DIALOGUE UNDERSTANDING AND DIALOGUE DESIGN: FROM SCIENCE TO ENGINEERING

Anthony Graham Lambie

University College London

1999

Thesis submitted for the degree of Doctor of Philosophy
at the University of London

DEDICATION

This effort could not have been made and the thesis which resulted, whatever its intrinsic merit, could not have been completed without the support and love of many people, and particularly my mother Agnes, my wife Sheila and my daughters, Davina and Esme.

ACKNOWLEDGEMENTS

It has been a very enjoyable and stimulating time at the Ergonomics and HCI Unit. Everyone has been enormously helpful and friendly. I want to thank, in particular, my supervisors: first, Dr Andy Whitefield, and then, and for the most part, Professor John Long. They encouraged me through the difficult times I had with an ailing parrot and teenage children. I have learnt a lot. Without John Long's interest in and reflections on knowledge – and its good use – I would have found it difficult to remain buoyed up. The work was supported by a studentship from the Engineering and Physical Sciences Research Council; Logica, as the industrial sponsor, contributed to my upkeep for the first two years, and started the thesis ball rolling.

ABSTRACT

The thesis exemplifies, illustrates and argues for an epistemology of Cognitive Engineering (CE). Others have proposed an ontology of the discipline. To give cogency to this ontology of CE with its own practices and method of validation, both radically distinct from Psychology and Cognitive Science, separate streams of science and technology have been asserted. Thus, design knowledge is not scientific knowledge, and one cannot specify the former with the latter: design problems with their own specific requirements define their kind of knowledge.

However, where the need is for the design or evaluation of a system which mimics human cognitive behaviour, how can it be met without scientific (or descriptive) knowledge? The project of designing or evaluating Natural Language Dialogue (NLD) systems presents just such a need. Two solutions are required:

(i) to show how a transition might plausibly be made from scientific (or descriptive) knowledge of linguistics to explicit design (or prescriptive) knowledge of NLD systems, differentiating resources which are usable from those which are not: an NLD framework; and,

(ii) to endorse this transition with an argument for the epistemological validity underpinning the move, and to provide a general foundation for the relationship between science and technology (comprising applied science and engineering): a foundational framework.

The thesis responds to both needs: promoting, criticising, and supplementing linguistic theories as the basis for the framework to satisfy (i). And to fulfil (ii), concepts adapted from Speech Act theory are combined with elaborated versions of key terms in the ontology of CE in order to argue for the commonalities of science and engineering. These arguments are situated via polemical disputes about the status, with respect to science, of Human Computer Interaction (HCI) and design disciplines more generally.

CONTENTS

FIGURES

PART 1

"Is it not possible that the next century may see the birth, through the joint labours of philosophers, grammarians, and numerous other students of language, of a true and comprehensive science of language? Then we shall have rid ourselves of one more part of philosophy... in the only way we can get rid of philosophy, by kicking it upstairs." ('Ifs and Cans', (J L Austin, 1961))

CHAPTER 1

The thesis is an investigation into design knowledge and its relationship with knowledge of other kinds. The introductory chapter describes the terms of this investigation, its subordinate goals and their motivations.

At bottom there is a practical difficulty which requires resolution. The route to this resolution is a conceptual one; the conceptual problems are wide-ranging. The reconfiguration of the conceptual elements which may satisfy these problems may, in its turn, have positive practical consequences. The problems, therefore, though abstract, are not analysed a priori or in a vacuum. The fact that the thesis is in two parts reflects this. The development of the general concepts in Part 2, the foundational framework, are never completely separated from the work in Part 1, where an answer to the practical problem is offered. Conversely, the solution in Part 1 is affirmed by the claimed general utility of the foundational framework in Part 2.

Thesis Introduction

Objective knowledge of the world has long been thought of as one and continuous, and this unity and continuity of knowledge has been the prerogative of science. Before the rise of modern science, knowledge of nature was thought to be the product of two processes – observation and reason, with the second preponderant. This was the Aristotelian attitude which prevailed into the Renaissance, and although a proper engagement with experimental investigation had begun by then, it is arguable that the attitude has maintained some momentum and that the observational/rational stance still holds sway. The possibility that knowledge of equal importance might be derived from the systematic design of utilitarian artefacts has been eschewed. It is only relatively recently that the question has been raised as to whether pure science and engineering are as simply associated as is suggested by the image of the application of knowledge of the former, thought of as primary, to the solution of problems of the latter, thought of as derivative. For example, Gibbons (1983) concludes,

> "...enough has been said to indicate the existence of a model of the interaction between science and technology in which both science and technology constitute *autonomous streams of knowledge*." (my italics)

I shall return often to this theme in the course of the thesis, paying particular attention to it in the second part (Chapter 11). This idea of the independence of science and engineering was much discussed in the 1960s, 70s and 80s. Historians and philosophers of science (and technology) have discussed it in general terms (and I deal with some of these in the penultimate chapter); and in the area of cognitive ergonomics important investigations of this question, of a more conceptual nature, were undertaken by Herbert Simon (1969, 1981); while, in the wider field of aeronautical engineering, many important issues have since been raised by Vincenti (1990).

Others, notably Carroll & Campbell (1989), of whom Carroll is an intellectual descendent of Simon, perceive the difficulties encountered by Human Computer Interaction (HCI) as the result of the straightforward (and unthinking) application of psychological knowledge; and they provide a unique solution to this design discipline's status through its re-categorisation as a peculiarly different kind of *science* devoted to design. I provide a critique of this approach along with that of Simon and of one other representative of design work (in the second half of the thesis). Though I believe Carroll's solution is wrong, it is an interesting solution, but it is a solution which has been attempted before the alternative expressions of the problem have been properly acknowledged. He rightly points out that the application of scientific psychological knowledge is a flawed approach to solving HCI problems, but fails to consider that this might be so because design is radically different from science. Instead he claims that scientific knowledge is somehow already incorporated in design (see Chapter 10), and I shall argue that his analysis, far from resolving the issue, creates further confusion.

A recognition of science and engineering as radically distinct disciplines does, however, pose a new problem, which may be what Carroll has chosen to ignore. If they are so distinct, in what sense do they interact at all? It is widely accepted that they do, and if they did not it is arguable that there would be problems even with *understanding* what had been engineered or designed. Among the principal aims of the thesis is that of avoiding the confusions which arise from the conflation of science and engineering/design; and while keeping the disciplines distinct, to argue that they have something in common, and try to make clear what that common denominator is.

The thesis attempts in the first part to bring science (in the form of linguistics) closer to engineering/design (in the form of Natural Language Dialogue (NLD) design) to solve a practical problem: that of evaluating NLD. It does this, however, not by conflating the disciplines, nor by the application of one to the problems of the other, but by arguing for a *plausible* transition from scientific knowledge (linguistics) to engineering knowledge (NLD system design). In order to succeed in this plausible transition, the analysis postulates criteria for distinguishing useful knowledge for design from that which is useful for science. This exercise, and these criteria, point the way (in the second half of the thesis) to a general solution to the problem of the relationship of scientific and technological knowledge, which I call the foundational framework. And this foundational framework, if it is correct, provides the justification for the conditions which underpin the plausibility of NLD framework.

It is important, therefore, throughout this thesis, since it takes some abstract twists and turns, to bear in mind that it is not simply addressing issues in the literature (although it may do this too), but that it is trying to solve a practical or operational problem, and to provide the context in which such a problem can be solved. Conversely, indeed, it is arguably a valid claim that the 'issues in the literature' such as the limits of scientific knowledge of cognition, what can be done in AI, scientific realism etc., are only properly examined when they are considered in relation to practical problems.

Operational Problem

The origin of the practical problem was the suggestion by the software company, Logica, the leader in the European-wide Esprit project SUNDIAL (Speech UNderstanding and DIALogue)[1], that some work needed to be done to elucidate what was involved in the stubbornly difficult task of evaluating NLD systems. Fraser (1991) had already suggested there was a problem in understanding how to specify such designs and he had commented, "clearly, a speech understanding system which

[1] Esprit Project No. 2218

modelled the exact behaviour found in the corpus and nothing else would not be useful"[2].

The implied challenge was how we *prescribe* from a mere *description* of a corpus of dialogue material, and this in its turn rested on the difficulty of discerning the salient features of natural dialogue which should be the goal of design – tantamount to admitting that the same difficulty would surface in the evaluation of the output from any NLD system. (The MMI^2 project[3], represented by Michael Wilson of the Rutherford Appleton Laboratories[4], had concerns similar to those of the SUNDIAL workers, when faced with question of how NLD should be specified.)

Logica used certain resources from linguistics and the philosophy of language, devising metrics from them. The researchers were unhappy with these metrics, and I judged the reason to be that the linguistic knowledge involved was not of the sort which could be used for design. I suggest in Chapter 2 and 3 why this might be so. The problem, prima facie, was connected with the way in which our descriptive and explanatory knowledge might be tapped for design purposes.

In the field of language processing, opinions vary as to the merit or correctness of moving from scientific or descriptive knowledge to that of design. In NLD design, after the early but doomed optimism about the imminent construction of talking machines, a new opinion was expressed by Nickerson (1976):

> "there are two contentious remarks that I would like to make regarding the notion of conversational interaction between persons and computers. The first is that the differences between the person-computer interactions that take place today and interperson conversations are far greater that the similarities between them. The second is that interperson conversation may be, in some respects, an inappropriate and misleading model to use as a goal for person-computer interaction".

---

[2] SUNDIAL involved the classification and tagging of dialogue corpora as a prelude to specification.

[3] Esprit Project No. 2474

[4] Exposed during a presentation given at the University of Surrey in '94

Two things should be noted about these remarks. The first is that Nickerson is careful to qualify the comparison he makes between person/computer and person/person interactions by confining the time of the comment – "today". So, the differences noted might not be so great in the future, or even now. That comparison, in any case, presupposes some way of making such an assessment, and it is difficult to see how this could be done without reference to linguistic knowledge. Indeed, his second point implies (by "in some respects"), and elsewhere in his paper he states, that some features of person/person interaction should also be considered important features of language interaction between man and machine. Again, a reliance on some linguistic knowledge for design and evaluation is implied.

At the other end of the spectrum, a recent paper by Perlis et al (1998) argues

> "that there may be a core set of meta-dialog principles that is in some sense complete, and that may correspond to the human ability to engage in 'free-ranging' conversation. If we are right, then implementing such a set would be of considerable interest".

Two comments will suffice: it is, first of all, not clear that we need go so far as to claim a complete description of essential principles in order to benefit from some knowledge of such principles in design, and, secondly, the key question is in what 'implementation' consists. It is this 'implementation' which bears on the kind of relationship which exists between descriptive/explanatory and design/engineering knowledge. Any implementation would presuppose a means of representation which could act as the vehicle for the transfer of knowledge of linguistic behaviour and its subsequent employment in the design process.

I do not support either of the positions outlined above, citing them only to illustrate the range of opinions, but the important point is that both raise the question of how natural language and designed language are related and, implicitly, how the knowledge of the former can be expressed and compared with that implemented in the latter.

Elicitation of Design Knowledge – a Mysterious Process

Two methods which allegedly extract this kind of knowledge are the 'Wizard of Oz' and the 'ecological' methods, neither of which are explicit in their workings. Fraser's (1991) comments on the difficulty of using data for design imply that the alternative would mean a search for what should ideally be a principled, i.e., rationally explicit connection between the data and the design. However, he has pursued a line of enquiry which sidesteps this question. In Fraser & Gilbert (1991), he is adopting a method which is *assumed* to make the transition from data to design – the Wizard of Oz (WOZ). The authors introduce the method with a backward glance at the problem: "The problem, according to von Hahn (1986) is that 'we have no well developed linguistics of natural-language man-machine communication'". In the absence of such principled and appropriate knowledge they proceed to expose a simulation exercise to refine the building of an NLD system. The argument for the *practical* value of the method is cogent, but it is not easy to rationalise it as based on anything less intuitive than previous approaches. What it *appears* to do well is focus the designer on the target area of the design problem. Where the intuitions come from which guide the simulation design itself are not better understood; and Fraser & Gilbert do not claim that they are.

The other solution to the problem of the source for design knowledge input is the 'ecological' study (cf. Dowell, Smith & Pigeon,1998). Although it is more explicit than the WOZ method about how the putative descriptive knowledge is employed, just how the appropriate knowledge is extracted is largely intuitive. It rests on the reasonable assumption that if you want to know something about how people should behave in certain circumstances then you should study them behaving in a situation which meets those circumstances as closely as possible. It emphasises the whole dynamic interaction, and witholds theoretical assumptions associated with a 'classical' mechanistic approach; and behind it lies the belief that the agent and that part of the world with which s/he is interacting are 'coupled': that there is an intimate relationship between the agent's representation and the complex of tasks of which the 'work' is made up. These properties underlying the method need to be more clearly and explicitly addressed, and they will be later. However, the method's immediate virtue, like that of Fraser & Gilbert's, seems only to be its better targeting of the area of the design problem.

Effective Design without Knowledge of Nature

Naturalness does not thereby imply *effectiveness*, but effectiveness is a key aim of systematic design (implicitly at least); and in any project of rigorous design, e.g., engineering, it should have explicit expression. We do, however, appear to have reached a point in the development of interactive designs when such qualities as naturalness are sought as part of satisfying that aim. We can identify this desire for naturalness on the part of designers with the need for linguistic knowledge to support the design of NLD systems, since it is believed that this knowledge would provide the specifications for the understanding of the relevant natural behaviours: to be effective the interaction must be more natural, i.e., it is a necessary not a sufficient condition. Perhaps the value of employing the WOZ or an 'ecological' method is that it answers the requirement for feeding the design with descriptive/ explanatory knowledge (natural knowledge); and the target behaviour is thereby scoped roughly for an *effective* design outcome. The difficulty then with the above methods is that measuring effectiveness is not only meeting the requirements of the design but also how well such knowledge, as is targeted, has been elicited and transferred.

If, however, one adheres to the strict view of cognitive engineering (CE) argued, for example, by Dowell (1993), then one should also hold to the view that engineering generates its own knowledge, and does not involve scientific knowledge of any kind[5]. The implication is that knowledge accrued to solve a cognitive design problem relies on a relatively mysterious process, which cannot be a good basis for a purportedly systematic explicit discipline such as CE. The argument might be, however, that these implicit methods are temporary stop-gaps until a more principled method can be found. I shall, however, criticise the kind of separate development described by Dowell (1993) claiming that it poses an epistemological problem.

---

[5] Though interaction is acknowledged to take place. This issue is dealt with in some detail in the penultimate chapter.

This criticism of Dowell is largely negative, but a positive view, based on a claimed better understanding of scientific and technological (including engineering) knowledge will also be promoted. For how can *any* useful knowledge be gathered, say, from the 'ecological' method, which appears to derive its usefulness in the design process by providing *declarative* linguistic knowledge roughly configured through scoping. That it to say, given that it works, how does it work? It is my intention to broaden the basis for knowledge interaction, to allow for such a rationale, while maintaining the quite distinct disciplines of science and engineering. I want to maintain and support the project of engineering, as distinct from HCI as applied science and craft, while arguing that there is a way of specifying the common denominator of science and technology (including engineering); and to provide an epistemological setting for cognitive engineering which draws on, and adapts, the central concepts of what Dowell (1993) and Dowell & Long (1989, 1998) have established as its ontology. In other words, the thesis arguments are not at all hostile to the ideal of CE as a distinct activity; rather they should provide consolidation for it, while providing a rationale for epistemological traffic between science and engineering consistent with, but not the same as, the widely held view of the application of science.

To summarise: the thesis adopts the perspective that the goal of designing a natural language dialogue application is a legitimate one and that the difficulty of understanding its principled design or evaluation arises from an unclear view of how scientific (descriptive and explanatory) knowledge can be interpreted for use in design. The argument thus permits the continued adherence to a strict and separate discipline of cognitive *engineering*, while admitting epistemic interaction with cognitive *science*; and what it offers, in addition, is some attempt at making explicit just what kind of interaction is involved.

So, one aspect of the relationship of science and technology might be expressed as the problem of knowledge communication. Another slightly different aspect, alluded to above, which has to be addressed is the issue of whence comes the *authority* for engineering knowledge – from science or from engineering itself. This thorny problem, I believe, impels researchers into the arms of science. These aspects will take on distinct analytical features as the thesis develops, but first, I shall describe briefly the genealogy of this leaning towards science.

Systematic Design's Dependence on Science

The first step which serious design thinkers had to take was the legitimisation of design as a systematic set of practices: to separate it from craft. Two writers (and two books) in particular stand out: Simon with "The Sciences of the Artificial" (Simon, 1969) and Alexander with "Notes on the Synthesis of Form" (Alexander, 1964). Both supported design as a rigorous discipline, Simon by conjoining scientific knowledge with the practices of the production of artefacts through his notion of the inner make-up of the artefact with reference to its context and the requirements or goals which determined the interface between the two: his aim, while recognising that science is concerned with "how things are" and engineering, for example, with "how things might be – in short, with design", is to show how "a science of the artificial is possible and to illustrate its nature"[6]; and I shall criticise his positive/normative stance in the second part of the thesis.

Alexander emphasised the distinction between the 'unselfconscious' design of craft and the 'selfconscious' and explicit methods of systematic design, linking scientific knowledge with that of design, by the provision of a concept which bridged the knowledge of the two kinds of discipline – the 'constructive diagram'; and, like Simon, described the duality of the artefact as an item in strict relations with a context. Both, however, while consolidating the security of design or engineering knowledge, did so by failing to emphasise the autonomy of design knowledge, leaning perhaps too heavily, in Simon's case, on the kind of knowledge guarantees which science traditionally provides.

Simon (1969) is led to this tacit 'collaboration' with science through his interest in cognitive behaviour as artifice in the sense that it is a goal-directed product. He, firstly, expresses the general claim that a science, encompassing objects requiring for

---

[6] It is worthwhile noting that, according to his autobiography (Simon, 1996), he uses the word 'science' in the phrase 'artificial science' to contrast with 'art' rather than engineering.

their understanding the assumption of teleological 'forces' (biological function, natural selection) happens to embrace the particular business of systematic design – one science of the artificial. Secondly, he asserts that design is of the essence of human cognition[7], and understanding it, for Simon, meant understanding human cognition: "the proper study of mankind is the science of design" – another more particular science of the artificial. This position is interesting, but judging by his treatment of the connection between the normative and the positive (examined in Chapter 10) he actually sees both the cognitive and the non-cognitive sciences of the artificial as concerned with *understanding*. His position, he writes, "is entirely consistent with treating natural or artificial goal-seeking systems as phenomena, *without commitment to their goals*" (my italics). I believe, however, it is just this commitment which one cannot set aside when making the distinction between science and engineering.

Alexander also treats design in an abstract fashion (static), and when he considers the process (dynamic) he falls into the error of describing the development of a design as like "hypothesising", which is a stage of *scientific* reasoning. Alexander's ideas seem less prone to mislead us to conceive of design as a species of science, since he is so wholly involved in the *design practice* of architecture.[8] Alexander is not trying to *comprehend* buildings as Simon is trying to *comprehend* individual cognition or social behaviour.

There is, therefore, in both Simon's and Alexander's work evidence of some vacillation about the kind of knowledge with which design is carried out. It seems to me that this ambivalent attitude which they exhibit depends on a picture of science and technology which is misconceived, and I shall attempt, in the second part of the thesis,

---

[7] "The laws that govern these strings of symbols (language), the laws that govern the occasions on which we emit and receive them, the determinants of their content are all *consequences* of our collective artifice." (my parenthesis, my italics) (Simon, 1969)This gnomic statement goes further than conventional cognitive science, but is not developed by Simon. I believe it is better accommodated by my foundational framework, is manifest in the NLD framework analysis, and commented on in the concluding chapter.

[8] As we shall see in Chapter 10, Dasgupta (1991) bases his reductive perspective of design on just such a use of the concept of hypothesis as part of the design process.

to weaken this misconception by recasting the models of science and technology to produce a more stable image of their distinctness.

Language Knowledge and Cognitive Engineering

The context of the source problem for the thesis is the culture of HCI as an engineering discipline (HCIe) (Dowell & Long, 1989; and Dowell & Long, 1998). That culture encourages the addressing of a particular design problem, i.e., a particular problem defined by a set of practical requirements. The design of an NLD system is at odds with this kind of project insofar as it aims at achieving naturalness over and above meeting practical requirements; and the rationale for the more 'austere' approach is that technology and science are distinct and developed in parallel. That is to say, since the design of an NLD system would depend on some knowledge of what a natural language system was, it would depend, to some degree, on systematic or descriptive (scientific) knowledge; but this claimed dependence is contested by the proponents of HCI (as an engineering discipline) and cognitive engineering (Long & Dowell, 1989; and Dowell, 1993). The question therefore of whether any operational or practical problem with respect to NLD design could be resolved would turn on the prior question of whether a communicable relationship between scientific and technological knowledge (in particular, engineering) can be established.

If my concern is the possible transfer of scientific knowledge of language to the sphere of NLD design, it is inevitable that I should pay some attention to the question of what constitutes that kind of knowledge. The introduction, therefore, to any approach to the question of design resources for NLD will address the issue of how linguistics knowledge should be understood.

I am going to take the object of scientific study to be that which is 'natural'. This point of view maintains the distinction between scientific knowledge and engineering knowledge as concerning nature and artifice, respectively, and directly bears on the convergence of the two implicit in the project to support the design/evaluation of NLD. That it is convergence hides more than it illuminates; and it is for the thesis to make the relationship between the two terms of the relationship plainer. However, any resolution of the relationship might have implications not only for the meaning of 'natural' in NLD design but also for the significance of 'natural' in the phrase 'natural world', and therefore what is meant by knowledge of nature or scientific knowledge, and how it is related to everyday knowledge of the world, if that is different. These then seem like reasonable assumptions and a good starting point. Things may turn out differently.

There are, of course, also different ideas of what scientific knowledge is and it is not my intention to arrive at a definitive answer to any speculations on this question, only to put forward a likely candidate both by virtue of its cogency and in terms of its authority – as derived from respected proponents; and most importantly for my thesis claim, that it should show, under at least one interpretation, how it might be consistent with an integration of scientific and technological knowledge, which solves the various problems I have sketched.

Evaluation, Design and Engineering

It should be said at the outset that I shall deal with evaluation in its general relationship with design; its dependence on knowledge of two kinds: requirements and values on the one hand, and features and qualities on the other – the prescriptive and the descriptive. Since I am not dealing with the evaluation of a specific system, in a specific setting, but rather with the general question of how evaluation is tied in with issues of scientific knowledge, I shall not consider the details such as are treated by Galliers and Sparck Jones (1993). These authors note that it is not surprising that "evaluation in NLP has lagged behind that in other areas of information processing and NLP application systems: *if you know already that your system cannot do very much...there does not seem to be much point in embarking on big-time evaluation.*" (my italics) By "big time evaluation" I take them to mean serious assessment of whether the language employed by the system is 'natural' enough, given, of course, the particularities of the case. Nowhere is this question addressed explicitly. The bibliography of Galliers & Sparck Jones' book has no other references than those concerned with language evaluation practice: satisfying some de facto set of requirements. It is in this latter sense that it is only a survey of what sort of evaluation has been carried out prior to publication. It is not unsurprising as this kind of global evaluation is not available, and will most likely develop with the development of the technology of natural language processing and the arrival of systems which incorporate what Galliers & Sparck Jones call ‘exigent’ (or thorough) processing. However, what can be done now depends to an extent on what is possible, and that can be determined to some extent by reflection on the nature of design and evaluation.

When Galliers & Sparck Jones come to sum up the major issues of NLP (Natural Language Processing) (a list which demonstrates that they have been involved in a survey of what evaluation practices are rather than what they should or could be, and therefore of what sort of knowledge might support evaluation) they ask,

> “Can evaluation criteria, measures and methods be generalised? That is, is an evaluation necessarily task – even application – dependent, or can specific evaluation techniques, as opposed to abstract notions, be applied across individual cases?”

and,

> “Finally, as a pervasive, underlying issue...*Should NLP evaluation be linguistically or computationally oriented?* That is, how far is NLP ‘just’ serving the machine simulation or emulation of present human language use, so evaluation refers to this, and how far is it serving new uses, so novel reference bases for evaluation have to be defined?” (my italics)

Addressing these last two questions will illuminate the possibilities for evaluation (I have alluded to Fraser’s comments on the inadequacies of a descriptive tagging of text. Galliers & Sparck Jones refer to Fraser’s remarks on SUNDIAL’s evaluative techniques, but do not quote his dissatisfaction.), and this should provide the framework for the issue of evaluative generality. Indeed, it is just such issues as these which are at the focus of attention of HCI as engineering. This conception of HCI as an engineering discipline (HCIe) emphasises the design of an artefact as the attempt at a solution to a *particular* problem, and its practices do not assume that another problem which has comparable features is necessarily solved in the same way: HCIe pays great attention to the manner in which the problems might be similar. However, this stress on *particularity* may pose difficulties for just how one can justify *generalisation* in design or evaluation. It is my intention therefore to amplify these virtues of *practice*, which HCIe possesses, by developing a framework or rationale, which justifies and binds the related instances of successful engineering in terms of certain neutral concepts common to science and technology; and to do so via an exemplar of NLD design or evaluation. The generality gained should be for 'mediated' evaluation (Scriven, 1964) – a rationalisation of 'pay-off' and 'intrinsic' evaluation. A good example, in practice, is to be found in Carroll, Singley & Rosson (1992).

Where Evaluation and Design Converge

The emphasis on the import of these issues is at a level of analysis which is removed from the kind of practices and methods of evaluation with which Galliers and Sparck Jones are concerned, and means that the investigation is equally valid for the concept of design as it is for that of evaluation. The specific practical or operational problem was identified as one of evaluation. However, the NLD framework I am proposing, and the epistemological setting (the foundational framework) within which this framework is constructed, is valid with respect to design also. My initial concern (the NLD framework) is with how language should be conceived, both for design and evaluation purposes: how, for example, this representation supports communication between design workers in the MMI^2 project (referred to above), as it does evaluation for SUNDIAL workers.

The framework highlights a static rather than a dynamic perspective on the artefact and its conception. Alexander (1964), in the chapter entitled "Goodness of Fit", sees the process of design, and by implication that of evaluation, as the fitting of *form* to *context*, where he defines form as "a part of the world over which we have control", and context as that part of the world which puts demands on this form", and, he goes on, "anything in the world which makes demands on the form is context". The latter authorises us to say that the fitting of form to context is common to both design and evaluation as well as the fitting of a representation of context to a representation of form, since it is not only the state or not of fitting but a report of that state; and we can see from the following quotation that design and evaluation are simply aspects of 'the fitting of form to context':

"It is common practice in engineering, if we wish to make a metal face perfectly smooth and level, to fit it against the surface of a standard steel block, which is level within finer limits than those we are aiming at, by inking the surface of this standard block and rubbing our metal face against the inked surface. If our metal face is not quite level, ink marks appear on it at those points which are higher than the rest. We grind away these high spots, and try to fit the block perfectly, so that there are no high spots which stand out any more. This ensemble of two metal faces is so simple that we shall not be distracted by the possibility of multiple form-context boundaries within it...." (In other words, not the myriad specific claims) "...If we wish to *judge* the form without actually putting it in contact with its context (analytic evaluation), in this case we may also do so. If we define levelness in mathematical terms, as a limitation on the variance which is permitted over the surface, we can test the form itself, without testing it against the context. We can do this because the criterion for levelness is, simultaneously, a *description* of the required form, and also a *description* of the context." (my italics).

Conclusions

There is a dichotomy at the heart of the thesis: that between science and engineering. This dichotomy can be formulated in several ways; and I try to develop some clearer bases for it. A dilemma lies behind the operational problem: if science and engineering are continuous then how are they to be separated when it is necessary, as it appears to be sometimes? If, on the other hand, they are not connected in this fashion, how do we talk about the one in terms of the other (and trade between them), which we apparently do? The dilemma is patent throughout most of the thesis, even when I drop down to its lower analytical expressions. In the end, I hope to resolve it by making a distinction between those aspects of the two components – science and engineering – which are common and continuous, and those which are not: "(O)ne reconciles...two views by the time-honoured device of making them apply in different situations" (Rawls, 1955); in this case, by making the two views apply to different aspects of the two discipline types.

The dichotomy poses a more specific and serious dilemma: if engineering gains autonomy from science it appears to lose its solid foundations, since, in pursuing an austerity which relinquishes what we know, i.e., our description of the world as it is, it may isolate itself epistemologically. I argue that this danger is present in HCI as an engineering discipline, and is not addressed in Dowell (1993). However, Alexander also (shortly after the passage quoted above) raises the question of how we "cognitively experience the sensation of fit" i.e., positive design knowledge, since he claims that a complete 'field description' of the artefact is unattainable and, in practice, the designer is faced with a finite and limited set of misfits on the way to the finished successful design. But scientific theory is, in a similar way, never *positively* confirmed, and this means that both science and engineering knowledge appear to rest *methodologically* on the Popperian paradigm of high refutation risk as the closest approach to certainty. How can we, from this minimalist position, arrive at the positive view we undoubtedly have of design through engineering, or, indeed, of the understanding of reality through science?

I believe that the foundational framework, which I offer, (a) resolves the original dilemma, while avoiding the conflation of what I believe are the *three* important components of what is referred to as science and technology: pure science, applied science and engineering; and (b) it may be that engineering and science can be of mutual assistance in this respect, that they too can converge in the achievement of secure knowledge, bolstering each other's weak points.

HCI, which I treat as a sub-discipline of CE is commonly described as multi-disciplinary, and this attribution allows not only the admission of different practices but also different kinds of, and perspectives on, knowledge: Carroll (1997) writes, "Perhaps the most impressive current feature of the area (HCI) is its fragmentation". If, however, the foundational framework provides a proper grounding and avoids epistemological isolation, the kinds and practices of fields of interest which might contribute to HCI (or CE more generally) will do so in terms of *the aims of the discipline* rather than in terms of the origins of the knowledge or the distinct practices of the contributing fields of interest, thereby, I hope, slowing down or reversing the fragmentation of the field referred to by Carroll. The thesis aims to be inclusive of the diverse methods and perspectives heretofore hostile to the idea of the traditionally rational paradigm of engineering, not by offering a discipline within which they have to lose their identity, but by broadening its conceptual basis to accommodate them.

Survey of the Thesis

Chapter 2 turns to the subject of how natural language might be represented. I briefly survey the modern view and its development, concluding the first section with a prominent philosopher's opinion that language has a dual aspect: representation and activity. The next section is a less abstract analysis of one of these faces of language – representation. I summarise and criticise a study of story grammar to assess its value as a semantic tool, i.e., whether it could be used in a constructive way to reproduce (or predict) a meaningful story. I conclude that whatever story grammars do, they fail to do this; the study itself makes suggestions as to the reasons why which may point the way forward.

Before, however, I undertake to consider the options of the way forward, I return to the source of the operational problem and set out SUNDIAL's suggested resources for evaluation. If the resources are of limited value, how and why would this be so; and, are there criteria for distinguishing what is and what is not useful knowledge, for the purposes in hand – design/evaluation? I devote Chapter 3 to investigating how representation might be understood, and developing rules derived from Searle's philosophy of language which could act as such criteria and be consistent with the understanding of representation. It is important to bear in mind that these rules are adapted with a different aim in mind; and I turn them on the speech act itself which Searle would not do.

In the following chapter (4), having concluded what is missing from the kind of resources which SUNDIAL draws on, I suggest a component – planning, which might be added to speech act theory, providing a 'causeway' from a descriptive and explanatory (scientific) theory to a design/evaluation framework for NLD. Such a theory exists, in the form of the Plan-based Speech Act (P-BSA) theory, and might serve the purpose. However, before examining it more closely, I have to confront some objections to planning as a fundamental component in cognitive representation. I conclude that they can be at least circumvented. In the context of planning and dialogue, I draw attention to another paradigm of planning and SA theory, but the decision about which paradigm to adopt must await a more detailed analysis of P-BSA theory. I undertake this analysis next, in Chapter 5, looking at how well the concept of planning, as conceived by the authors of P-BSA theory, fits with a proper breakdown of the concept of 'illocution' central to SA theory. It seems that it does not. I close the first section with a revised view of planning and cognition.

Even revised, is the P-BSA theory adequate to be the framework for NLD? I consider, in Chapter 6 how it might be categorised (along with how its authors categorise it), and if it is so categorised what implications this might have for how it should be supplemented by a theory of a more general kind, e.g., Relevance theory. The chapter also argues for this categorisation as an essential aspect of the representation of such a conceptual framework as the thesis is attempting to develop.

Chapter 7 deals with Relevance theory, sketching its claims, examining its status with respect to design, and finding it wanting in a similar way to the other resources turned to by SUNDIAL. This analysis is another opportunity to work with the lower-level dichotomy which was developed from Searle's philosophy of language; and, with the help of a critic of Relevance theory, to elucidate better that dichotomy and its constituent parts. It is also the occasion for employing another developing concept, that of the domain, to elucidate the alleged flaws of Relevance theory as a comprehensive account of cognition.

The next chapter (8) is the conclusion to Part 1 of the thesis. It is the fulcrum of the argument, which rests on the assumption that NLD is a kind of engineering which can draw on kinds of scientific knowledge. In this chapter, I recapitulate the ideas which have been developed through the analysis and criticism of the previous chapters and which make up the NLD framework, and I call on a modern account of language to endorse it. I show that it maps to the established ontology of HCI as an engineering discipline (CE); and take it as equivalent in form to a framework of CE. I then re-express the established terms of this CE framework in ways consistent with the development of the NLD framework, and with the subject of Part 2: the foundational framework.

The next two chapter (9 and 10) are devoted to consolidating the epistemological validity of the CE framework: a more thorough articulation of the idea of the 'domain' (with an instance of its employment in public policy) in Chapter 9, and in Chapter 10, a detailed consideration of the relationship between descriptive and prescriptive knowledge, in the light of three design thinkers' views. This chapter concludes with an adapted version of Searle's (1969) conception of 'institutions' and 'brute facts' as the joint basis for the normative element of practical reason. These two chapters prepare the ground for the foundational framework.

Chapter 11, in its first section, tries to position scientific knowledge in a more balanced relationship with that of technology, arguing that there is much to suggest that the traditional view of science is partly what distorts the conception of engineering; and the ideas of a philosopher of science are invoked in order to support what might be described as a more neutral perspective on scientific knowledge. In the second section of this penultimate chapter, I describe the elements of the foundational framework, which allows for an even-handed share of knowledge going to science and engineering – a symmetry making sense of, and being made sense by, the above perspective on science.

I conclude, in Chapter 12, with suggestions and implications of the arguments, which might point to future work, setting the theme of the thesis in a historical context, and hoping that science and engineering can collaborate effectively: fulfilling the promise of the birth of modern science and technology; and understood by, I believe, at least one renaissance man, some four hundred years ago.

CHAPTER 2

In this chapter, I shall refer to, and briefly discuss some pertinent ideas of modern philosophy , of a general nature, on the status and function of natural language (later, in the first part, I draw on and adapt concepts from the philosophy of language). It is important to separate out concepts of representation, in language and, of language; and mark the differences. Representation is central to the work of explicit design and it is essential to establish what that representation is with respect to language. Since I am assuming the radical division of the science and engineering disciplines, it is necessary that I ensure that the representation entailed by the framework for evaluation or design is fit for the purpose in hand – not necessarily deriving from linguistics. The examples in this chapter should set the level of representation and the problems at that level. The subsequent chapters will deepen the analysis after defining some general categories of representation and their associated rules.

Language Representation

Introduction

The chapter will provide two views of language: first, an abstract account and classification of language; and, second, an attempt at testing one view in an operational setting. Developers of LP (language processing) systems are trying to formulate specifications or representations of dialogue for that purpose; and I must bring the argument round from the purely conceptual to some expression of representation for design which (in accord with the aims of the thesis) should facilitate linking an operational framework and some account of language. I cannot begin to do this without illuminating the possibility of such a connection. As well as this example of the application of one view of language, which is considered at the discourse level, I shall briefly touch on the analogous view of language at the sentence and word level. I shall consolidate my criticism of the language view in question by citing the arguments of a well known researcher in the field of language design. Next, I shall expose SUNDIAL's resources for language design, and, in the light of the foregoing analysis, I shall raise the question of what sort of resources we should be looking for, with language design in mind.

Language and Meaning

The problem of evaluating or designing NLD is intimately tied up with language's status as the bearer of meaning, and what this status has to do with its mode of representation as a factor in NLD design. Language and thought, which it represents at least in part, are problematic in their apparent mutual dependence. Questions such as, "Can we think without words?", and, "Can we *speak* without thought?" (where 'speak' means something more than make noises which can be taken for words) bring this mutual dependence into focus. These questions are related to what we consider to be the main function of language. Is it essentially cognitive or communicative? Strawson (1973) refers to the two approaches as the question of whether language is essentially 'formal semantics' or 'communication-intention'; and he ends his essay with the remark, "as theorists, we know nothing of human language unless we understand human speech." I want to consider the ideas of someone who deals with this commonly accepted dichotomy, and perhaps deepens its analysis.

Dummett's Reflections on the Dilemma of Language's Function

Michael Dummett is one of the foremost contemporary philosophers who is concerned with language. Some academic thinkers are concerned with more predominantly technical matters, and their thinking reflects a view of reality which is alien to non-specialists. Dummett, however, although his arguments are dense and difficult, adopts a more usual worldview. His interests are those which might concern someone interested in understanding the context of NLD, being one of two philosophers[9] "who have studied the relationship between truth, language and reality" and who "demand attention in virtue both of their originality and their influence" (Passmore, 1985). He sees philosophy as a theory of meaning, and thinks that we can only understand things imperfectly as long as we do not grasp the relationship between language and the world.

Dummett (1989) begins his paper by writing, "language, it is natural to say, has two principles: that of an instrument of communication and that of a vehicle of thought."; and, he notes, the question has been posed as to which is primary. He brings together key figures who have addressed questions of meaning in the philosophy of language. Dummett introduces Strawson's endorsement of Grice's doctrine that linguistic meaning is centrally associated with the intention of the speaker to get the hearer to grasp his aim to communicate, and that what is communicated is a belief of the speaker's: what is referred to above, by Strawson, as 'the communication-intention' approach. It is contrasted with Frege's view that language is a 'vehicle of thought': a

[9]The other is Donald Davidson.

version of what Strawson calls 'formal semantics'. After a detailed critique of Strawson's position, Dummett concludes that the opposition is a shallow one.

He suggests that "the true opposition is between language as representation" (i.e., as corresponding with thought or containing it[10]) "and language as activity". The question remains: Is one or other primary? He likens the belief that language is primarily communicative to that of holding that once one has learned the game of bridge by playing with others one can then go off and play bridge by oneself: a reductio ad absurdum by analogy with talking to oneself as, in some sense, derivative of conversing with others. In other words, he writes,

> "the question at issue is whether it is because it can be used to communicate with others that it can also be used as a vehicle of one's thought, or whether, conversely, it is because two people are able to use the same language as a vehicle of thought that they are also able to use it to communicate with one another".

The problem with the comparison of playing bridge and having conversations is that it is not at all clear that communicating with oneself may not indeed be, like playing bridge by oneself, a degenerate form of the activity; and therefore no basis on which to discriminate between language as primarily either a vehicle of thought or means of communication:

> "It is indeed true that to describe someone as communicating with himself is to obliterate the whole distinction between using language as an instrument of communication and as a vehicle of thought....The true opposition is between language as representation and language as activity: and it is operating as an activity in soliloquy as much as it is in dialogue."

In spite of Dummett's criticism of Strawson, his own preferred dichotomy – representation and activity – cannot be treated as an *exclusive* disjunction. He judges that they are interdependent, and writes that

> "the representative power of language is both genuine and central. The illusion is...that this representative power can be isolated from all the other features of language..." and "...that those other features can be explained in terms of it".

[10] Language, it is implied, also stands for or represents – in some sense – facts or states of affairs in the world, but this representation has to be mediated by thought.

He has concluded therefore that language as representation cannot rest on language as communication, nor is the converse the case.

Dummett qualifies his idea of language as activity, writing, "not the banal point that to utter a sentence is to do something...rather, this: *that the significance of an utterance lies in the difference that it potentially makes to what subsequently happens.*" (my italics) He sees communication as "consisting in a complex interplay between linguistic exchange and related actions".

His position, then, is that the 'formal semantics' supporters (those in favour of the 'vehicle of thought' model and those promoting more skeletal models of language) are "shirking their responsibility" by not taking sufficient account of what he calls this " interplay between linguistic exchange and related actions". The formalists have often believed that something substantial can come of their work: that they can strictly derive substantial conclusions from merely formal premisses. One might be forgiven for understanding Dummett to mean that the Chomskyans' tendency to confine themselves to a narrow definition of linguistic knowledge – a very bare vehicle of thought – as what constitutes 'shirking responsibility'.

Reviewing, therefore, what is itself a key review of the literature of the philosophy of language, allows one to conclude that at least one prominent modern view of language sees it as possessing activity as an essential, but not more fundamental, aspect than its representative function.

For the time being therefore and for my purposes, it is enough to assume these attributes of language to be both important *aspects*. It is encouraging to find a writer whose reflections on the essence of language should conclude an active role for it[11], and who in employing the notion of 'responsibility', shirked by some, suggest that there might be some practical consequences thought to follow from further investigation of the mechanics of language. Throughout the first part of the thesis, the

[11] Allowing easier integration with the world of work: tasks as requirements of design to be fulfilled in and through language.

conflict of the two views which Dummett criticises and dismisses as superficial is repeatedly evident. As Dummett tries to resolve this opposition, so does the thesis' search for a framework for design of NLD, but in an operational context.

Representing or Acting

The gap between language as activity and language as representation may well be spanned by language as engineering: the 'artificiality' of which Simon (1969) attributes to human mental behaviour[12]. The philosophers, however, have been pursuing a general account of linguistic phenomena. The purpose of designers and evaluators is different. It may be desirable for them to arrive at views consistent, at some level, with those of the linguistic and philosophical inclination, but their immediate aim is to provide some form of specification – a more or less rigorous description – of what they seek in the realised system (designers), or clear means of knowing whether the finished system has achieved the properties required (evaluators). This has to rest on the means of representation for those purposes. Thus language, in the shape of a more or less formal representation of that of activity, in the senses alluded to above, by such as Strawson and Dummett should be examined, and its adequacy judged. I have taken a well known paper by Black & Wilensky (1979) which deals with one representational aspect, in a comparatively 'applied' setting as an illustration of the issues.

The polarity of representation and action as characterisations of language itself stands for two different ways of relating language to the world, and these different ways must be representable if we are going to solve design problems with them. Representation as a mode of language is usually thought of as tantamount to its property of *correspondence* with the world or with facts: on analogy with a picture. Language as activity, or action, has a less precise metaphor, but is usually thought of as *cohering*. It may indeed be the case that *formally* they are the same, but it is important to point out that that this formality may conceal a substantial difference

[12]Simon's view is contrary to that of Chomsky, for whom the behaviour is prefigured by, for example, linguistic competence: "that there are only a few 'intrinsic' characteristics of the inner environment of thinking man that limit the adaptation of thought to the shape of the problem environment"(Simon, '69).

which is relevant to design. Winograd (1975) makes the following point in the context of procedural/declarative controversy, which is a related dichotomy:

> "We must go below these labels to see what we stand to gain in *looking at* it as one or the other. We must examine the mechanisms which have been developed for dealing with these representations, and the kind of advantages they offer for epistemology." (Winograd's italics).

Black & Wilensky's Survey and Argument

Grammar as a Formal Representation of Discourse

A critical survey of story grammars carried out by Black & Wilensky (1979) is worth going into in some detail since they examine possible notations or representational schemes. The argument for this paper's relevance is that if grammars for stories are inadequate in some way then this inadequacy will also be true for dialogues, since stories may form part of dialogues; and dialogues, part of stories.

Story grammars, say Black & Wilensky, purport to fulfil three functions:

(i) they distinguish stories from non-stories

(ii) they act as models of story comprehension

(iii) they are memory models for stories

They assess their adequacy from a formal and an empirical point of view; and we are interested in (i) and (ii) principally, although (iii) might figure since dialogues are carried out over time and what Barbara Grosz (1981) calls 'focus' (carrying the thread of the discourse) must be maintained throughout.

What is meant by a grammar is a set of rewrite rules such as the following parsing grammar:

$$\mathbf{S \Rightarrow NP + VP}$$

$$\mathbf{NP \Rightarrow Adj + N}$$

$$\mathbf{VP \Rightarrow Vb + Obj}$$

where S = sentence, NP = noun phrase, VP = verb phrase. These symbols are all non-terminal. Terminal symbols would be 'man' in N = man, 'sat' in Vb = sat.

These rewrite rules could be of various types and Black & Wilensky take the following as typical examples for their purpose:-

(i) Finite State Grammars (FSG):

A ⇒ ab (where a is terminal & A is non-terminal);

(ii) What Black & Wilensky refer to as, Phrase Structure Grammars (PSG):

A ⇒ BCD – Context Free Grammar (CFG)

and

ABCDE ⇒ ABFDE – Context Sensitive Grammar (CSG);

(iii) A Notional Story Grammar

Story ⇒ Setting + Theme + Plot + Resolution

e.g. one non-terminal further expanded

Theme ⇒ (Event)* + Goal, where the '*' represents iteration.

Black & Wilensky recognise that Chomsky and others have concluded that FSGs, CSGs and CFGs are not up to the task of expressing NL sentence grammars. They think it, nevertheless worthwhile to make clear that neither are they up to the task of encapsulating story grammars, though it seems to them unlikely that they would be. Their aim is to bring out tacit implications of the story grammar's application. The exercise Black & Wilensky carry out is therefore useful to my aim because of this tacit level of analysis which their critique targets – comprehension of a radical kind and the criteria which enable this comprehension.

I want to take only certain arguments from Black & Wilensky. The comments on the formal inadequacy of particular grammars is a double-edged sword, without a more principled argument, and this view is reinforced by the vulnerability of Black & Wilensky to technical criticisms (see below). I believe, however, that they provide this more principled argument, and there is some point in introducing the principled argument via a review of formal grammars, if only to make precise what we mean by grammar and with what this might be contrasted.

The problem with the formal adequacy approach is that any conclusion is open to the objection that the class of possible grammars has not been defined. For example, Black & Wilensky note that Finite State Grammar cannot be essentially self-embedding, that is to say, if a symbol appears on the left hand side of the syntactic rule then it cannot also appear on the right hand side among other symbols. They argue that since stories can have goals embedded within goals FSGs cannot be adequate formal representations. However, a legitimate extension to FSGs exists, in the form of the Augmented (State) Transition Network (ATN) (Woods, 1970). It is of course arguable what counts as legitimate e.g. the ATN's degree of formality when compared with finite state automata (the machines which corresponds with FSG), as the ATN can even be seen as corresponding with transformational grammar (see Garnham 1985, p88). It is this difficulty which renders the argument from formal grammars inadequacy "double-edged", and calls for a more principled approach.), since the grammar equivalent of the ATN permits dropping down into sub-networks and recursion, thereby allowing essential self-embedding. Lyons (1970), on Chomsky, also mentions that phrase structure grammars with addition of a deletion rule become what Black & Wilensky call URSs (unrestricted rewrite systems). That is to say, there may be legitimate extensions of these formal grammars, but these extensions only postpone the application of a more principled criticism. The value of proceeding from the formal expression to the principled lies in the elucidation of the components of representation and their function. The detailed examination of the grammars helps one to understand what Black & Wilensky mean by their principled objection to the idea of the story grammar as capable of *fundamentally* discerning the meaning of a text.

When we turn to their arguments against the adequacy of the Phrase Structure Grammars (PSGs) we are approaching the real difficulty of *explaining* the story by the application of syntactic rules. In the case of PSGs, their first move is to point out that lines in stories may be interrupted by both contextually relevant and contextually irrelevant material, which interruptions could only be distinguished by a Context Sensitive Grammar, therefore excluding CFGs. To establish that examples of stories with 'irrelevant' lines[13] are not manufactured for the purpose, they point out that they abound in the literature: "The Canterbury Tales", "The Decameron" and "The Thousand and One Nights". So what about CSGs? The problem here is that Context Sensitive Grammars (see above) must have at least as many symbols on the left-hand side as on the right-hand side; so precluding a deletion rule which seems to be a prerequisite for story understanding. This latter conclusion is derived from an analysis of the following example:

(a) 1.John learned that his wife wanted a divorce. 2.John was upset. 3.He went out and got drunk.

and

(b) 1.John learned that his wife wanted a divorce. 2.John was overjoyed. 3.He went out and got drunk.

In (a) we have a story in which a possible transformation would be to omit component 2 and the story would still be acceptable. The latter attempts to address the problem of applying some kind of transformation rules to account for deletion and re-ordering of story components, but the difficulty is that the rules rest on the whether the omitted component, in this case, can be inferred from the surrounding components (a semantic

[13] I use quotes because it is clear that the stories are not disjointed.

issue). We can even imagine that the second component could be omitted from (b) if the wife's behaviour towards her husband had been cruel and the husband depicted as sympathetic.

They go on to analyse more examples from other writers' attempts to produce rules to aid comprehension, for example Rumelhart's, but their point is essentially the same. They conclude,

> "Once again, we are caught in a quandary. To apply syntactic grammar rules, we must first know what the events are that constitute the story. Finding these events may require the use of inference procedures to postulate implicit events. These inference procedures can determine that an event needs to be inferred based on knowledge about the possible semantic relationships among events. Thus it is only after these inference procedures have run, unguided by syntactic considerations, that we can hypothesize the actual propositions to which the syntactic story grammar rules are to apply. But by this time the inference procedure has already determined the semantic relationships among the events. This was the task we assumed would be aided by the story grammar rules. So the argument that a set of syntactic story grammar rules would help a reader to understand a story is circular. In order to determine the constituent structure of the story, we need to first have understood the story. But in that case, analyzing it into its constituent structure becomes unnecessary." (cf. also, but with respect to pure Speech Act theory, Shanon, 1993)

Their recommendation is that some sort of "content-oriented grammar" would seem to be the goal to aim at, and *they note that planning theory has been employed as a device in this context.*

Black & Wilensky's paper contains the essential ways in which language might be said to be represented more or less formally: from the simple grammars dealing with the elementary and abstract features of sentences to the more complex and more concrete representations at the story or discourse level. It has been criticised (Frisch & Perlis, 1981) for mistaken descriptions of those formal elementary grammars: phrase structure and finite state grammars, for example. However, most of the points Black & Wilensky make are cogent and it is generally accepted that these extensional grammars (PSGs, FSGs etc.) are not adequate to the task of representing story structure let alone that they can properly represent all and only legitimate sentences.

More important is their attitude to attempts at more sophisticated story grammars such as those of Rumelhart (1975). Frisch & Perlis (1981) claim that Black & Wilensky have overlooked the purpose of story grammars, and add that the grammar would be only one knowledge source among many which would contribute to the task of understanding. However, in the absence of a very precise explanation of just which knowledge sources do what, this position leaves the status of the story grammar in a very weak position, and Black & Wilensky's general conclusion of a principled nature appears solid.

It may be that their claims, as Garnham (1983) says, go beyond a simple rebuttal of the story grammarians and make further claims which are not substantiated. However, it is difficult to see how a grammar which is a set of rules for structuring language, normally at the sentence level, can be adequate for retrieving the meaning or import of the story. I do not perceive a fundamental distinction between Garnham's view, on the one hand, that there is nothing equivalent to a sentence/clause lexicon to which one can resort during a story's analysis and serving an analogous function to a dictionary (as there are an infinity of sentences but a finite quantity of words), and Black & Wilensky's view, on the other, that there is no way of knowing what rule to apply until the import of the components (implying also their inter-relationship) is grasped. That is to say, I believe both Garnham and Black & Wilensky are saying that there is a decidability problem here, and they are expressing this in different ways.

In fact, both expressions are partial. Garnham, by saying that the problem stems from the infinite class of propositions, is ignoring the possibility that the size of the class is irrelevant if the mechanism is (as it must be, I think) one resting partly on a salient selection, i.e., of the apprehension of a particular relationship between the lexical components. The Black & Wilensky argument, as Garnham points out, is concerned with the dominance of the semantic component as a determinant of how the story is structured, and these authors are ignoring the 'formal' requirements of completing an empty structure with appropriate values. In short, what separates Garnham and Black & Wilensky is a view of what counts as a legitimate argument: a properly linguistic one, as Garnham exposes, or one more broadly based, as argued for by Black & Wilensky.

Both may be inadequate from a given point of view, but it is interesting, finally, that both Garnham and Black & Wilensky concur in their general conclusion: Black & Wilensky write, "...the important issue for investigation is the nature of understanding, not grammaticality."; and Garnham concludes, "Once the role of other kinds of knowledge has been clarified, these grammars will be redundant." Further, Garnham claims, "The hierarchical structure of texts, for example, reflects the hierarchical structure of *plans and goals*"; and Black & Wilensky's final sentence is, " These content oriented approaches [including *planning knowledge*] are examining the important issues in story understanding, whereas the story grammars are not."

Sentence Level and Context

The comments in the last section apply at the discourse level. As mentioned above, it is accepted by followers of Chomsky, for example, that PSGs cannot deal with analysis at the sentence level, but they believe that PSGs can be supplemented by the addition of transformational rules (as mentioned by Lyons (1991), and referred to above). This matter cannot be dealt with 'head on', so to speak, since we do not have any examples of clear or formalised versions of transformational rules, and it is really an expression of a conviction on the part of the Chomskyans. In the absence of these rules which would permit and explain deletion and re-ordering, for example, it would be at least of practical importance to underline the essential part the context plays, i.e., what is meant or intended, or understood as meant or intended, by the players, e.g., the author/reader or the dialogue participants.

At the sentence level, Waltz (1982)[14] illustrates very well the lack of any necessary internal constraints on the meaning of sentences by means of the graphic alternatives of the sentence,

"I saw the man on the hill with the telescope" (Figure 1).

---

[14]Waltz may have based this series of images on that in Simon (1969), which he uses to illustrate, and cite as an example of syntactic ambiguity.

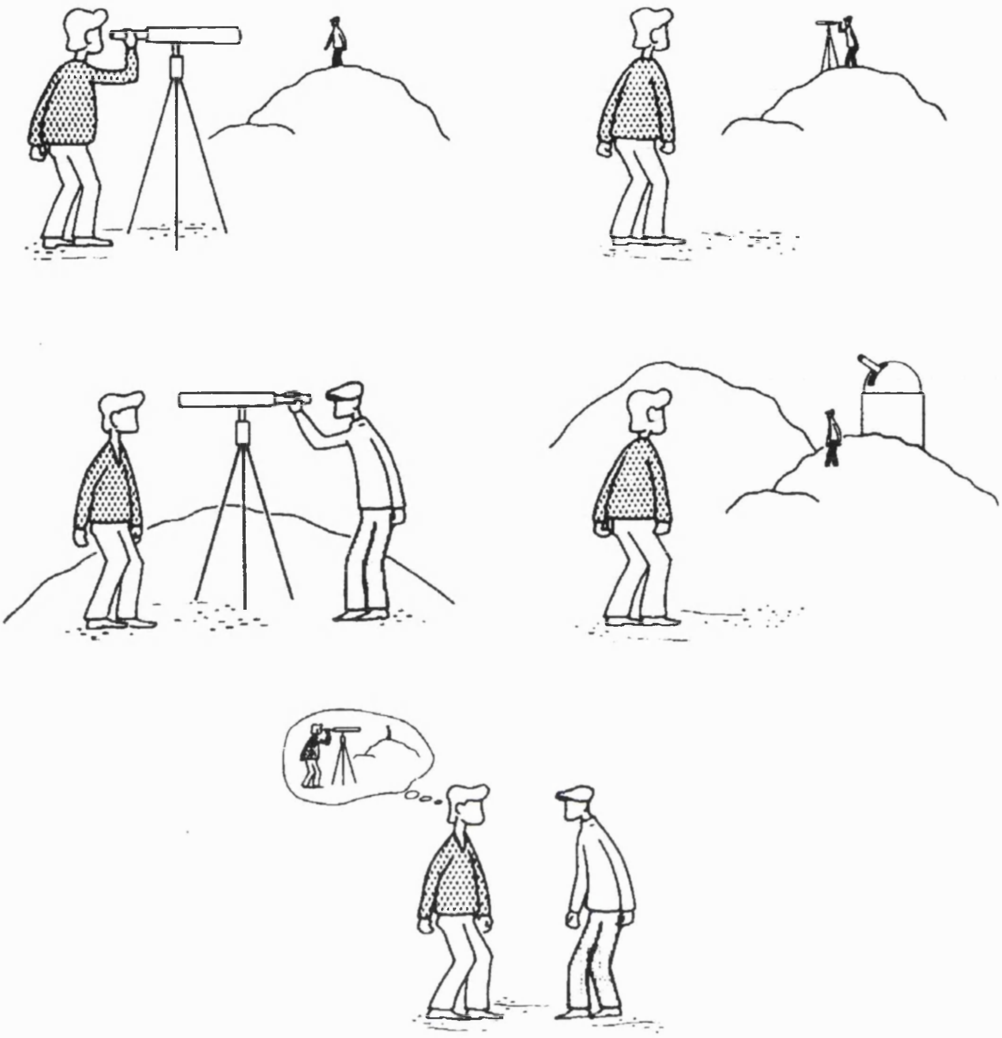
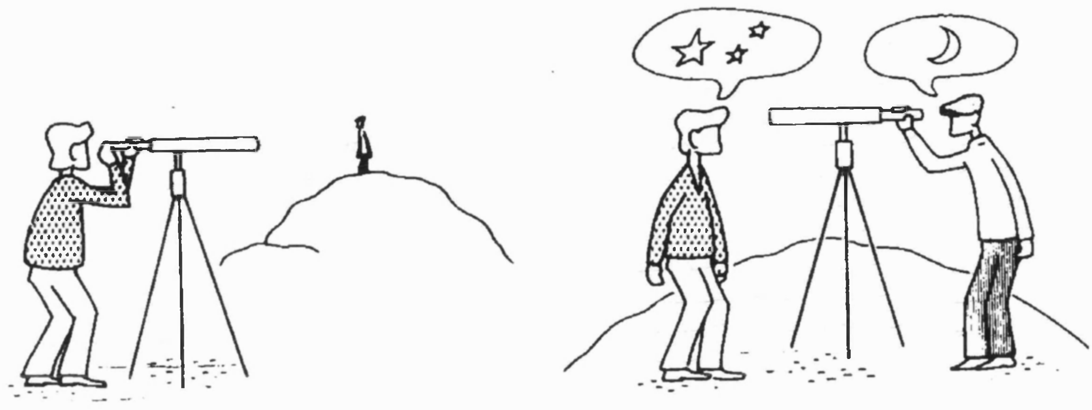


Figure 1: A Syntactically Ambiguous Sentence

There are clearly grammatical differences discernible in the different interpretations. Chomsky, for example, would not consider these differences relevant to his theory, since he is concerned with the deep grammar of the distinct interpretations: misunderstandings are a matter of *performance*. However, designers and evaluators have to concern themselves with just such ambiguities and their determinants. This is not meant as an argument against the Chomskyan position, but it does highlight the limited scope of such a position as a design resource, yet it is a position which aspires to a grasp of the essence of language; and as such, his perspective influences the conceptual background of those involved in language understanding *and* design.



"I cleaned the lens to get a better view" "We talked a while about astronomy"

Figure 2: Resolving the Ambiguous Sentence

The various levels are finally inextricable, and it is clear that the sentence in Figure 1 is, like the phrases or sentences in a story or dialogue, affected by the sentences which are uttered in the same context. Garnham (1983) also made the point that Black & Wilensky were wrong to claim that the structure of a given proposition in a story was determined by the propositions surrounding it. Instead, he asserts that the proposition's structure depends on its own semantic content. But how is this (its own semantic content) determined except via the semantic content of the surrounding propositions? Both the structure and the semantic contents of the sentence variously depicted in Figure 1 clearly depend on such utterances as exemplified by Figure 2.

The Word Level and Context

At the word level, comparable difficulties of language understanding arise. In Shanon (1993), there are many examples, but I shall take one: the word "paint". He points out that this word might be defined as "to cover a surface with paint", then goes on to cite an explosion in a paint factory and dipping a paintbrush into paint as instances of a reductio ad absurdum. What we seem, then, to be 'given' as language users are tokens which are employed *in meaningful ways*, and understood *relative to the participants' intentions, values and the specific contexts* in which they are uttered. However, although there may be more and more people recognising the crucial part context plays, there is still a tendency to interpret this context as really part of the syntax or structure of the utterance. For example, Barwise & Perry's theory (1983), which is reviewed by Winograd (1985), reveals this tendency and is, I think, convincingly criticised as a instrument for *understanding* natural language.

Winograd refers to this shifting to include the context as if it were a simple extension of the 'content' of an utterance or sentence, as "moving the fulcrum" of meaning, the last in a long line of moves from the assigning of truth values to propositions and its part in the determination of meaning, via the mechanism of assigning objects to sets in order to account for the meaning of, for example, 'Socrates is mortal' all the way to the creation of 'possible worlds' to accommodate intentional objects. It is therefore not enough to represent sentences and their context if, at the same time, these both become part of a 'grammar' which is to be applied to *return* the meaning. The difficulty, Winograd says, is not comparable with the idealised model of Newtonian physics (an implicit reference to Chomsky?). He prefers the analogy of economics in order to account for the inexactness discovered in the employment of 'bachelor', 'lemon', 'tiger', 'bird', 'water', 'on' etc. dealt with by Winograd, Hilary Putnam the philosopher, and Searle. If we imagine, Winograd suggests, attributing fixed values to commodities (in other words, treating values as primitive) instead of accounting for them in terms of supply and demand, then, although the "theory is not logically incoherent", it is unlikely to work since qualification will be continually required to satisfy each particular occasion. Likewise,

> "a theory that begins by taking properties and relations as primitive will need radical augmentation to deal with those cases where it is the linguistic activity that determines the meaning. Those cases are not the exception, but the norm."

Winograd concedes, however, that such a grammar, or linguistic theory, may well be put to good practical use, if its limitations are understood. Winograd's general point is that formal descriptions always involve the eventual moving of the 'semantic fulcrum', and that is why I commented above that the use of a critique of these formalisms is "double-edged". It is better to take the principled and deeper course, as Winograd does, or 'patches' will be continually added to solve problems as they arise. Winograd's argument highlights the boundary which separates the formal from the substantive or the purely syntactic from the semantic. It is this distinction which Black & Wilensky (1979) attempt to exemplify.

SUNDIAL's Design Approaches

In the next sections, I would like to expose what methods SUNDIAL offered as approaches to these questions of representation of dialogues. Following this brief exposition, the issue of how representation might be characterised, at a more abstract level, will be broached in the next chapter, with a view to discovering guides to the proper constraints which should be imposed on the kind of resources which a design/evaluation framework can employ.

Dialogue Design: 1st Approach

The project was led by Logica, Cambridge and the UK team was responsible for the design of dialogue management. Their initial approach was to design on the basis of abstraction of dialogue features found in corpora of dialogues. They collected 100 human/human and a 100 simulated human/machine dialogues(created by WOZ scenarios based on H-H dialogues). To fulfil the requirements they had set themselves for the dialogue design they abstracted a lexicon from the limited vocabulary of the corpora. Likewise, they abstracted syntax and semantics employed in the dialogues of the corpora. Finally, they provided the abstract description of the dialogue grammar, taking account of such properties as turn-taking and adjacency pairs (questions/answers, greetings etc.). Two related issues of particular interest were raised in the course of this process (one of which was mentioned in Chapter 1) by Norman Fraser (1991), a member of SUNDIAL'S U.K. team:

> "how best to make the transition from data to design is an open research question, one which deserves much more attention than it is presently receiving in the NLP community"

and;

> "clearly, a speech understanding system which modelled the exact behaviour found in the corpus and nothing else would not be useful."

Both quotations address the problem of specifying the design from some base of knowledge about the field in which the design takes place, and they were not thought to be satisfactorily answered by the method adopted in these first attempts at devising a description. In particular, they address the difficulty of moving from a position of knowledge, via some *principled* application of the knowledge to a specification which will instantiate a more general abstract description: 'principled' because it should bring some kind of guarantee that this will be so. It is clear that such a guarantee was not available in this case. Evaluation in this area of NLD is dogged by the same difficulties, that is to say, designers lack a clear expression of the essential features of a dialogue so that they can know whether they have achieved what they were aiming for. In the absence of which, some more empirical and craft-like approach has to be adopted, e.g., Fraser & Gilbert's (1991) Wizard of Oz or Dowell et al's (1998) ecological method [15].

Dialogue Design: 2nd Approach to Design

The next attempt at a characterisation of the features of a natural language dialogue involved calling on resources from work done in philosophy/linguistics and sociolinguistics: respectively, Grice's Cooperative Principle (CP), and Conversational Analysis (CA). This approach was more evident in the Sundial team's thoughts on evaluation. Two of the metrics which they suggested during the evaluation stage are in part derived from these resources: they are 'Contextual Appropriateness' and 'Turn Correction Ratio' (The terms are more or less self-explanatory; and, respectively, they are based on Grice and CA theory). Contextual Appropriateness required a team of experts to agree on the degrees of appropriateness of dialogue turns; and assessing the Turn Correction Ratio of user and system needed some judgement as to what counted as a high or a low ratio relative to some measure of naturalness or effectiveness. Both therefore depended on the expert's implicit knowledge, the terms of Gricean or CA theory only acting as a focus or common language. Could something more of an 'objective' and explicit nature be extracted from these resources?

Available Resources

Grice's work (1967), like Searle's Speech Act theory (1969), were both part of a general move away from a universal application of the logico-deductive view of language. Grice tried to establish what the general rules were which govern conversation, and Searle and Austin (1962) before him, examined the conditions

---

[15] See Chapter 1.

which underlie the correct or meaningful use of several classes of utterances: promising, for example, falling into one of them (the class of commissives). In some ways, the two are complementary (indeed, there are overlaps – indirect speech acts and implicatures[16]). Speech Act theory might be seen as providing the conditions which underpinned the consistent and systematic use of language by the user: a prerequisite for coherence across the speech interaction of dialogue participants, which was the area analysed by Grice. I shall come back to Speech Act theory in the constructive assessment of its place in an approach to solving the technical problem; and later I shall look more closely at the question of amalgamating theories like those of Searle and Grice.

To examine the plausibility of using such resources for the design and evaluation of the design of dialogue, we must examine what such theories claim. I have already made some reference to CA above, and I shall point to some of their own explicit statements, in the next chapter, which will tend to support my contention that CA resources, like unelaborated forms of SA theory or Gricean theory of meaning, are inherently inappropriate, if they are treated as 'design-ready' knowledge.

Example of Grice

In place of the unreflective acceptance of conversation as made up of propositions, which were meaningful by virtue of their truth value, and were generated and interpreted by the participants according to some vaguely understood information processing model, Grice tried to show that what was of fundamental importance was what the participants were aiming for. The relationship with truth would be subtler, and the primary requirement was that the participants were working together: that they followed the CP:

> "Make your conversational contribution such as is required at the stages at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged." Grice (1975)

---

[16]The relationship between the two becomes an important issue in Chapter 6.

In order to maintain adherence to this CP, he postulated a number of maxims which the conversational participants would have to follow: the maxims of Quantity, Quality, Relation and Manner. Respectively, they posit as normal that the quantity of detail contributed is appropriate; that what is said is known to be true by the speaker and/or that he has evidence for what he is saying; that the remark is relevant at that stage of the conversation and, finally; that it is expressed clearly. Because it is clear that when people joke or are sarcastic they do not follow these maxims, Grice argued that the CP was operating at a higher level and that therefore what he called "flouting" of the maxims could also be meaningful. For example,

Flagseller : Would you like to buy a flag for the Royal National Lifeboat Institution?

Passer-by : No, thanks. I always spend my holidays with my sister in Birmingham.[17]

This could count as a flouting of the maxim of relation. Under normal circumstances the conversation would not break down, and the remark would be understood as a humorous one. This intentional move is embodied in what Grice calls an implicature. The speaker is implying something by breaking a maxim, but it is not a logical implication, and if it is meaningful it is held to be so because it can be understood as an expression of being conversationally cooperative.

The problem, in part, is that these maxims are guides, not rules in any strict sense: that they are defeasible by their nature and subject to the overarching principle of cooperation. There are different views as to the connotation of 'cooperation', but it appears to be most appropriate for Grice's purposes if it is understood in a most general sense; and it is clear that conversation is not simplistically cooperative. But, without rules which are specific in some sense, and, therefore, not ultimately

[17]This real fragment of dialogue is attested to by Sperber & Wilson (1982)

defeasible it is difficult to see how its connotation can be other than *co*-operative, i.e., just the way conversation is – two halves of a dialogue *operating with* one another. It is not that we do not recognise something new in Grice's account. It is simply that when we apply the description to a conversation we only bring our intuitions of where and how it is cooperative to bear on any explanation. In such a post hoc account we presuppose certain conversational mechanics rather than discover them. It is difficult then to see how, as they stand, they could be of scientific or design use.

What Kind of Resources?

Rules which govern the make-up of language have been central to this chapter. I have drawn on arguments (both formal, grammar; and informal, maxims) which undermined those rules as the basis for understanding text or dialogue. In order to exploit resources such as Grice's principle and maxims in the design or evaluation of systems employing speech interfaces we must first understand what kind of rules they are; and we must understand any limitations which are inherent. And to see what kind of rules are required in design and evaluation it is important to consider what sort of knowledge these rules support. In the next chapter, I want to look at the general characterisation of representation, since, at least intuitively, Fraser's point about the need for a better understanding of the move from data to design requires a representation of that data and a representation of the artefact, i.e., a design specification. At present there is an ill-understood process taking place between those two representations. In the context of a general view of representation it might be easier to explain the criteria for discriminating those knowledge resources which support design in terms of the kind of rules which they exhibit.

Conclusions

With the benefit of hindsight, it could be said that designers of language systems lacked the conceptual apparatus to ameliorate NL design in any significant way, either because the conceptual work had not been done or the linguists and designers had not taken it up. As an introduction to my own suggestions about how the issue of using conceptual resources should be tackled I, therefore, engaged in a brief critical survey of current ideas of what sort of entity or process language might be. I adopted Dummett's view that it consists of two aspects: the representative and the active, which some thinkers consider as opposing positions. I, therefore, took an example of an argument against the formalist stance as in itself able to deliver the means or criteria for understanding; and elaborate other short arguments against this attitude. Employing it as an adequate representation of natural language for design would provide no substantive step towards generating nor comprehending natural language.

I next consider what kind of conceptual models are adopted by SUNDIAL and reject them in the form they are proposed. In order to see how these resources can be employed we must first understand why they are inadequate and yet appear to say something informative about language. To achieve this aim, we need to recruit some criteria which will allow us to articulate what the resources offer and perhaps, further, indicate what is required to remedy the inadequacy. And to fulfil this goal, we first require some understanding of just what range of relevant representations are needed for the purposes under review. Both of these tasks will be undertaken in the next chapter.

CHAPTER 3

The aim now is to refine an abstract notion of language representation as described in the last chapter, and with the help of concrete examples to orient the search for a representation with the desired properties. I adopt a high-level description of a scale of kinds of representation, and then pose the question: what kind of criteria would permit the integration of the kinds of representation and yet allow us to distinguish the extremes of that gamut? The criteria derive from Searle's philosophy of language and define speech acts. They can, however, be adapted for other uses; even with the speech act itself, as they will be later.

The rest of the chapter is devoted to clarifying these rules which will act as the criteria, and justifying their employment for this purpose. Some time is expended in the explanation and illustration of these rules as they play an important part in the thesis argument.

Rules and Representations

Introduction

The job of the last chapter was to draw attention to the relationship between representation and understanding as it is exemplified, at different levels of granularity, in more or less structured kinds of representation. The conclusion has been drawn that understanding is not captured in any such simple mapping as instanced by those representations considered, but I have considered only language as a candidate for representation. I want, in this chapter to broaden (and deepen) the analysis to include representation more generally, in order to understand it better and in a more abstract manner than that of the last chapter.

My aim is to examine the possibility of employing a shibboleth (in the form of a very generally applicable distinction) to enable us to distinguish what sort of resources, briefly described in the last chapter, can be exploited for the purposes of design from those which are inappropriate – at least in the form in which they are presented. My principal interest here is to establish a simple extension of a rule distinction made by Searle as an essential part of his theory of speech acts. Between the introduction of the distinction and its ultimate use there will be its intermediate role as I examine Speech Act (SA) theory supplemented by planning theory (Plan-based Speech Act theory (P-BSA)): how a descriptive and quasi-scientific account of language might serve as the basis for a framework for design and evaluation. I shall underline, on analogy with science and engineering, how this intimate relationship between different kinds of rules can explain how communication may take place. It is an important illustration of the power of the distinction that it will also allow us to reconcile some of the claims of ethnography, in the form of Conversational Analysis (CA), with those of a framework for the cognitive engineering of Natural Language Dialogue (NLD).

Representation

In its broadest sense a representation is a transform of something perceived or apprehended, a re-presentation of something in another form or mode, used for a purpose. Patrick Hayes (1985) lists, "logical calculi, some programming languages,..., map-making conventions, circuit diagrams etc." as examples of formal schemes of representation because they bring with them "a definite notion of *well-formedness*" (my italics). He acknowledges other forms of representations which he claims lack this quality (and, thus, cannot be *ill-formed*), but are still "ways which humans have of conveying meaning", such as: "...drawings, photographs, poems, conversational English, musical performances TV pictures etc...". More generally, he distinguishes between representations which can be stored and used by a program and others of the informal kind "requiring the deployment of knowledge for their successful interpretation". Hayes, however, stresses that even the formal sorts of representation – 'schemes' – require interpretation by the application of semantic theory. In this sense, he continues the theme of the last chapter: that grammar, which is a 'scheme' in Hayes' terms still requires semantic input for its completion as a representation.

His paper is, in addition, a plea for the recognition of semantic interpretation as relative to a purpose. He is not content to see his partition of representation into 'scheme' and others compromised by the introduction of fuzzy concepts or logic. His intuition is that meanings are not in themselves imprecise. It is rather the way in which they are used. On the other hand, he is not content with formal 'languages' which employ categories such as 'substance' and 'attribute' without more specific explanation of what these entities are. Nor is he prepared to impose an extensional semantics on natural language, but with respect to those who deny the usefulness of an extensional, he recognises that it "is a perfectly respectable philosophical position: but I submit that it is bad engineering." In other words, he is agnostic in his world-view, holding that his tools should be those which are going to progress his technology.

Hayes' (1985) argument is one for which I have a lot of sympathy. However, I believe a better partition can be made of representational types, which will allow us to accommodate the informal category (including language which Hayes sees as marginal) and avoid the following example – an implication of his position: he writes, "(T)here is no such thing as an ill-formed photograph". It may be, though he does not qualify this assertion, that it is not as bald as statement as it looks, but it deserves stronger support than he gives it. The difference, then, between what Hayes calls 'schemes' and other informal kinds of representation is supposed to be self-evident, but, with this assertion above, he appears to take up a problematic stance. Is there not, perhaps, another way of considering representation which would circumvent this difficulty, consistent with the conclusion of last chapter; and still allowing us to progress the argument.

An Alternative Representational Schema

One might characterise the different forms of representations by envisaging them as consisting of a range of mappings from, on the one hand, the classificatory or phenomenological to, on the other, the generative or predictive. That is to say, the simplest representation, qua representation, is something like a Linnaean or Aristotelian classification which tells you nothing of the mechanisms or connections underlying the order perceived, and the richest which reveals an underlying mechanism and which resembles scientific theory. The one may be regarded as the first methodological step to the other. The first, at its simplest, is the recognition of similarities. It is tentative and revisable and, by its nature, non-predictive and non-explanatory. The second is essentially predictive, or generative. See Figure 3.

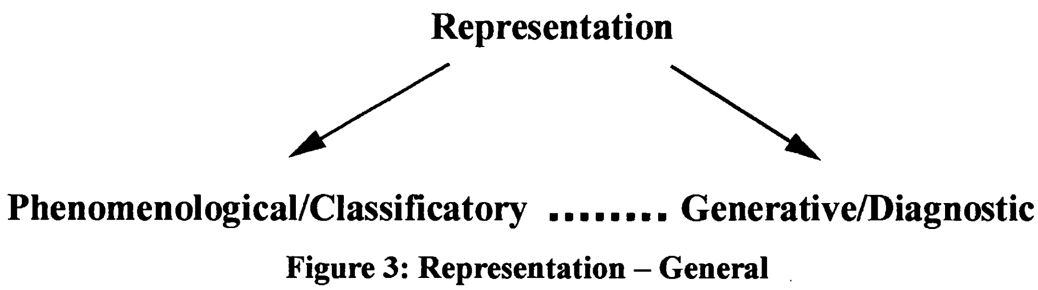


Figure 3: Representation – General

Fraser's (1991) appeal to the research community to address the question of "How best to make the transition from data to design..." could be viewed as a question about the required representation. He approaches the problem of determining the characteristics of NLD in an 'information provision' application such as SUNDIAL, by identifying and tagging the vocabulary, syntax and dialogue grammar – a process of classification, and provides the rejoinder, "Clearly, a speech understanding system which modelled the exact behaviour found in the corpus and nothing else would not be useful." In other words, although a form of representation has been achieved it tells you nothing new when implemented in vivo. This, then, is not an example of a generative model, but of one which is only classificatory or phenomenological.

Hayes' assertion that "(T)here is no such thing as an ill-formed photograph" appears to assume the requirements of a representation of the generative/predictive kind while finding that it falls short of them. But what about one in which there is figure and ground? For example, the ground is a scene and the figure takes up 80% of the picture, and the figure is out of focus. Now, two questions might arise: Was it done intentionally? and, Does it work? The questions appear to arise because of the possibility, in the first case of an error of operation, or, in the second, of an error of judgement. Here, surely, are two sorts of ill-formedness. This, then, also looks like an example of the cut-off point on the continuum separating the merely phenomenological or descriptive from the generative/predictive or prescriptive, in Figure 3. It is the point when the necessity of a semantic theory, relating the representans to the representandum, begins to intrude. Because Hayes has admitted that all 'schemes' need interpretation by the application of a semantic theory he cannot allow himself a notion of well-formedness which is totally abstract. If it is absolutely formal it is absolutely empty and is not, in his terms, representation at all[18].

Is there, therefore, some way in which we might characterise the two broad categories of representation without running the risk of this dilemma? If possible we want to recruit a concept of representation which will be apt for the purposes in hand: representation of language and representation for design.

---

[18] As Haugeland (1988) says, "A formal system as such is completely self-contained and, viewed in that way, is just a meaningless game".

We should be looking to define the required form of representation in the most general terms, and since all forms of representation conform to rules it might be helpful to examine rules of a very general kind to arrive at some representational schema.

Searle's Rules

Searle (1969), referring to Rawls' (1955) rule distinction when he introduces his own, defines the difference in the following way,

> "Regulative rules regulate a pre-existing activity, an activity whose existence is logically independent of the existence of the rules. Constitutive rules constitute (and also regulate) an activity the existence of which is logically dependent on the rules."

Examples of constitutive rules are game rules, without which the game cannot be said to be played: for example, that the bishop, in chess, can move and take only diagonally. By contrast, a regulative rule might be: Keep your legs together and bent, when lifting a heavy weight. Even if you do not follow the rule you can still be described as performing the activity of lifting a heavy weight, by contrast with the constitutive rule of the bishop's move. By breaking the rule the move does not count as such. Further, as Searle defines constitutive rules, it does not make sense to ask the question: When is the appropriate time, in chess, to apply the rule 'Move the bishop diagonally?'

Chess, however, is an interesting case. And, in this context, constitutive rules are not only of this specific type, with clear conditions of application. In the game of chess one meets rules such as, 'In the opening, take control of the centre.' What is the status of such a rule? There are, of course, explicit and expected ways of taking such control of the centre, for example, pushing the centre pawns to the 4th rank There are also less explicit and more unpredictable ways of doing it: by, for example, fianchettoing the bishops so that they gain control from a distance. There may be yet other ways which have not been thought of as examples of such control: more complex positions which have not been described to date. This general form of rule may be constitutive *but contextually determined.* We cannot exclude a move as non-centre-controlling without waiting to see how the game develops and/or without understanding in what overall state of the board the move takes place. These rules are constituted by the 'external' context, not by the 'internal' connotation. The rules, which are constitutive may not be 'ground' rules only.

The point here is that, because the game of chess though complex is finite in its ramifications, there are potentially rules which are tightly linked to the end or goal of chess – checkmate, i.e., they are not logically independent of their application. Two things should be noted: firstly, how rules are represented or categorised is related to the setting in which, and purpose to which, they might be put; and, secondly, rules may have to be seen in a different light in finite and fixed systems by contrast with open-ended ones, but this relativity is not indeterminate.

The use of "regulate" in the parenthesis within Searle's quote is indicative of an underlying equivocation of Searle's which may be viewed as reflecting either, the bearing which the setting has on the rule's application, and which we shall understand better after an examination of the Plan-based Speech Act theory's limitations; or, the implied relationship between a semantic and a pragmatic theory with the implications that has for the relationship with the science/technology distinction, addressed later. However, it is worth dealing with the equivocation in relation to Searle's surmised position. Searle defines a constitutive rule as one which is not logically independent of the activity it qualifies; he gives examples of rules of games which are clearly at least partly definitional of the activity of the game. He appears at times to claim that the game is constituted of the rules in the same way that a chess piece is constituted by the rule of how it moves and takes opponents' pieces. Now it is clear that the game transcends the rules in the way that the chess piece does not. The game can be changed and yet remain the same game, perhaps by changing the rules of pieces, by defining new ones into existence and old ones out of existence. There is a residual, but not really satisfactory, sense in which even a piece would remain, say, the bishop and be allowed to move in some modified fashion. The game could retain its character with changed rules, but not, say, the bishop: the piece, of course, would look the same. I think one can fairly allow that relative to Searle's view of the rules as defining and, in a sense, governing the game of chess, he maintained an agnostic attitude to the meaningfulness of the interest or desire to play the game of chess[19], and its bearing on the force of the rule distinction. In other words, he was not asserting, as Schwyzer (1969) was implying, that chess was only the sum of its rules.

As a contrast, that someone should lift a heavy weight with their legs bent, or not, really makes no difference to the description of such an activity as weight-lifting. At least, in this sense of logical independence Searle is, I think, correct; and, for the purposes of the thesis argument, this stark distinction, expressed in this fashion, is enough. It will become clear how these distinctions play their part in the larger scheme of things as the thesis argument develops. A propos, Shimanoff (1980), raises the question of the legitimacy of having a constitutive rule which also regulates. After demonstrating scepticism about the whole project of maintaining such a distinction as Searle does, she writes, "Apparently, even Searle has trouble keeping them separate, because after carefully distinguishing constitutive and regulative rules, he writes that constitutive rules constitute *and* regulate an activity" (my italics). Of course, a distinction can be made by keeping one term distinct: regulative rules do not constitute, and this stricture may be all that Searle needs for his purposes.

Rawls' Understanding of Rules

Rawls (1955), who I have noted draws a similar rule distinction in the context of moral philosophy (with particular reference to criticisms of utilitarianism), when alluding to the 'practice conception' of rules, writes that they "are logically prior to particular cases", or "those involved in a practice recognise the rules as defining it". This then is Rawls' version of constitutive rules (although I, think, he nowhere refers to them as such); and, by contrast with this 'practice conception', he defines what he calls the 'summary view' of rules. The 'summary view' is his equivalent of Searle's regulative rule[20]. For Rawls, it is the recognition of these different conceptions of rules which allows him to resolve different and conflicting attitudes to, for example,

---

[19] This problem of the relationship between constitutive and regulative force raises its head again in the differences between Searle and Grice.

[20]The term means that Rawls has in mind a moral rule arrived at by a kind of induction. Searle pays no attention to the origin of regulative rules so we can ignore this slant in Rawl's analysis. The important thing is the relative status of the rules, and the analytic aspect of Rawls' practice conception rules and Searle's constitutive rules.

punishment and its justification. It is important to note, however, that the distinction is powerful enough to fill the requirements of Rawls' forceful argument. His particular employment of the types of rule is, as he describes at the beginning of the paper, in order to "to show the importance of the distinction between justifying a practice and justifying a particular action falling under it".

Primacy of Constitutive or Regulative Rules?

There is a difference in the way in which Rawls employs his concepts of rules, and it is worth going into a little detail. His categorisation appears to be the converse of the one I have adopted from Searle. Rawls' 'summary view' is what we would naturally call the principle, which has the property of generality, and which is, in this case, adduced or inducted over instances. On the face of it, then, this notion seems to correspond more to a law of nature, which I shall characterise later as constitutive; but for Rawls the 'summary view' is regulative. His 'practice conception', on the other hand, concerns the particular case, and is constitutive, i.e., there are certain practices which follow 'necessarily' some event: the punishment fits the crime. I have been referring to practices as regulative, by contrast, because they justify and support the law or principle, where it is the law or principle which is constitutive. The apparent inversion of the status of the rules is due to their application. The application of the rules is governed by the activity predicated. Ethics is prescriptive and as a consequence its driving force is regulative. In addition, axiological rules are not generally supposed to reflect reality (except in the case of Platonic forms) as science does. Science is descriptive, and so its orientation is constitutive; and so, one might say is the philosophy of language or linguistics.

I think it is important when comparing Rawls' and Searle's conceptions of rules to bear in mind that once Searle has made the distinction he casts aside the 'regulative' having no further purpose for it. It is true that Rawls does not develop it much further[21], he acknowledge that they both play a role in ethical reasoning; and, indeed, he uses their co-existence as the means of reconciling conflicting viewpoints arising out of utilitarianism. He also accepts that "there will be many border-line cases about

[21] Although he notes at the end of the paper that there is room for such an analysis.

which it will be difficult, if not impossible, to decide which should be made if one were considering other questions". Like Rawls, I intend to take both as significant, and integrate them in the NLD *and* the foundational framework.

The fact is that we feel obliged to keep our promises for reasons which transcend the constitutive definition; to do with, for example, how others in society may behave as a consequence of, or with regard to, others' rights: "This rule we may think of as the rule of promising; *it may be taken as representing the practice as a whole. It is not itself a moral principle* but a constitutive convention" (my italics) (Rawls, 1971). This quotation expresses the difference between justifying a rule and justifying an instance which falls under that rule (see Rawls, 1955). Searle nowhere addresses this ambiguous role played by otherwise constitutive terms such as 'promise' (which is at the heart of this thesis), but its resolution is the key to both the role speech act theory, via the plan-based version, plays as a representational resource in the process in evaluation and design of NLD, and the role science, generally, plays in its relationship with engineering, and vice versa, as I shall try to show.

Representation and Constitutive Knowledge

I am hoping to derive something more than an application of Searle's distinction. I want to generalise the ideas of the rules to that of constitutive and regulative knowledge and corresponding representations. This generalisation is simply the stipulation that knowledge of a part of the world which depends primarily on constitutive rules is to be understood as constitutive knowledge, and its representation is correspondingly constitutive. In the course of the thesis, I have (and will again) refer to knowledge being constitutive relative to some projected purpose. Later, as I have hinted I refer to scientific knowledge as essentially constitutive. Yet it is clear that scientific knowledge is not vacuous. I want, therefore, to illustrate what I mean when I call some piece of knowledge constitutive.

I have, in the last chapter, said of Grice's Cooperative Principle (CP) that it seems, by its nature, to be uncircumscribable in a way which might allow a scientist or a designer to proceed and apply it. I want to *re-express* this account of conversation as constitutive knowledge, since it means simply the juxtaposition of individuals' activity (the two sides of a conversation). It does not imply that the conversation's protagonists are being helpful, for example; and, as pointed out before, reducing the idea of cooperation to the mechanics of the maxims' operation may only make sense post hoc. In the last chapter, in the dialogue between flag-seller and customer the maxims are flouted meaningfully, and this notion of cooperation is acceptable as an explanation partly because the exchange is friendly: it seems plausible that a principle of cooperation is at work that gives direction to the dialogue. What, however, of a conversation in which someone unsuccessfully attempts to borrow money from another unsympathetic to the cause of the putative debtor, or a dialogue in which the antagonists argue over rights. Where is the cooperation? Surely this is only an expression of *co-operation*, i.e., of some kind of joint activity. The *cooperation* of the first example is now at one remove. It appears more correct to say that we understand what the cooperation means only when we have grasped the point or direction of the dialogue or conversation. This means that relative to the purpose of design the characterisation of a dialogue or conversation offers nothing regulative. Paradoxically, it might be thought, it remains meaningful – as part of philosophy; and as part of a move in pragmatics it may represent progress (see Chapter 7).

Let me, therefore, take an instructive example from literature to show how even an apparently absurd example can be constitutive but not entirely empty. In "Le Malade Imaginaire" by Moliere, the learned doctor is asked to explain how opium induces sleep. He answers that it does so because it contains a 'virtus dormitiva' (a sleep-producing power). This is a paradigmatic instance of a piece of constitutive knowledge. It seems utterly vacuous. In fact, a little reflection shows that it too, though constitutive, is not without some import. Imagine the conceptual framework extant, say, three thousand years ago. Even the idea of physical causality was not fully formed. Agency was primarily quasi-personal. Gods controlled events. Things were intermediate. This absurd answer by the learned doctor still carries information which has not always been known. It carries with it assumptions of a relatively modern idea of causality. It is not absolutely meaningless or vacuous – only relatively so. It is a good example, therefore, of what might be meant by this idea of constitutiveness relative to a project. Such a principle as Grice's may, nevertheless, be helpful. It may be potentially regulative, but some systematic relationship has to be set up between such an account and its exploitation as practical knowledge. Treated as regulative knowledge Grice's maxims and his CP are, then, vacuous.

The maxims are, therefore, as Suchman (1987) describes them, applicable post hoc, and it is begging the question to arrive at the meaning after applying the rule. The CP is constitutive because there is no further and *final* rule by which the maxims shall be flouted; and it is *primarily* constitutive because any regulative component is tacit. It is not, however, meaningless as many philosophers who admire Grice would affirm. But it *could* be said to be meaningless to the NLD designer.

Revised Representation

Is there, therefore, some way in which we might characterise the two broad categories of representation, based on such general rules, and then relate them to the business of design and evaluation? I introduced this section with a reference to a particular use of rules by Searle. The definition which he gives in his book "Speech Acts" contains an equivocation that expresses a dilemma, which either vitiates the basis for the distinction or points to a nuance which is important but which he does not, in fact, address. Since these rules play an essential role in his view of what speech acts are, I believe it is the second horn of the dilemma which must be exposed. The equivocation, I believe, rests on the relativity of the distinction between the merely descriptive and the diagnostic, which I have alluded to above. The concept of the constitutive readily stands in for the limiting notion of the descriptive and that of the regulative, the limiting notion of the diagnostic (or prescriptive) (see Figure 4).

That is to say, in terms which Searle uses elsewhere in "Speech Acts" (Searle, 1969), the 'constitutive rule' for x states what 'counts as' x (This is classification at its simplest and least reflective), and the 'regulative rule' for x, what one must do with x (for which there are usually reasons). The first is conventional, arbitrary, or 'just the way things are' and the second is 'the way things should be', absolutely or relatively. Since it is my contention, on analogy with arguments in the field of ethics that theoretical reason arises from practical reason, and that arguing from 'is' to 'ought' is natural and, ultimately, unproblematic, might we not express this transition in terms of the move from the constitutive to the regulative? This is not as straightforward as it sounds, but will be dealt with later.

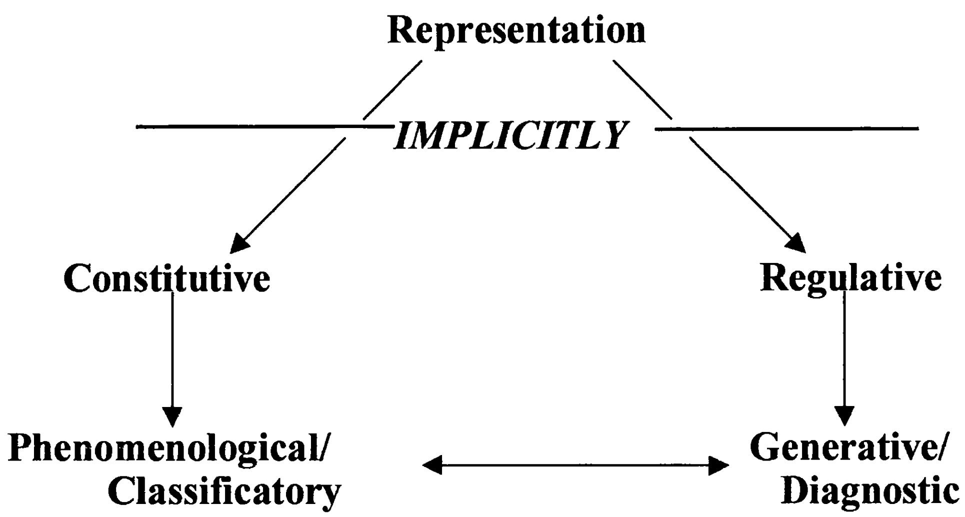


Figure 4: Representation Incorporating Constitutive & Regulative

The 'is', the purely descriptive, is the constitutive, and acts as the limiting point, that is, the abstract and the theoretical (the pure as opposed to the applied). Although it may be the starting point for formal reasoning, it is the finishing point for knowledge which starts out as unconscious and practical. It is this inversion of the usual roles of theory and practice (e.g., ethics or design) which resolves (by making nugatory) the problem of deriving design from declarative knowledge (or 'ought ' from 'is'). The formalisms involved in the process of design are thus made up from the re-constitution of the declarative with the procedural; the theoretical with the practical; the constitutive with the regulative. The important point is that this synthesis is not the artificial introduction of the normative into some idealised world of pure description, but simply the re-application of never-completely-separated regulative component – back where it belongs, so to speak. The 'constitutive' end of the representative continuum stands both for classification and formality. The expression is tabular or formulaic, and adopts a rigidity. It is this rigidity which is tested to destruction and is witness to the theory's falsifiability. This is the logical end of the hypothetico-deductive system. It is at the other end of the spectrum that evidence is proffered for hypotheses, and speculation is alive. However, the constitutive end cannot always be treated and developed in this way. Many philosophical principles or observations are of this kind, and consequently disciplines which inherit such notions, such as linguistic and cognitive science, have to be approached cautiously when recruited as epistemic resources for either science or engineering/design.

Science tells us how things are, not how they should be, and although it might seem as though, if you knew how things were, then you could easily move from there to how things should be using the laws of nature, rather in the way you might use a map, this is partly because it has been forgotten that a map is a tool for the purpose of getting from A to B. This analogy, therefore, begs the question which we are trying to answer. What is it about a map which makes it useful and what is the map's relationship with our knowledge of the world? The comparison might be expressed in the following way. With scientific knowledge we know where we are, but discovering how to get to another place requires establishing a route.

What might make linguistic theory useful to design and evaluation knowledge, or what is the connection between this latter knowledge and our general knowledge about language? In the absence of substantive representation, i.e., one which *explicitly* captures the regulative rules governing conversation, what could we suppose would be a promising model to adopt in order to discern a route to NLD design from what we might be said to know of language? I have pointed out that some who are critical of a 'formal' (e.g., syntactic) approach have argued for a recognition of the importance of motivation and goal-oriented analysis, namely, some version of the application of planning theory.

SA theory is an account of language, and *Plan-based Speech Act (P-BSA) theory* (Cohen & Perrault, 1979, which I shall deal with in the next two chapters) *is, I would claim, the mediator between the abstract SA theory and the design activity.* It is the introduction of plans which allows the move back from the theoretical abyss to the sunny uplands of design. *It is, therefore, an example of the move from the constitutive to the regulative at some generic level.* The move is systematic. That is to say, the claim would be that the link between the conventional/arbitrary/artificial speech acts, which function as tokens, is that they are carried on the back of the formal vehicle of planning theory, and plans are entities in the world – the *realisation* of intentions, to be distinguished from their *implementation*. The vehicle is, however, not necessarily 'powered'. That is why Cohen & Perrault (1979) think of their P-BSA theory as a 'competence' not a 'performance' theory[22]. We have, however, a way of characterising specific exchanges in terms of SAs embedded in a context, once we know when agents are likely to want to do certain things. These wants and their timing will derive from the cognitive and environmental constraints. I shall address these matters in the next chapters. The P-BSA theory is, nevertheless, the bridge which takes us from the simply descriptive, provided by linguistic science, to the prescriptive activity presupposed by design.

Application of the Rule Distinction

Searle accepts that SA theory essentially concerns constitutive rules, although, as noted above, he equivocates on his definition of 'constitutive'.[23] So I shall not dwell on a separate criticism of rules of speech acts as claiming regulative status. In any case, as far as I can see, they are not drawn on in any explicit way by the SUNDIAL team. If they were, it seems that the same comments would apply, but with more authority.

I have also left out an exposition of the other principal resource, CA. I believe it falls to the same criticism. Indeed, it does so more obviously since its essence lies in the emphasis on local management of the emergent conversation. The rules in turn-taking, like those of Grice, can be contravened meaningfully: the rule-following or contravention being constituted by the apprehension of the meaning, i.e., post hoc. I think it is quite possible that the practitioners of CA would not demur if their rules were described as constitutive.[24] This is what I believe their claims about the local

---

[22] I return to this theme in Chapter 6. The terms 'competence' and 'performance' are Chomsky's and might be broadly described as epithets of theories which concern themselves with the essential rules or structures of a given field of study, language, (competence); or the behaviour supported by these structures when operating in a real context with all its contingencies ('performance'). Chomsky further attributes reality to the rules or structures of his 'competence' theory. Cohen & Perrault use the term analogously, but only for the model of the dialogue.

[23] Another way of interpreting this equivocation is to see it as the admission of the tacit use of (regulative) knowledge. Searle does not express an opinion on this matter.

[24] Sacks, Schegloff and Jefferson, in Schenkein (1978), refer to "the character and organization of the rules" (of turn-taking) "that *constitute* the system as a local management system..." (my italics); and in Footnote 46, which excludes their use as regulative rules, they write, "Thus, while an addressed question requires an answer from

management of conversation amount to: one cannot derive generally applicable rules for conversations from the observation of turn-taking (which would be helpful in design) because these rules are integral to the conversation in question, i.e., at least partly constitutive of that conversation.

In terms of my extended diagram, Figure 4, and in conjunction with the above application of Searle's constitutive/regulative rules, the argument is that, understood strictly, the formal approach is vacuous, and can only be of substantive use if some other knowledge is brought to bear in conjunction with the formal rules. Likewise, with respect to story grammars and the like, it is a short step from the expression of 'formal' rules, such as the rewrite rules discussed in Chapter 2, to the recognition that they are good candidates for the category of constitutive rules, where the left-hand side is constituted solely of what is decomposed on the right-hand side. The important point which arises for the story grammarians, or designers who exploit SA theory, CA theory, Gricean theory is whether they would acknowledge the danger of leaving unspecified the tacit knowledge which they inevitably bring to their design task.

It important to be aware of this use of tacit knowledge as part of self-conscious design, since getting things done in a systematic way and being able to communicate what we are getting done to, say, other members of a design team is an all-important presupposition for successful specification of the design task and for its eventual evaluation. What is required is the re-establishment of design continuity between the interface work and the application work, with a view to achieving the necessary level of effectiveness. If language in its regulative aspect is activity then the 'mechanics' of this activity should be grasped to assess its effectiveness and continuity with the effectiveness of the application.

---

the addressed party, it is the turn-taking system, and *not syntactic or semantic features* of the 'question', that requires the answer to come next" (my italics).

Conclusions

This chapter has been devoted to taking a stand on how representation, in a general sense, can be characterised; and then proposing rules the application of which might serve as criteria for distinguishing instances of representation in the terms characterised. The rules are not uncontroversial, so some space has been devoted to examining what their status is; and both the rules and the corresponding representational schema proffered will be employed throughout the thesis to satisfy the more fundamental goals of the argument. It is important that the general distinction, along with the motives for exploiting it, is well established. Now it is time to initiate an examination of the 'exemplar' of a design/evaluation framework as an instance of the general possibility of the move from the descriptive to the prescriptive.

I have qualified the resources, Gricean and CA theory, as embodying constitutive rules, and I claim this amounts to saying that the knowledge is purely descriptive 'as it stands': that to make it useful for applied purposes some tacit knowledge must be brought to bear in the process of its application; or the knowledge should be made explicit in some way. The next stage, therefore, is to propose P-BSA theory as the means of introducing this knowledge explicitly, in order to add the regulative component to Searle's speech acts (which expressly only conform to constitutive rules) to allow the representation to meet the requirements of a design/evaluation specification and enable the principled move from data to design.

Of the two general approaches to representation adopted during the design stage of SUNDIAL, one is explicitly classificatory and Fraser (1991) accepts that it generates nothing new from the data; and the other is, as I have characterised the term, constitutive, that is to say although it *appears* to offer general knowledge of a helpful nature (Gricean maxims), it does not engender a substantive rule. In *practice*, both approaches may result in something greater than the originally conceived specification, but they will do so through a deployment of tacit knowledge. At the stage of evaluation, this deployment is exercised by the experts, whose tacit knowledge is only elicited *in practice*. In sum, the form of the HCI discipline being operated here is one largely of craft, as defined by Long & Dowell (1989). The authors claim that this is the current state of HCI, but that the aim should be discipline practice, but on the basis of explicit knowledge leading to a process (specify-then-implement) of the design, moving forward from data to design through principled representation. So the question is: Can we confirm P-BSA theory as a representation of language (in particular natural language dialogue (NLD)), resting on the conceptual distinctions which I have made – a representation which will aid a more formal and explicit evaluation of the quality of designed dialogues?

CHAPTER 4

Having adopted planning as the means of introducing the regulative element into a purely descriptive or constitutive account – SA theory, I must investigate how robust is the connection between planning and utterances. The idea of planning as a tool in the representation of cognitive behaviour is not new but it is controversial. It is important, therefore, to understand its limitations. I address some of the criticisms of its employment, but it is not my intention to support it without qualification, and I take the work of Power (1979) and that of Cohen & Perrault (1979) to illustrate how planning in NLD might be positively conceived (in the case of Power). One can arrive at this positive conception if the different manner in which meaning is treated by Cohen & Perrault and Power is articulated with respect whether speech acts are constructed with building blocks or formed by contextual pressure. I point out ways in which the arguments of SA theory and P-BSA theory bias this issue and, alluding to Power, mark the critical difference. Completing this critique awaits the next chapter's examination of speech acts and their part in communication.

Planning and Speech Acts

Introduction

In this chapter and the next, I shall comment on and examine the idea of combining planning and SA theory. I shall deal with some obstacles to planning's use as a representation of human cognitive behaviour. However, there are different ways in which planning can play a role in the production of linguistic behaviour: two approaches, in particular. I want to try to make these differences clearer and suggest some of the implications of such different approaches – over the next two chapters. In this chapter, I am going to concentrate on the details of one approach, partly because it emphasises the part planning plays, underlining the influence of planning on meaning in dialogue and its relationship with the wider issues of context and rational behaviour goals. Although planning is the main subject of this chapter, I shall introduce SA theory and lead into P-BSA theory, with the following chapter's work in mind.

These two approaches are adopted for the purposes of exposing and developing my own views. I intend that my interpretation of Power's approach be a reasonable elaboration of the 'literal' position as offered by his paper (Power, 1979) and his PhD thesis (Power, 1974). But I believe that his paper (with its emphasis on meaning and speech as action) and that of Cohen & Perrault (1979) (with its identification of the planning schema and the speech act syntax) represent a fork on the route of the development of the integration of planning and language.

In the next chapter, and in the light of the exposition of planning in this chapter, I shall examine Cohen & Perrault's P-BSA theory more closely, using, as a basis for my criticism, their courageous attempt to make explicit the integration of planning and the utterance in the very structure of the speech act. This examination leads to a rejection of their approach, and endorses what I believe to be a potential development of the approach adopted by Power (1979).

P-BSA Theory as the Integration of Planning and Speech Acts

For Cohen & Perrault, adopting and integrating planning theory with SA theory, and treating intentions as plans, is what allows them to show "how plans can link speech acts with non-linguistic behavior", and for this reason they limit themselves to "'instrumental dialogues'". By which they mean "such dialogues (as) arise in situations in which the conversants are cooperating to achieve some task-related goal", and they refer to Deutsch (1974). This declaration might give the impression that their approach avoids the real issues by not addressing conversation head-on. However, it should be borne in mind that the correct implication may be that their approach to task-related dialogue is the right place to start, and that it is wrong-headed to consider conversation in its free-ranging configuration, as a subject of study for design.[25]

There have been several attempts to provide some knowledge representation for human cognitive behaviour, in general, for example, Frame theory (Minsky, 1975) and Conceptual Dependency language (as a representation of Natural Language) (Shank, 1975), so why adopt Cohen & Perrault's plan-based speech act theory as the point of departure? The principal reason is to make a clear distinction from the beginning between the broad categories of the 'dynamic' and 'static' components in language *while allowing for their potential integration.* The introduction of plans distinguishes the 'static' conditions of speech acts (those of a constitutive nature) and those which are 'dynamic' and goal-related (those of a regulative nature).

---

[25]Some sort of systematic approach such as the planning/task related dialogue model will be required for the specification of any dialogue. Everyday conversation might be thought of as a complex interplay of 'instrumental' dialogues, in Cohen & Perrault's sense of the term, but where the tasks include recursive dialogue management as well as personal and affective goals etc.

The secondary reason is that the P-BSA theory can be understood as a general expression of what these other kinds of knowledge representation are attempting to achieve in their particular ways. In other words, its development introduces a *generalised* notion of the 'content-oriented' vis à vis the use of frames, scripts, schemas etc. – declarative modes of representation – via the more fundamental concepts of plans/intentions and their association with values, goals or purposes, which also permits the designer to target specific task requirements. With respect to the thesis aims, it is appropriate because the idea of P-BSAs originates in an analytically derived theory and may be used as the model for a general framework of evaluation or design. Thus it may exemplify transition from descriptive or declarative knowledge to a prescriptive form. The prospective NLD framework is, therefore, in principle adequate to stand as champion to the cause of NLD design, while addressing such design and evaluation work in terms of fundamental concepts which will, in addition, allow me to develop a foundational framework to consolidate the NLD framework project.

This chapter will try to show that although Cohen & Perrault's project of extending and testing Searle's SA theory by marrying it with planning theory is valuable in its own right (e.g., as a testing of what Searle (1969) thought of as a Speech Act *hypothesis*), they may have interpreted planning too narrowly or 'inserted' the planning structure at too superficial a level if their aim is to reproduce NLD. The analysis is not meant as a *demonstration* that their approach is wrong, but rather as a strong suggestion that comparison with another approach which has an equal commitment to implementation shows up inadequacies with respect to the higher level concepts of intention and planning (to be taken in conjunction with the comments on Suchman's arguments, later in this chapter), cognitive description and reductionism; and the issue of compositionality of sentence meaning. However, the next chapter will concentrate on some of the details of the 'mechanics' of Cohen & Perrault's P-BSA theory to show how it falls short in its detailed account of communicated meaning, a criticism which may also have implications for SA theory and its status: a result which endorses what I believe to be the better perspective on the integration of planning and speech acts.

SA and P-BSA Theory

In "Speech Acts" (Searle, 1969), the author is concerned with the relationship between facts and values – a traditional philosophical problem – and how it can be resolved by an understanding of rules. Though he introduces and adds his own refinements to Grice's theory, and though he is concerned with communicational conventions and their connection with meaning, he formulates the conditions of making a speech act, and the rules which constitute it primarily to develop the implications of these rules. Cohen & Perrault go further by embedding these rules in planning systems, to test their adequacy for the generation of utterances, and they recognise that this method needs a corresponding theoretical formulation for the understanding of such utterances; a formulation addressed by Allen & Perrault (1980). Cohen, Perrault and Allen (individually or severally), on the one hand, flesh out, and to some extent, reformulate Searle's theory; on the other hand (together: Cohen, Perrault & Allen (1982)), they also attack the shortcomings of computer systems with language interfaces. Their view is that question-answering systems frustrate users because the latter expect, in quite simple cases, that the system will grasp their unstated aims. They believe that techniques needed to achieve this second-level understanding should be special cases of more general abilities: these abilities include both recognising the interrogator's plan and using plans to provide a helpful reply. It is the integration of these general abilities, *specialised by their application*, which results in the translation from the constitutive to the regulative: the descriptive to the prescriptive. At least part of that specialisation is exemplified in Cohen & Perrault's P-BSA theory. Thus, they take meaningful expression beyond the merely (with respect to design) constitutive stage, and locate the dialogue acts with respect to the tasks about which the dialogue participants are both, to an extent, aware and interested parties. The question that remains is whether this formulation of the SA conditions, and even its elaboration by the infusion of planning, increases significantly its adequacy as a model of human *communication.*

Austin, who laid the foundations of speech act theory, and whose version, it might be said, is the subtler (at least partly because he postpones the constraints of systematisation, leaving it to others such as Searle), emphasised that speech was action. This action was of a significant type. It was not simply an absent-minded action such as running your hands through your hair (as Dummett (1989) points out). So speech was not *just* the appropriate production of an utterance. Potentially, it had to make some intentional change to the world, and there were rules to be followed. Problems posed by the non-existent king in "The present King of France is bald" had been tackled by Russell (1905) in the older truth-functional tradition by its reduction to a statement in the existential predicate calculus. This unnatural, though logical, solution was tempered by Strawson (1950). However, the refreshing alternative, without the logical presupposition offered by Strawson, was the more direct expression by Austin (1962) of an ineffective because inappropriate act: that the utterance, 'The present King of France is bald', he suggests, one might assimilate to the class of actions like "purporting to bequeath something which you do not own." In other words, though it was difficult to fob off logicists with the observation that the statement was neither true nor false, it was relatively natural to characterise an action as 'void', e.g., the action of bequeathal (Austin uses 'infelicitous'). There is, of course, the consequence, of this new way of looking at language, that if speech is composed exclusively of acts with *performative* criteria then even truth may be compromised. I shall briefly address this issue later, in Chapter 11.

The broad dichotomy developed by Austin was defined by the existence of the 'true' and the '*felicitous*'. The import of this dichotomy was that apart from utterances which could represent the facts or not (be true or false) there were so-called '*performative*' verbs such as 'promise' which were appropriately used if they fulfilled certain requirements, and were therefore felicitous or not. For example, (negatively with respect to the requirements) you could no more (and correctly mean it) say, "I promise to be there at 3pm", and *sotto voce* "but I'm not coming", than "I was at home sick in bed yesterday" when you were, in fact, at Wembley. However, the first one is not false and the second one is. It is true that one might say of the deceitful promise that it was falsely made, but the use of 'falsely' is misleading. Consider the difference between "I'll be there at 8 o'clock" meant as a promise and the same utterance meant as a prediction. It is the different ways in which the utterance can be fulfilled which mark it out as either performative or, what Austin originally contrasted it with, '*constative*'[26]: respectively, though the same *locution*, their *illocutionary* forces are promising and stating.

It is that aspect of speech act theory which I shall take up later. That is to say, the significance and extent of the above distinction. In some ways, this is more the work of Austin than of Searle. But I want to pick up from Searle who drew heavily on Austin's insights, particularly where these extensions and variations on Austin's original ideas have been exploited by some designers of computer systems. Leaving aside the finer distinctions and their significance, the salient property of language noticed by Austin was, as noted above, its social and active role. It is this feature of language which allows it to be embedded in systems which might carry out tasks

Speech Act Theory – an Interpretation

Although they have been insightful and helpful in a general way, SAs have not been without their problems (Searle, who has tried to systematise Austin's ideas, admits that there are no fine lines dividing one sub-class of SAs from another, for example.), and I think we should consider SA theory as resting on the fundamental belief that speech is not something entirely over and against the world but part of it – a class of acts: communicative acts.

I shall understand the speech act, at this stage, as the encoding of particular intentions (Bruce, 1975), in terms of presupposition, conventions and goals. It was the recognition that language was not exclusively propositional and truth-functional, but rather fell into the class of actions, which led to the development of the ideas embodied in SA theory. Indeed, if it did not belong to that class of actions there would seem to be a problem in relating language and the world of tasks and values. As Bruce (1975) writes, "A string of words, per se, is not associated with any plan or goal. But an action is; *in fact, the full representation of an action seems to require a representation of both its actual and its intended effects, its actual and its assumed*

---

[26] His early view had been that there were quasi-observational utterances which were not performative (see Austin (1971)). He came to believe that all utterances were performative.

preconditions."(my italics). This view prefigures the work done by Cohen & Perrault, and Allen (e.g., Allen (1979)), at the University of Toronto.

Speech Act Conditions

Let us take the act of promising as a class of speech acts. It is defined, by Searle, in a particular way, and is enacted under the following conditions:

•Normal input and output conditions

•The utterance is composed of a propositional element and an illocutionary device

•In the propositional element the subject predicates a future act of him/herself

•H prefers S doing A and S believes H prefers S doing A (to distinguish threats from promises, for example)

•It is not obvious that S will do A in the normal course of events

•Sincerity condition

•Essential condition: that S understands that the utterance puts him under an obligation to do A

and some further conditions, which are general to communication, to do with the recognition of the intention on the part of the hearer and the part this recognition plays in the fulfilment of the meaning of the utterance. From these conditions Searle derives five rules, of which only Rule 5 – the essential rule – is treated explicitly as constitutive in Searle (1965). The others are either preparatory, deal with the content (the predication of the future act) or with the sincerity of the utterance. These rules are, however, all treated by him as constitutive later in his book "Speech Acts" (Searle, 1969). In Searle's terms, this illocution is an activity whose existence is not logically independent of the rules. What it is resides in the rules. Particular promises, however, cannot be generated by these rules, since the latter do not concern particulars nor are they constrained by a context. This is the point where the plans will come in, and it is relatively easy to see that the above breakdown of the elementary parts of the act of promising lend themselves readily to translation into the planning idiom.

Searle's elaborated view, however, of Speech Act theory is not required for the thesis purposes. Indeed, it seems somewhat restrictive. Since it is an aim of this project to provide a framework for the possible evaluation of NLD it should be as uncommitted as it can be, offering as few constraints on the conditions of language use as possible:

- The theory normally operates with sentence-level utterances . It is, however, the case that in normal conversation people use one word utterances which are actually construed as full-scale illocutionary acts. In fact, it is the one-word utterances which draw attention to language as action. Indeed, Searle (1969) in his book "Speech Acts" states,

> "...the characteristic grammatical form of the illocutionary act is the complete sentence (it can be a one-word sentence)..."

- Speech Act theory implies that the immense variety of utterances can be reduced to a limited set of acts, which is unrealistic, at least, if not incoherent. A means of accommodation, which is appropriate in the context of design, is to see speech-act classes as defined and constrained by consideration of their area of application.

- Searle's notion of the speech-act is often thought of as carrying with it the belief that meaning is conventional, in some sense. However, it is thought important to assign meaning, to a significant extent, to the sphere of inference. This theme is taken up again in the course of examining the SA theory in some detail. To handle this criticism is also to deal with the phenomenon of Indirect Speech Acts (ISAs).

- Finally, the more general criticism that speech act theory is ill-founded because conversation cannot be based on rules can, I believe, be dealt with[27].

Adaptation of Searle's Rules or Conditions

My view, therefore, of the matter is that every utterance that is meaningfully expressed as, or understood as, a communication is a speech act, for the same reasons as given by Bruce (1975): it is thus they have effects in the world. The speech act is composed of the *illocutionary* and the *propositional* component ("I promise I shall write" is made up of the illocutionary act 'promise' and the possible propositional 'state of affairs' of my writing): the illocutions are the primitives as depicted by Searle's language of representation. Where I differ from Searle is in my adaptation of the application of constitutive and regulative rules (see Chapter 3). If we consider the relative status of the illocutionary and the propositional components I want to say that the propositional component can only be (embedded as it is) constitutive[28] since it can only *represent*; but the speech act of promising (the illocutionary component) follows a regulative rule with respect to it, and it is constitutive only in its general form or make-up. Likewise, *statements* (propositions which may be true or false) according to SA theory do not simply stand for and, *therefore*, assert a truth. They are not constitutive – not simply and automatically true, any more than a symbolic formula is. Like other performative utterances, they posit requirements for their fulfilment, and are regulative.

Just as the propositional component, *relative* to its embedding illocution, is constitutive, we can say that certain rules which seem regulative and thus might provide procedural rules which aid design, are, in fact, constitutive *relative* to some proposed use. To say that something is constitutively true is not to say that it is meaningless, just that it is a vacuous rule where rules are expected to provide *knowledgeable guidance* in a given setting. In the foregoing, I have separated out

---

[27] Schegloff's (1988) views, though interesting, may not be relevant. NLD is domain-related design. *Conversation* is arguably, if pointedly not domain-related (or domain-constrained), not directly subject to design. However, it remains to be seen if recognisable conversation emerges from the domain-constrained approach to NLD design.

[28] This idea will be developed and further illustrated later.

Searle's description of the SAs from his more general reflections on rules, and, given the transition to the P-BSA, I have developed these rules for the purposes of discriminating kinds of representation. I am now bringing these rules back to bear on the speech act components themselves. I believe this step is justified by my aim, but it is also appropriate because of the equivocation[29] which Searle himself exhibits towards these rules. It has also been argued in Chapter 3 that it is difficult to see how the illocutionary force, for example of 'promising', can be supported only by rules of a constitutive kind – without the help of goals which transcend the set of practices, e.g. chess; and perhaps this limits the SA model's ability to embrace human communication.

My interpretation, therefore, of supplementing of SA theory with planning is that this extension fulfils the role of adding the regulative element, while providing a practical connection with the tasks which make up the requirements demanded by an NLD design problem, and which have to be satisfied in and through language. It remains is to be seen how well this can be done, employing P-BSA theory as the instance of this incarnation of planning. Before this assessment can take place, it is necessary to investigate what we mean by the planning component.

Planning and Cognition

Introduction

It is not obvious that the effects of using planning as a test-bed for SA theory will be without problems. We may have to pare away, or otherwise modify, less useful features of speech act theory, as well as positively enable a 'real' model of NL dialogue from it. The important point is that the basic features of speech acts be acknowledged and that planning be accepted as an essential characteristic of human behaviour. Even when we are not indulging in explicitly rational behaviour we all assume we are being reasonable, or if not, then we have to take responsibility for our misbehaviour, and recognise that it might, for example, have to be accounted for causally or behaviouristically. Our typical behaviour is, therefore, more or less saliently intentional or goal-seeking, and these functional categories can be kept apart

[29] See Chapter 3, 'Rules'

when necessary, e.g. we can be hungry and not intend to eat. But if we are conscious of our hunger there are usually reasons for the abstinence (there *might* be causes); and what that means is that there are defeasible plans and higher goals or intentions are at work. All we need as a structure of the representation of rational behaviour of the most general kind is a possible mapping between plans and intentions.

Planning Theory

Planning theory means the adoption of means-end analysis which has been traditionally used in problem solving, as a way of representing intentional, rational behaviour . As problem-solving instrument, the end is the state that counts as the solution and the means are the systematic operation of rules + predicates (operators) on the initial state. An example of an operator might be **unstack(a,b)** or **putontable(a)** in the Blocks World (e.g., Winograd, 1972). The goal is thus made true by the application of the constraint rules, the inference rules, plus the operators. This primitive version of planning has been subject to many sophisticated improvements, and the original notion of a plan needing to be fully worked out before implementation is no longer required, as a result of the introduction of ideas such as 'hierarchy of abstraction spaces' and 'partial ordering' of operations/subgoals (Sacerdoti, 1977). It is still a problematic area in which research continues to make headway. A good survey of the 'plan recognition' (i.e., the use of planning as medium for understanding in linguistic communication) literature is provided in a paper by Carberry & Pope (1993). Although there are many who would disavow the approach because of the problems that are obstacles in the way of its progress, it is, nevertheless, still widely viewed as a promising area for investigation. However, certain assumptions underlie research in planning and some of these may prove to be obstacles to its application in the area of NLD. They need to be identified.

There are claimed to be principled arguments against planning's use as a radical representation of human behaviour. Notably these arguments are levelled at the AI community who employ planning theory, by people such as Dreyfus (1985) and Suchman (1987). Since it is my intention to accommodate planning theory in the context of the evaluation of human behaviour I think it is important to address what I take to be the essence of their position. Suchman (1987) appears to use two arguments, the second of which resembles the kind of argument employed by Dreyfus, e.g. in Dreyfus (1985). My aim with the following brief comments is not to claim that real problems cannot be identified within AI conceptualisation, but is to limit the constraining effects of such arguments on the progress of systematic design in the cognitive field.

Planning and Rule Representation in Cognitive Engineering/Science

Suchman's Two Angles of Attack on the Planning Model

In "Plans and Situated Action" (Suchman, 1987), the author denounces planning theory in AI. However, although she is perhaps right to draw attention to limited views of planning in AI theory, her argument spills over into one of attacking the very idea that any real sense can be attached to cognitive artefacts which do things in a similar way to humans. This position, I believe, is unwarranted. I want to show simply that it is conceptually coherent to work on the assumption that a programme of cognitive engineering need not acknowledge that there are any such a priori obstacles to its progress. I am sympathetic to the opinion that planning theory can be misconceived and I shall try to show why; but I do not want to throw the baby out with the bathwater. Suchman's agenda is, however, more radical and must be carefully addressed. I would also like to point out that one can only be positive and constructive with this issue up to a point, in any case, since the whole subject of intention and its relation to action is problematic. It is enough to observe that intention, like planning, is arguably hierarchical and partial, and there are planning schemes which are both. I cannot treat any such analysis in the way which would be required if it were the central issue.

The following are the two ways in which Suchman attempts to undermine planning theory:

(i) Analytic argument

> "If plans are synonymous with purposeful action, how do we account, on the one hand, for a prior intent to act which may never be realized, and, on the other, for an intentional action for which we would ordinarily say no plan was formed ahead of time?"(Suchman, 1987)

I call this argument analytic because it is concerned with teasing out the implications of the relationship between the concepts of planning and intentional behaviour. So it is important to agree on the meaning of the assumptions which she makes. Suchman is taking as read the synonymy of purposeful behaviour and planning, but this assumption cannot be accepted without qualification. There are clearly plans which are means in the context of other plans, but are subsumed as operators within the larger plans. These sub-plans might have effects of which the system is 'unaware' (because the level of the sub-plan is too remote from the primary goals of the system) and could not be properly described as intentional. A system could be designed to have a reference set of values, arranged by priority, which in conjunction with its knowledge would allow any reasonable observer to say that certain actions were intended and others not, while making way for a sense in which the system was 'responsible' for some effects but not others. Equally, we should be able to say that a robot had made a plan but not carried it out, perhaps because it conflicted with another plan. Since plans can be sub-plans or means to other plans they can be "synonymous with purposive action" without being synonymous with any action (connected with primary goals). In the case of a conflict of plans, the formulation of the means to another plan is itself characterisable as a purposive action.

In fact, analogously, human behaviour allows a wide tolerance of what counts as intentional. We often blame ourselves and others for actions the results of which we ought to have been aware, or which simply because they are *our* actions are *our* responsibility. In the human sphere of behaviour, separating plans and purposive action is also difficult: e.g., actions such as braking 'instinctively' to avoid a child in the road (an example in Suchman's footnote), although not explicitly planned, can be said to be intentional because the high priority of the value precludes the necessity of reasoning about it. We may say here that there is a sliding scale where the value assigned to the action is such that the plan need not be articulated over time with prior intention. *The notions of responsibility, intention and planning are just as problematic for us as they are for computerised systems.* It may be that Suchman is correct to claim that there is "confusion in the planning literature over the status of plans", but it is not true that this confusion is because plans *cannot* be employed meaningfully in the business of the design of HCI as representative of the user's mental behaviour or of the intelligent system's 'mental' behaviour. Suchman writes,

> "As common-sense constructs, plans are a constituent of practical action, but they are constituent as an artifact of our *reasoning about* action, not as the generative *mechanism* of action."

But who believes that plans are "the generative mechanism"? The system, or the 'agent', is 'the generative mechanism' in the case of the behaviour of intelligent machines, and it is the person in the case of human activity. Plans are identified with purposeful action because the intention is supervenient on the plan (as Suchman (1987) points out, in an earlier chapter, that according to Dennett "it is in part our inability to see inside other's heads...that makes intentional explanations so powerful"), and seeing inside computers is not much easier. We may still distinguish the plan as the representation of the action and the action as the realisation of the plan, and further, the action itself as distinct from both; likewise, with intention. Indeed, these distinctions allow us to tell apart actions which are physically indistinguishable by virtue of the supervenient purpose or intention.

(ii) Epistemological/cognitive argument

This argument is expressed in several ways by Suchman, and is related to Dreyfus' concept of 'background' (Dreyfus, 1985). It is also connected with the 'frame problem' (McCarthy & Hayes, 1969), although this frame problem is difficult to pin down in one precise form (see Dennett, 1990). The problem concerns the apparent need to continually qualify the axiom system; to deal with the need for non-monotonic reasoning; or cope with the open-ended quality of world knowledge. It implies that any computable representation of human behaviour can only be specified incompletely. No representation, Suchman argues can ever be adequate to the task of capturing human cognitive behaviour. Suchman quotes Garfinkel's example of an exercise given to his students which involved their description of a dialogue. As Garfinkel "imposed accuracy, clarity and distinctness, the task became increasingly laborious", until finally they gave up. Suchman's gloss is that "the task of enumerating what was talked about itself extended what was talked about, providing a continually receding horizon of understanding to be accounted for."

I think there are two issues confused here: the paradigm of reasoning about what we know with that of reasoning about the representation of what we know. The confusion is then dealt with by argumentation which conflates exposing the technical difficulty of describing what we know with the philosophical problem of itemising tokens without a specification of types. If you are asked to count all the 'things' in a room then the difficulty is not to complete the task but how to begin it. However, there are always default categories, and it is difficult to see how one could demonstrate that the task, with these default categories, could not be completed *in principle.* The argument, in short, is an attempt to form an apprehension of the boundless complexity of our knowledge of the world, but the most that is shown is that it is unbounded in abnormal conditions, and it remains to be seen what this 'unbounded' means.

Planning and Dialogue

I want now to look more closely at attempts to carry out the project of combining planning and SA theory, and to see if there is anything to be learned from examining two such projects side by side.

The effectiveness associated with Austin's notion of 'felicity' concerns only the minimal speech act. To complete the picture of *dialogue* it is necessary to indicate how plans play a part in this interaction, not only as simple directed action but as agent-interactive directed action. Both Power (1979) and Cohen & Perrault (1979) incorporate plans in their models of task-oriented conversation (see Deutsch, 1974) (Cohen & Perrault explicitly refer to the importance of adopting 'instrumental dialogues'). Cohen & Perrault write, "The value of studying such conversations relative to the structure of a task is that the conversants' plans can be more easily formalized", and Deutsch has shown how the structure of the task might assist the business of tracking reference. In such dialogues, Power (1979) draws attention to what is meant by a plan-based theory (although he does not actually use this expression) by contrasting it with SHRDLU (Winograd, 1972) which only employs a planning system, adapted from PLANNER (Hewitt, 1969), to carry out the operator's commands. SHRDLU employs planning in the traditional way, as a means-end problem solver, but this problem solver works in one direction only: when the operator interrogates SHRDLU the latter has no notion of what the operator might mean or intend. The machine does not interpret the operator's intentions – *recognise* his plan – but simply carries out the order (the input is a coded trigger for the plan implementation) by problem-solving. Power's answer to this inadequacy (partly on account of "the processes which give the conversation its overall structure" being inside the operator's head) is "to create a system in which the conversants are two sections of a program rather than a program and an operator...They are able to achieve simple practical goals in this world" (their world, that is) "either by working separately or by cooperating". Like Cohen & Perrault, Power chooses conversation which has as its function practical cooperation: "This type of conversation was chosen because its purpose is definite: a remark is relevant if and only if it is related to the process of achieving the goal". Power chose to ensure that "conversational procedures are run jointly, planning procedures privately". This distribution of roles was partly determined by practical consideration (cf. Novick, 1988, who attempted to tackle the complexities of 'turn-taking' and 'repair') and partly to reflect his adoption of the Schegloff & Sacks (1973) concept of 'adjacency pairs' such as question/answer and to simulate this local management. Power characterised this cooperation as partly based on the fact that the agents "construct and execute a single planning tree". Accompanying this single tree, however, is the separate identity of the agents: Power writes, in his PhD thesis,

> "The robots are...identical in almost all respects, being derived from a general robot program which is compiled twice with different variable names". (Power, 1974)

Each agent possesses a copy of the tree and although they need not be identical, they must be consistent. Agents can hold a dialogue if, given a practical task, they have the same expectations or can reach those expectations. It is clear that if agents do not have a minimum of resemblance between their plans, and a minimum of motivation in common, they cannot cooperate and will not communicate with respect to a specific task. The simpler the task the easier this is to see. Of course, the human agent can perceive more easily when its counterpart needs help and when not; and the method of establishing the nature of the counterpart's tree is largely inferential. In Power's model, the robots establish privately whether they can proceed up the tree, and if not the planning procedure calls a conversational procedure (run jointly) to engage the help of the other robot. The resemblance of the program's behaviour to that of human agents is weak because simple, but there is a mapping of key features.

Power goes on, in the second half of the paper to criticise the program's shortcomings, and it has to be borne in mind that we are interested only in the basic correctness of the model of planning with which to combine speech act theory. In the first half, he has been concerned with the task interaction, and now he draws on the insights of Austin and Searle to resolve certain problems of conversational discontinuity. Power wants to examine the resultant dialogue and "spell out some of the processes which lie behind it" thereby introducing a model of planning in NLD somewhat different from that of Cohen & Perrault.

Description and Reduction

The conclusion of Chapter 2 and 3 was that language is not literally and simply 'representative', since although it is composed of symbols and an articulation – its syntax – which allows it to stand for events and processes in the world, it must also play a role in that world: a mediating role to effect change. The requirement was for some content-oriented representation, and this would cover the two aspects of language and perhaps provide a representation of natural language for design purposes. If an explanation and description of language resided in, for example, SA theory, then adding the content in the form of planning provided the bridge to design. But this is not a simple technical device. The planning component makes explicit the danger of a purely descriptive approach. It highlights the reductive tendency of such an approach, and Power's analysis is worth examining in this connection.

The *dialogue state,* a "new notion" (Power, 1979), might be likened to "the board position which has arisen between two chess players as a result of their previous moves". Power is pointing out that a mere description of the 'stack' of utterances or plan moves (the computational history of the conversation) will explain the current position but only with the setting of goals and expectations. Cohen & Perrault have to import an artificial 'cause-to-want' link between the request and the desired action. Power, on the other hand, in his criticism of robotic conversation, tries to account for the continuity and inertia of the speech act and its intended perlocutionary effect; and Power's account is the general one of assuming extant desires driving human action: this is what gives the point to an utterance.

Another perspective on this reductive approach to language can be discerned as follows, providing a more fully cognitive view. Language could, of course, be simply representative, if we were to allow that individuals are not really self-consciously decisive in their actions: the fact that John believes that he will go to Jane's party might simply stand for a state of mind but it would be sadly reductive in its implications if he could not be said be going to Jane's party *intentionally.* That is to say, what we know about John's behaviour (and perhaps his statement that he wants/will go to Jane's party) depends on our understanding his intentions, and how they are related to his commitment to go. In its turn, it depends on Jane's possible disappointment if he does not in fact turn up; or on the means that he might employ to get to Jane's house, and that I might also depend on him for a lift since I am going too.

Point of an utterance and Planning

In other words, we know what John means when he says that he is going to Jane's party not because we appreciate some descriptive property of his nervous system or mental state (constative) but because John's statement implies the making of plans, and plans are public (performative) and can be shared. His statement is meaningful because of its link with what is happening, happened or going to happen in the world, but not by simply standing for what is happening, happened or is going to happen. It can do that too – but never simply and solely. Thus Power's *dialogue state* should not be treated reductively any more than the 'descriptive properties' of a cognitive or neurological state.

Power emphasises the circumstance both factual and intentional by supposing that Bill would be *disinclined* to utter the words, "Will you put the fire on?" to Mary, if the circumstances were any of the following:

1. Bill has just asked Mary to put the fire on and she has agreed.

2. Mary has just announced that she will put the fire on.
3. Mary is walking towards the fire with a box of matches.
4. Mary has just said that she doesn't want the fire on.

and so on. And Power suggests that there might be a separate rule for each, for example in the case of 1: that Bill should not repeat a request that has already been accepted. However, he proposes that "it is more parsimonious to suppose that he knows the following fact: that a request for X to do A causes X to do A by first causing X to INTEND to do A."; and that if this knowledge is taken along with two general principles of rational behaviour: (a) Don't try to achieve goals which are already achieved, and (b) Don't try to achieve goals which are unachievable." In cases 1 to 3, (a) applies and in the final case (b) applies. This principle of parsimony is something which the sharing of plans allows us to comprehend. Power's aims, in the context of his attitude to the 'descriptive property' of the dialogue state, are distinct from Cohen & Perrault's: he puts forward planning not simply as the happy juxtaposition of two individuals' plans (with the awkward addition of an artificial 'cause-to-want' link), but emphasises that the individuals participate, in some sense, in the same general plans. Perhaps Power's position can be best expressed as the view that the dialogue is, *as a whole*, plan-based, though he does not use these terms. By contrast, Cohen & Perrault's characterisation is of each component of the dialogue as plan-based; and intentional in a relatively narrow sense.

Planning Paradigm and the Principle of Compositionality

Cohen & Perrault (1979) is more precisely a testing of Searle's SA theory, which is less wide-ranging in its discussion of SA theory and planning, but allows us to evaluate both SA theory and traditional planning theory in a more systematic manner. The systematic manner or address, however, brings with it certain dangers, in particular, the threat of reduction. A related issue, therefore, to that of the last section is a principle adopted by both Searle (1969) and Cohen & Perrault: that of compositionality. I would briefly like to explain the concept and its connection with the way in which planning is conceived in its partnership with SA theory.

Cohen & Perrault follow the route of compositionality, i.e., the claim that small chunks of text go together to make larger chunks until meaningful units are formed.

Usually this process finishes with the sentence and the principle of compositionality binds in a lawlike manner the relationship between the meaning of the constituent expressions and the meaning of the sentence. The difference might be exemplified by the distinction between, on the one hand, 'construing' a sentence in a foreign language where one follows rules of grammar together with lexical input (compositional); and, on the other, the process of 'interpreting' the import of larger portions of text[30]. The principle is often attributed to Frege[31] and glossed by Dummett (1973) as: "Only in the context of a sentence does a word stand for anything" (note, however, that in SA theory one word may be an utterance because it is a 'concentrated' sentence). That it is capable of different interpretations might indicate the assumption that syntax and semantics are inextricably bound together (see also Ramsay (1990)); Stalnaker (1987) puts it thus, "intentional semantic relations are to be explained on the level of sentence and proposition rather than on the level of simple expressions and their referents". It is true, finally, that in Searle's case his actual mention of the principle of compositionality is in discussing the act of reference and its connection with the sense of an expression being established only as part of an utterance. However, taken together with his view of the contribution of *conventional* meaning of expressions to the meaning of an utterance it is arguable that he adheres to something very like that principle.

The interpretations can be represented as on a gamut from Dummett's, above, all the way to Schwayder's (1994): "the 'Sinn' (meaning) of a complex expression is determined by the 'Sinn' of its constituent expressions" but the elements of the view, if we interpret Schwayder broadly is that the larger chunks of language determine in some way, as well as are limited in some by, the constituent chunks of language.

There are two complicating problems with this formulation: firstly, the importance and extent of the relative influence of the constituents and/or the context on the chunk of language; secondly, that extra-sentential features play an analogous role with respect

[30] It is this difference which is crucial to the distinction between the possibility of a sentence grammar and a story or dialogue grammar (see Chapter 3)

[31] In "Dic Grundlagcn dcr Arithmctik", Brcslau, 1884, according to Scarlc (1969).

to the sentence as the sentence/statement (or utterance) does with respect to the constituents of the sentence/statement (or utterance). Garnham's (1985) example is sufficient to make the point. Compare:

(a) John and Mary got married. Mary became pregnant.

(b) Mary became pregnant. John and Mary got married.

It certainly seems reasonable to argue that we should look at the role that even larger portions of text/discourse play in the determination of the meaning of the sentence/statement (or utterance).

I introduce this notion of compositionality for two reasons: firstly, because it occurs in Cohen & Perrault's P-BSA theory and that it is used there as a constraint on the thread of the analysis; and, secondly, to show that it is both widespread and used by people of very different linguistic persuasion. It demonstrates, at the same time, how difficult it is sometimes to disentangle the conceptual apparatuses of very different philosophies of language. When Cohen & Perrault adopt it in their paper (1979) they do so to underline the systematic nature of their theory: that they should be able to substitute like for like without changing the import of pieces of dialogue. If such a substitution fails then either their own theory needs modification or there are flaws in the SA theory which they are embedding in a planning context. However, a broader contextual influence is clearer when it comes to indirect speech acts and the way their meaning is determined. Just exactly what the differences are in these two approaches (the top-down and bottom-up compositionality) is not easy to understand until some such extrinsic structure as planning theory is employed as a vehicle for SA theory.

For Cohen & Perrault, the rigours and cogency of the planning context reinforces a seminal theory of speech acts by testing it in a rigorous, but as yet, non-empirical way; and that planning answers the requirement posed by the inadequacy of story grammars demonstrated in Chapter 2: that the scope of the determination of meaning be larger than the sentence/statement (or utterance). As Cohen & Perrault put it in the closing remarks of their paper (1979): "It may provide a more systematic basis...a basis that would facilitate the tracking of conversants' beliefs and intentions as dialogue and task proceed." Just what that means will become clearer as, first, the P-BSA is examined more closely and, second, SA theory's illocutionary act is analysed as compositionally top-down or bottom-up. Finally, what needs to be understood is how this evidently fluid principle can be fixed in order to provide a more solid conceptual foundation for the understanding of language representation in NLD design and evaluation. Power's understanding of the more fundamental role played by planning highlights a top-down compositionality, and also underlines the way in which presuppositions about the import of adopting a semantic approach (which appears to have the virtue of formal consistency) might mislead the modeller. I shall return to the issue of the semantic/pragmatic dichotomy in the chapter after next.

Both Power (1979) and Cohen & Perrault (1979) highlight the importance of a planning context for SA theory and it is with this elaboration of SA theory that the most important advance is made. Some of the particular enhancements may not have general impact but at least serve as examples of how the elementary conception of SA theory can be adapted in a purposeful and reasonably formal way. And the variability of the interpretation of the principle of compositionality hints at the variability in the conception of planning which might be adopted; and also suggests that a theory such as P-BSA, if interpreted as based on a narrow conception of the principle of compositionality, will need to be complemented by some theory which is more broadly based and pragmatic. These limitations of the P-BSA theory will be exposed in the next chapter.

Conclusions

The central theme of this chapter has been the place planning might have in a model of the dialogue. I started with a brief description of the relevant elements of Searle's SA theory, and also briefly restated my adapted application of Searle's constitutive and regulative rules.

I considered some objections to planning representation and gave some reasons for rejecting the criticisms as fatal to my project. It may be that there is validity to the criticisms of *some* uses of planning as representation, but as they stand, they are, in my view, too strong.

I next looked at the different approaches to the integration of planning and utterances (or speech acts) which might be adopted. Power's (1979) paper underlines the primacy of planning and, only after developing some view of interactive planning, does he consider how SA theory might offer a way forward. Thus, I believe that he arrives at a subtler view, and links it to contextual factors and the characterisation of what the utterance or speech act is getting at (the 'point of the utterance'). I considered the related implications of the make-up of the utterance: how the meaning of an utterance may be derived from its parts or from surrounding influences: Is the composition determined bottom-up or top-down? This has a bearing on how and where the planning element is situated with respect to the utterance or speech act.

Having looked at the larger picture of planning and speech acts, I want to turn now to the examination of the detailed processes of the dialogue act in order to see how it might bear on the validity of the narrow plan-based (Cohen & Perrault) or broad plan-based (Power) views of dialogue.

CHAPTER 5

I have considered several aspects of planning and its relationship with speech acts, and am inclined to the view that it is unlikely to be narrowly associated with the intentions of the speaker as appears to be claimed by Cohen & Perrault. An examination of the mechanics of linguistic communication should support this view. The question of whether it is conveyed by means of conventional resources or tokens, as claimed by Searle, or via the workings of more general forces or principles, as Grice insists, is addressed: in other words how these general forces, expressed as plans, might operate on the structure of the speech act. This examination of planning in conjunction with speech acts is considered with the help of another researcher's work, and a problem is identified. The illocutionary act as the essence of communicated meaning is analysed and a position is developed, supporting the the previous chapter's conclusions, which is consistent with planning and convention.

The conclusion that the P-BSA theory is, at least, inadequate compels a closer look at the Gricean formula as a complementary theory, and leads to some tentative conclusions, in the light of alternatives described in the last chapter, about how plans in this context should be viewed.

P-BSA and Communication

Introduction

Cohen & Perrault's operationalisation of SA theory tests details of SA theory as well as the concept of planning. Some of the responsibility for problems which I shall identify have to be borne by Searle. Therefore, I start the chapter, which is going to examine P-BSA theory, with a view of SA theory and its conceptual cousin, Gricean theory, which deals with similar problems but in a contrasting way. The suggestion is that Searle's theory is not quite up to the task of representing language in a dialogue, even when complemented by planning.

I describe next Cohen & Perrault's analysis of plan-base SAs and then subject it to some criticism with the aid of another investigation into utterance planning. What issues is a picture of a flawed P-BSA theory, and I follow it with a close look at the act of illocution, concluding the fault lies with the concept of planning and its misapplication to the speech act.

Adequacy of SA Theory for Dialogue Representation

The two theories, Speech Act theory and Gricean theory, which were suggested as useful means of developing evaluative/design methods for the output of the SUNDIAL project, played a major part in the philosophy of language in Britain of the 1950s and 1960s, and subsequently in the development of the subject of pragmatics as part of the discipline of linguistics. Dascal (1994) argues that, though they grew up together and appeared to complement one another, this intimacy may have been more apparent than real. There were differences of course, but they were treated as minor. However, "This divergence 'about the details'", Dascal argues, "amounts in fact to quite different conceptions about the tasks and form of a suitable pragmatic theory, about the division of labour between pragmatics and semantics, and about other issues." Notably, the point of contact between Searle and Grice, which is best known, is where Searle has to find a way of explaining 'indirect speech acts'. As Dascal says, "there are a number of ways in which the theory is incomplete *qua* theory of use", and one of these is that "there are many cases where a speaker does not say exactly what he means". Searle, he goes on, contends that they "are not theoretically essential to linguistic communication" (Searle, 1969).

For Dascal, speech act theory, in its treatment of indirect speech acts, is related to Gricean pragmatics "as if Grice's theory provided merely a sort of 'transformational' component that relates 'surface' utterances to their 'deep structures', the latter being fully and unambiguously described by the semantics of speech acts": Searle, by contrast, is arguing that what is theoretically essential concerns the 'competence' of SAs. When Searle in "Speech Acts" deals with 3 'fallacies', including what he calls the 'naturalistic fallacy fallacy' (sic), he claims that they arise because philosophers confuse 'meaning' and 'use'. Thus, he, like Chomsky, brackets out 'performance'. And, as Dascal points out, though it may be true that this confusion arises and it is important to be aware of it, Searle himself does not have the resources (or has not yet developed them) to account for 'use' in his theory. "Use involves, then, something irreducible to semantics, even to the all-encompassing semantics of speech acts", writes Dascal. Because of this disinclination on Searle's part to venture armed with speech act structures in order to tame what he (Searle, 1983) calls the 'inarticulate background', he stays in the region where he is most in control. This position has implications for what his theory might be likely to explain or account for. Dascal concludes his paper by saying that there is no doubt that the two theories *can* complement each other because their areas of application are complementary ("because each needs the other"). However, he is doubtful that their different theoretical presuppositions will *allow* them to accommodate each other. Dascal cites their theories'major features as follows:

Searle	*Grice*
Monological	Dialogical
Formal	Informal
Conventional	Non-conventional
Grammatical model	Non-grammatical model
Constitutive rules	Heuristic rules, presumptions
Implication	Implicature
Semantic	Pragmatic
...	...

Conversation, since it is 'dialogical', is thus one of the phenomena, according to Dascal, which Searle's theory cannot deal with, and which for him means that "such a structure derives from external principles, alien to the principles of linguistic communication". As Dascal says, "Searle's statement simply expresses his own presumption that his is the only possible model for a theory of use".

I want to suggest that the attempts of Cohen & Perrault (1979) and Power (1979) to extend speech act theory by blending it with planning theory show up some of the inadequacies (indicated by Dascal) of SA theory and that this 'application' of SA theory brings out the contrasts of Searle's and Grice's positions but might permit a reconciliation in an operational context[32]. The significance of this reconciliation is that their elaboration could be from knowledge based on constitutive rules (a kind of conceptual operationalisation) allowing them to be considered on an equal plane, and able to fulfil their tempting compatibility. Their sole raison d'être as philosophical instruments would therefore be overridden. SA theory is affected the most by the modification, which is partly due to its technical detail, but mainly to its role as a means to solving general philosophical problems, unlike Gricean pragmatics which set out on a more modest mission.

[32]Planning, as a means of representing both the illocutionary and the perlocutionary, unites the theories and, perhaps, exemplifies the integration of high- and low-level notions of the principle of compositionality.

Searle introduces his argument for SA theory with the slightly odd claim that he is proposing a hypothesis (philosophical systems are not normally thought of as hypotheses) and he will consider its validity in the light of its success when it is applied to certain problems. Dascal does not think he solves those problems satisfactorily, and I believe that my attempt to 'apply' his hypothesis using Cohen & Perrault's and Power's adaptations of SA theory, which is therefore done in the same spirit as that espoused by Searle i.e. testing his hypothesis, also shows up weaknesses; and in addition, the kinds of weakness exposed suggest a way in which the integration of Searle's and Grice's intent might be achieved.

Cohen & Perrault's P-BSA Theory

Cohen & Perrault (1979) mean the theory to be a 'competence' theory, i.e., it is concerned with the conditions underlying the employment by agents of speech acts as planning operators. They treat intentions as plans, and it is important for them that these plans permit the linking of speech acts with non-linguistic behaviour. One can describe what Cohen & Perrault are doing as mapping the constitutive rules for, say, promising onto the planning variables: pre-conditions, effects and the actions which lead to those effects.

My intention is to concentrate on how well the two main components of the model – speech acts and planning – can be integrated, with a view to realising the general conditions for recruiting linguistic knowledge to design purposes. Leaving aside, then, these issues of formality which arise because Cohen & Perrault intend to create a computer implementation of the model, it is my intention to focus rather on the framework as a matrix for such a model among others.

Searle writes,

> "The procedure which I shall follow is to state a set of necessary and sufficient conditions for the performance of particular kinds of speech acts and then extract from these conditions sets of semantic rules for the use of the linguistic devices which mark the utterances as speech acts of those kinds."

Some of the conditions which Searle invokes as necessary and sufficient for a speech act are controversial and their exact status is questioned by Searle himself; indeed, Cohen & Perrault feel that their analysis endorses Searle's doubts and raises others. I shall take only central ones to illustrate how Cohen & Perrault undertake their project.

An example of one of the speech acts analysed by Searle is that of 'promising'. Some of the distinctive conditions of promising are as follows: (a) it concerns a future act; (b) the speaker believes the hearer can do the act; (c) the hearer is able to do the act; and (d) the act would not be undertaken, in any case, in the normal course of events. The rules derived are that for such a promise to count as such these conditions must be adhered to, among others. They are what Searle calls the 'constitutive' rules for promising, because they appear to be related in an analytic way with the very notion of promising – as a general concept.

Cohen & Perrault approach this list of conditions with caution, and, indeed, claim that their examination of Speech Act theory set in a planning context might do the job of refining this list better than Searle did. Condition (a) relates to the proposition which is the object of the request so is not directly connected with the form of the request. Treated as a part of the plan it is marginalised. However, (b), (c) and (d) might contribute to the form of the planning operator. (b), (c) and (d) are what Searle calls 'preparatory conditions', but Cohen & Perrault sideline (d) because, like Searle, they suggest that it is a general feature of all speech acts, i.e., a feature of rational behaviour rather than language. So, they come down on (b) and (c) as what they posit as preconditions of the planning operator. In their terms, and taking the example the illocutionary act as that of 'requesting':

REQUEST (body)

CANDO. PR: Speaker believe Hearer can do Act; Speaker believe Hearer believe Hearer can do Act. EFFECT: Hearer believe Speaker believe Speaker want Act

The planning operator has the sequence: preconditions, request utterance then effect. There is an issue of what the effect should be, since it might be some change in the world or simply a change in the state of the hearer's mind: respectively, the distinction between the perlocutionary and the illocutionary effect. I deal with this distinction later in the course of analysing illocution and communication.

Cohen & Perrault set out to test, in a way (as I said above) which is consonant with Searle's own application of his SA hypothesis, the structure of Searle's system by employing it in a planning setting. They acknowledge the work of Bruce (1975) and Bruce & Schmidt (1974) as inspiration. In short, their aim is to reunite SA theory with an intentional framework, exemplified by planning theory. As I have said already, I cannot enter into a thorough discussion of relevant similarities as between intention and planning. My claim is only that it is possible to *adapt* planning as a representation of intention; and that intention and planning are intimately related. My general feeling is that mapping intention to planning clarifies intention primarily because it assumes that, for the former to be used meaningfully, it must be public. However, it is important to understand that planning can be used to represent intention at different levels, and these levels should not be confused. As with all artefacts (and planning is a component in cognitive artefacts), they are artificial, i.e., the components are not going to have the exactly the same characteristics as the 'originals'; but this difference does not mean that they cannot be functionally adequate.

I choose Cohen & Perrault's theory because it provides a prima facie model which overcomes the attenuated version of the constitutive theory by, as it were, offering the syntax of the SA theory with a potential semantic structure, thus answering the suggested need of a content-oriented grammar (see Black & Wilensky, 1979). Moreover, it is offered as a critical theory, and, therefore, at the same time, reveals other ways in which this 'application' of SA theory might both enlighten the linguistic/philosophical problem *and* serve as a 'plausible' transition from descriptive knowledge to a prescriptive form. I do not claim that Cohen & Perrault are interested in this 'implementation' for the same reasons as I am. Their aim is more practical and targets the basis for a solution to a particular kind of design problem. Their importance for me lies in the emphasis they put on planning and speech. Bruce

(1975), by contrast, is concerned with planning and belief, and his work is a continuation of the research to provide tools to aid *understanding*.

There are, however, ways in which Cohen & Perrault's model is inadequate as an aid to the design of an NLD device. Some of these inadequacies are, I believe, evident when compared with Power's work. So, apart from reporting what Cohen & Perrault discovered with their 'testing' of the Searle's SA theory in a planning context, I want to suggest ways in which these inadequacies become apparent *and* how, *while continuing to use SA resources*, we might be in a position not only to sketch out a framework which is able to offer the systematic nature of a 'competence' model – all that Cohen & Perrault set out to build – but also look forward to one which answers the requirement of 'coherence' (while still remaining a 'competence' model). As it stands (or, as it stood then) Cohen & Perrault saw that their job was to deal with one half of the dialogue, the generation of plans to *support* the 'production' of utterances. Those utterances, or others like them, would be deciphered (to produce the other half of the dialogue) by the correlative planning inferences in the manner of plan-based recognition of speech acts – work done by Cohen & Perrault's colleagues, including notably James Allen (& Perrault, 1980). Thus, something like Grice's theory (Grice, 1957) appears to be needed to bridge the gap. In the concluding remarks of the P-BSA theory, Cohen & Perrault write,

> "we have so far discussed how a planning system can select illocutionary force and propositional content of a speech act, but not how utterances realizing it can be constructed nor how illocutionary acts can be identified from utterances" and they go on, "extending the plan-based approach to the first area means investigating the extent of the 'pragmatic influence' of linguistic processing".

This position is more tentative than the one they held in a paper[33] published the year before: "We view the plan related aspects of language generation and recognition as *indissociable*, and strongly related to the process by which agents cooperate in the achievement of goals". I think that by the time of the Cohen & Perrault (1979) they were developing a semantic theory (of 'competence' as they say), and the pragmatic

[33]If judged only from the title: "Speech Acts as a Basis for Understanding Dialogue Coherence" (Perrault, Allen & Cohen, 1978).

aspects were still outside their grasp because they had not broached the issue of cooperative planning, as Power had attempted to do. However, these efforts were themselves, I believe, the beginning of a process which would reveal what these pragmatic aspects might be.

A planning context, as I have noted, allows us to take a natural step from Searle's conditions for speech acts which are, as he puts it, 'constitutive' of that speech act: in other words, these conditions are analytically related to that speech act, understood as a general concept. The conditions are developed by Searle into rules for these speech acts. But for the purposes of transposition from simple speech act theory we need only think of them now in terms of conditions. Thus there are what Searle calls 'preparatory conditions' which in a planning context translate easily into preconditions of an act, or of a particular action. Traditionally, a 'planning operator' comprises preconditions, bodies and effects, which may be the goal of the behaviour or some means towards that goal. Ordinarily, this operator would amount to a minimal plan. The speech act differs slightly from this model and this difference may be significant (we shall come back to it later): the body of the act is the utterance (or utterance type), the preconditions are the necessary requirements for such an utterance (or type), and the effects are the way the hearer's model of the world is changed.

This distinction between the different kinds of consequence of the utterance will be significant, and the actual characterisation of this difference in formal terms is problematic, because we do not have enough knowledge of the determining cognitive factors; but the argument is not affected. The distinction was captured by Austin in the terms 'illocutionary' and 'perlocutionary': respectively, what one is doing *in* saying something; and what one is doing *by* (or *as a consequence of*) saying something. The distinction is not absolutely clear, but something can be made of it – enough, for the present, and for my purposes. Austin tried various ways of defining the distinction between the illocutionary and the perlocutionary without any real success. However, it is a distinction which is widely used and, in practical terms, understood. I shall come back to the separation of the two when I have looked at the attempts to 'apply' speech act theory (in 'Planning and Illocution').

In referring to Dascal's paper (1994), I was drawing attention to the obvious complementarities of the two theories (Searle's and Grice's) in their 'explanatory' coverage of language, while pointing out that there were fundamental difficulties in their arriving at an accommodation of each other's concerns. The conflict turns, at least partly, on their respective view of meaning. Searle believed that speech acts were conventional and defined by rules which were constitutive of these acts. The conventionality of the acts was what allowed the participants to communicate with reliability. Searle argues that Grice (1957), whose stance on meaning in dialogue is that it is 'non-natural' (and resides in the intention by the speaker to mean something by some utterance by intending the hearer to understand or believe the speaker's utterance by, in part, recognising this intention), derives his 'stability' from the assumption that the communicative intention is focused on the perlocutionary – not the illocutionary act (Searle, 1969). That is to say, it is the intention (and the probable effect it is believed to have) which is paramount in determining the meaning of an utterance. There is some dispute between Strawson, Grice and Schiffer on this issue (see Schiffer, 1972). However, given the difficulty in distinguishing the illocutionary from the perlocutionary, and given that the part convention plays is only one of the bases that Austin employs to distinguish them, we should regard the issue as open. The important point that we may take away from this conflict between Grice and Searle (mediated to some extent by Strawson, 1964)) is that there is a need to found meaning on some stable basis. Searle's analysis, however, in "Speech Acts" is largely formal.

Searle provides examples but his effort is concentrated on the formal description of the structure of speech acts. As a consequence, he has to derive the stability he desires from components of the utterance, otherwise he would have to look to the context. Because the conceptual distinction constitutive/regulative is so important to Searle, for the solution to greater philosophical problems, he eschews this route. I believe that it is possible to take advantage of some of his arguments and look to the 'context' for the requirement that there be a foundation for meaning and communication. The starting-point is the reasonable adoption of planning as the setting for speech acts. Hornsby (1994) writes, "The true significance of *illocution* is shown when speech act theory is located in a broader, social context...". With a view to widening the application of planning, I shall assume that the "social context" is equivalent for my purposes to the part played by goals and tasks, on the reasonable assumption that social interaction is goal- and task-based in the same sense that language might be plan-based. Both social interaction and dialogue may, of course, set goals and implement tasks: instigate them as well as be instigated by them. I shall come back to Hornsby's analysis after a closer look at Cohen & Perrault's paper. The import of her quote is that there is some sense in which the illocutionary and the perlocutionary cannot do without one another. For a framework's rationale we do not need to drop to the level of detail which is practised by such as Strawson, Grice and Schiffer.

(The introduction of planning allows Cohen & Perrault to indicate differences between 'pre-conditions' and reasons for a speech act, e.g. Searle's so-called 'sincerity condition' appears to be a pre-condition for an INFORM speech act (SPEAKER BELIEVE PROPOSITION), but, in the case of a REQUEST speech act (SPEAKER BELIEVE SPEAKER WANT HEARER do ACT). It may well be that on the model of REQUEST, we should think of the precondition of INFORM as SPEAKER BELIEVE SPEAKER WANT HEARER BELIEVE PROP., but, in that case, at least revealed is the fact that there are two possible sincerity conditions, contrary to Searle's paradigm. It may be that one has to be more flexible in one's conception of sincerity conditions: that they are neither simple constitutive elements nor mere psychological states but determined by more remote goals or needs. In other words, a difference is highlighted which might indicate that the important constraints on meaning are not only illocutionary, but also perlocutionary.)

Criticisms of P-BSA

The particular points on which I shall comment are made by Cohen & Perrault on the following:

(i) differences between utterances with the same illocutionary force

(ii) differences consequent on speech act utterances in two-party and multi-party settings

(iii) 'side effects'

Cohen & Perrault set aside the problem of generating the utterances, and concentrate on the planning conditions which they claim could be said to underlie dialogue behaviour which operates according to the principles of SA theory.

(i) above refers to Cohen & Perrault's noting the claim by Searle that "if two utterances have the same illocutionary force, they should be equivalent in terms of the conditions on their use". They go on to cite the possible equivalence of, "please open the door" and "I want you to open the door". They write by way of explanation:

> "That is, they *can* both be planned for the same reason.", and add, "our treatment does not equate the *literal* speech acts that should realize them when they should be equated" (my italics).

In terms of the equivalent of Searle's conditions, REQUEST is distinguished from an INFORM OF WANT by its precondition SPEAKER BELIEVE HEARER BELIEVE HEARER CANDO ACT. By manipulating the conditions of the two utterances one can force equivalence. However, they observe, "a speaker (S) having previously said to a hearer (H) 'I want you to do X', can deny having the intention to get H to want to do X by saying 'I simply told you what I wanted, that's all.' It appears to be much more difficult, however, after having requested H to do X, to deny the intention" (goal?) "of H's wanting to do X by saying 'I simply requested you to do X, that's all.". The solution to what they see as an anomaly is to define in some sense the 'meaning' of the speech act by linking it to the achievement of the desired effect of the speech act (the task goal, for example) – the perlocutionary effect. Cohen & Perrault's work does not satisfy this requirement since it is concerned with the conditions for the *generation* of speech acts. It is an important observation by Cohen & Perrault, but it is not entirely clear that they understand its implications for the distinction between the illocutionary and the perlocutionary. Their reference in this section of their paper to "*literal* speech acts" (my italics), and not *locutions*, points to a difficulty engendered by the need to explain so-called indirect speech acts, and the connection this need has with the issue of the general phenomenon of implicature and Grice's 'non-natural' meaning; it does not seem as though their plan-based scheme can account for the distinction which they note. I shall return to this problem, and to some other general problems raised by the introduction of planning to SA theory after examining other features of P-BSA theory.

Next, I want to combine points (ii) and (iii), since some of the details of (ii), in Cohen & Perrault's treatment of multi-party speech acts are less important to my main aim. At the outset, Cohen & Perrault had assumed that their "plan-based approach will regard speech acts as operators whose effects are primarily on the models that speakers and hearers maintain of each other". Thus they adopt a 'point-of-view principle', which since speakers are trying to change the world via acts of speech, is as follows: "preconditions begin with 'speaker believe' and effects with 'hearer believe'". Then they illustrate a multi-party speech act such as that constructed to get John to ask Tom to tell the speaker where a key is, on the assumption that to ask someone for some information one (i.e., the requester) has to believe that the requested agent knows where the key is; if the REQUEST speech act is considered on its own this would be a required precondition. However, it is clear that the originator of the multi-party speech act need not assume it to be a precondition of the intermediate's request. The request is effectively just being passed down the line. For Cohen & Perrault, this observation is important because they believe in the principle of compositionality which states that it is the components of language which, working together in a rule-governed way, determine the meaning overall, and ensure its stability; and the components – the speech acts – cannot change unpredictably from what they are to something else. To maintain consistency, their solution is to modify the 'point-of-view principle' to a "more neutral" one: that no speech act's CANDO precondition[34] should begin 'speaker believe etc.'.

---

[34]The other kind of precondition which Cohen & Perrault employ is the WANT precondition, but the 'causal' relationship of speaker's and hearer's wants is not examined, being dealt with by intermediate actions such as 'cause-to-want' and 'convince' with their own precondition and effects.

Almost as a piece with this refinement, and in the context of multi-party speech acts, they note without much comment that, although they had assumed that there were effects specific to speech act types these effects were encapsulated in the illocutionary structure and consequently in the plan-based schema,

> "by the very fact that a speaker has attempted to perform a particular speech act, a hearer learns more – *on identifying* which speech act was performed, a hearer learns that the speaker believed the various preconditions in the 'plan' that led to that speech act held" (my italics).

These Cohen & Perrault call 'side effects', and they claim that since they are peculiar to the particular speech act and not the speech act type *they cannot be specified in advance*. They allude to what the hearer does with respect to the speaker's act as "infer" (the first mention of inference of speech act plans). I believe that with the admission of this oddity, and the introduction of the 'neutral' *point-of-view principle*, they are changing the whole basis on which the planning of speech acts can be made. This oddity, I shall show, is picked up by Ramsay (1990) and its import is generalised.

With (ii) Cohen & Perrault are removing a feature of the plan-based utterance and, as a consequence, making the plan-based schema even more stark and impersonal, but they avoid thereby addressing what exactly is communicated through the intermediary. However, when it comes to the bit (in (iii)) they have to acknowledge that some meaning is communicated without any marker signifying it in the structure of the plan-based utterance, but signalled by the introduction of the unaccounted for 'infer'.

Speech Acts as Plans versus Speech Acts as Plan-based

The author's project, in Ramsay (1990), is to get computers to understand English. The basis for his analysis is what he describes as a 'game-theoretic' semantics. It is a semantics which is rigorous: an epistemic logic in which knowledge is something that *agents* have and it is secured knowledge insofar as it is defended, as in a game. Ramsay, like Cohen & Perrault, believes in 'compositional' semantics. He believes that the process of grouping words into meaningful structures is governed by rules, "that every semantic distinction corresponds to some combination of constituents which is permitted by a syntactic rule", and "every combination of constituents that is permitted by a syntactic rule generates a unique commitment" (by 'commitment' Ramsay signifies meaning). One of Ramsay's motives for adopting the principle of compositionality is that "it shows the distinction between syntax and semantics to be largely illusory", which means that whenever an expression exhibits a difference of meaning we would expect some distinct 'syntactic marker'. Ramsay, I think, means the marker to be evident in the formal expression of the natural language, not necessarily in the surface structure.

Now Ramsay accepts that he has to account for the way in which linguistic actions affect others' behaviour, including their linguistic reactions, and why and how people go about producing utterances. His approach makes two assumptions: "(i) people's behaviour is generally rational, i.e., directed towards achieving their goals; (ii) linguistic action is not intrinsically different from other kinds of action", and he cites Cohen & Perrault and Perrault & Allen (1980) as among those attempting to discover the principles of such planning as concerns him. After a general introduction to the issue, he takes the embedding of speech acts in planning as representing "most AI work in the area".

Ramsay gives as examples of speech acts:

> Do you know the time?
>
> Can you pass the salt?
>
> The milk's boiling.

It is clear that such utterances can be understood in various ways: "that the intended effects can often be different from their face-value effects", and "it seems as though we need a mediating level of description to connect the surface effect of an utterance and its intended effect in a situation". Such a mediating level is the equivalent of Cohen & Perrault's speech act operator – an utterance type with preconditions and effects – and would provide the structured mediating description which could be integrated with a planning context. Thus, the hearer can judge that the "face-value" of the speech act type has no utility as such and s/he can 'chain' forward to possible goals and backwards to confirm the correctness of the speculation until a plausible result is discovered. That is the position with the 'indirect speech act'. However, *the assumption is that with the standard speech act it has the effect it does because that is the convention.* What this means exactly is not explained by Searle or Austin, but the form of the speech act is seen by the planning theorists, by contrast, as comprising preconditions, a body and an effect. What Ramsay's analysis questions is whether the canonical speech act can be sustained in this form.

This argument is difficult because it is not entirely clear what its terms are. I believe it addresses more than one issue and separating these issues is what ultimately makes plain what is going on. Ramsay first provides what he takes to be a version of a plan-based speech act schema, which signifies that X informs Y of P, with certain preconditions, and that the outcome is the X and Y have the mutual knowledge that P:

inform(X, Y,P)

preconditions: KNOW(X,P)

notKNOW(X, • [KNOW(Y,P)])

effects: $\mu(X,Y,P)$

> "In other words, X can only properly inform Y of P if X knows P. and does not know that Y knows it. The effects of the action are that X and Y should be mutually aware of P."

(Elsewhere Ramsay takes a more logical view of mutual knowledge (μ) as something which we cannot know that we have, but here it does seem that we cannot understand human communication unless we accept (know, in some sense) that mutual knowledge exists.) What he observes is that once we have this mutual knowledge what we know is that the preconditions held; yet this mutual knowledge is *inferred*[35] from the preconditions of the 'act' once that act is successfully carried out and the preconditions are thus known to have held. "In other words" he writes, "there is no

[35]'Concluded' is perhaps a better word since, in the context of a logical analysis of English, 'inference' is too strict. Note though that, as in Cohen & Perrault (1979) the concept of 'inference' intrudes.

need for *inform* to have any effects". He considers the following utterance spoken by a wife to her husband,

You've been in the pub.

and notes that this sentence, of declarative form, would in normal circumstances have to have the preconditions: *KNOW(X,P) and KNOW(X, •[KNOW(Y,P)])* These are not the preconditons of *inform*, and he suggests instead that they might be the preconditions of the schema for *nag*. He comments as before that "neither action has any effects beyond mutual awareness that the action has been performed", and further that "there is no syntactic marker which indicates whether something was an instance of nagging or an instance of informing": both examples have the form of a declarative sentence and as such can be understood either way. However, it is worth noting that, whatever the anomaly of the plan schema, if one considers the 'plan-based' structure there are quasi-syntactic differences in the preconditional properties.

Ramsay concludes that the schemas for the various speech acts are not going to help us work out what is happening in a given situation, because (a) the way we know what action was performed is by knowing what preconditions held, and (b) the only effects of the action are grasping of the preconditions – the mutual awareness of the preconditions. The argument is circular, and is not going to help us. That is to say, STRIPS-like (Fikes & Nilsson, 1971) planning allows backwards chaining from the effects to the preconditions, but here, Ramsay is arguing, the consequence of the speech act is, at the same time a mutual awareness of the pre-conditions. He suggests that "we are, therefore, driven back to a rather simpler action":

declare(X,Y,P)

preconditions: KNOW(X,P)

effects: Nil

So, Ramsay's final conclusion is that

> " any extra effects that X wants to produce will have to be based on information which is available to both parties already. If X wants to inform Y of P then X and

Y must already be mutually aware that X does not know that Y knows P. If, on the other hand, X wants to nag Y about P then they must be mutually aware that Y does know P"

So far so good, but an obvious question needs raising, to clarify Ramsay's argument, and to resolve the dilemma for anyone interested in the project which Ramsay calls 'epistemic planning'. For, in spite of his critical remarks he is not prepared to abandon this project, writing, "we cannot possibly explain how

It's OK, we can have another game.

could possibly be a response to

Do you know the time?

without considering the way the hearer might reconstruct the speaker's plan and goals". The obvious question then is: Why is the effect of a speech act not the changed situation which results from the utterance? A change *does* take place as a result of the utterance. So why does Ramsay insist that the 'effect' place-holder in the action schema is empty? The most obvious explicit feature of the schema to which he is drawing attention is, I think, best seen as the indistinguishability of the preconditions and the speech act. It is not, for me, so obviously the effect which is missing but the act as something which *follows* on from the preconditions. Certainly, an *utterance* might be said to follow on from the preconditions, but the speech act cannot be separated from the preconditions. That is what, from the speaker's point of view[36], defines it. Theoretically, and in the absence of an application, one has the choice of either formulation. The importance of Ramsay's analysis is that STRIPS-like planning (Fike & Nillson (1971)) is not the route to take.

We have proceeded further down the road which was taken when Cohen & Perrault's observed that 'side effects' existed, i.e., the awareness of the preconditions of a given speech act, but the results seem to be much more radical, appearing to undermine that whole idea of speech act theory, i.e., as realised and extended by the introduction of

---

[36]The hearer plays a part too.

planning. What we should consider first is whether the concept of planning requires modification and perhaps then we might be able to pursue the original goal.

This outcome of the examination of Cohen & Perrault's P-BSA theory, therefore, has led to a view consistent with a more radical understanding of what a 'plan-based' system should be like; and it is perhaps more correct to characterise the project of combining SAs and planning as one which actually incorporates the speech act as intimately bound up with planning, rather than *merely* 'plan-based', in the sense of being dependent for its significance on the precondition of the plan of which it is merely an instrument, or intermediate element. It seems that the manner in which planning is joined with SAs is crucial. What bearing might this have on the analysis of the heart of the human communication process – illocution? I shall broach this matter after some consideration of the implications of the above on the concept of planning.

Is Cognitive Planning Different?

I noted the use of the word 'infer' in both Cohen & Perrault's paper (1979) and in Ramsay's chapter (Ramsay, 1990) on epistemic planning. Cohen & Perrault had adopted a fairly uncompromising interpretation of planning, citing STRIPS (1971), which is essentially the formalisation of problem solving as systematic decomposition into simpler problems/sub-tasks. The idea was that the solutions arrived at should be provable, and this emphasis on formal rigour is apparent in the following quotation from "Artificial Intelligence" by Winston (1984), in a section entitled Planning Operator Sequences: "Logic shows what is true as a consequence of what is given, but curiously, logic can also show how to achieve truth by using operators to change things". In Winston's example of planning as 'real-world problem-solving' the initial situation might be (where A and B are blocks):

On(B,A) & On(A,Table)

and the final (or goal) situation,

On(B,Table) & On(A,Table)

The difference between this kind of scenario and the one examined by Ramsay, which allowed him to claim that no effect was achieved, was the non-existence of a distinct operator such as *Putontable(x...)*. It is notable that the British Empiricists Berkeley and Hume emphasised the importance of distinct terms in a causal relationship for such a relationship to be comprehensible. The difficulty then, with P-BSA, as it stands, is that it is based on the assumption which Hobbs (1995) makes that "a plan is essentially a representation of causal structure", and for him "it is useful for explaining *not just human behaviour* but other phenomena as well" (my italics). It is clear that the whole notion of planning, as distinct from decomposed problem-solving, is derivative of human action as *intentional* action, but it is not clear how far planning, as Hobbs views it, can be taken as 'explanatory' of human behaviour, or indeed as supportive of design or evaluative activity. The problems highlighted by Ramsay might be interpreted as a reductio ad absurdum of a too intimate integration of planning and human behaviour. However, I believe a return to the ideas implicit in Austin's analysis of language will allow us to reformulate the relationship of planning, SAs and communication; and, further, these ideas will accommodate the need to *complete* the P-BSA project by dealing with *coherence*, and, consequently, a theory *something like* Sperber & Wilson's Relevance Theory (1982, 1986 & 1995): in effect, reconciling Dascal's monological and dialogical models.

Speech Acts, Convention and Planning

I referred before to the difficulty of establishing a 'cut and dried' distinction as between the illocutionary and the perlocutionary, but that it is felt that the distinction is still important. Hornsby (1994) argues that Austin was not fully aware of the importance of illocution, but made several attempts at marking out the difference. Austin himself realised the "slipperiness" of the etymologically correct distinction, i.e., the difference between what one might achieve *in* saying something and *by* saying something, and as Hornsby points out, "it is possible to think of someone *both* as having ϕ-d <u>in</u> ψ-ing *and* as having ϕ-d <u>by</u> ψ-ing". She deepens the distinction by supposing that what Austin was getting at was the degree of 'basicness' inhering in the action: that "*saying how cold it was outside* is more basic than *persuading John to stay indoors*", but concludes that employing this idea to chart the difference between the locution, the illocution and the perlocution (in degrees of basicness), though it could, she thinks, be argued, still leaves us with the difficulty of drawing the exact line between the two 'prepositional' forms. So what is it that allows us to say that some act is more basic than another other than simply feeling that it is so? Hornsby introduces the idea of 'convention' which Austin (and Searle) both emphasise as an important indicator of the illocutionary, and she does so because Austin spends some time trying to separate clearly the illocutionary and the perlocutionary, invoking an 'in-the-course-of' sense of 'in' as well as a conventional one and a criterial as well as a 'means-to-end' sense of 'by'. She suggests that one theme underlies the gamut: 'simple', through 'conventional' to 'consequential': that there are conventions involved in the uttering of the locution words and phrases etc. with agreed meanings which we require as a foundation for communication, but she cannot agree that the illocution of warning, say, is necessarily performed in accordance with conventions. She envisages that using the words 'There's a bull' "relies on conventional significance...to get into the open the thought that a bull is present", but, she writes, "it is obviously wrong to say that there is a convention that one expresses the thought that something F is present to warn of the presence of something that is F". I want, however, to observe that if the first can be allowed as conventional, and it is a 'constative' (which, after all, Austin came to see as just another kind of performative category), then why not the warning too? The second *seems* inadequate as a conventional communication because employing the variable F deprives the act of its force by omitting to say that it is a question of a bull's presence that is at issue. I do not want to go into this matter any further, except to comment that it may be a mistake to lose sight of what *could* be construed as the attribution of convention to illocution[37]. In any case, I agree with Hornsby that Austin's early concentration on clear examples of performative utterances which form part of institutions such as christening ships, marriage, etc., probably helped to distract him from other structural features of illocution which illuminate what language is. Hornsby wants instead to introduce another property which characterises illocution as communication. I think it is an interesting development but this new feature of illocution which she introduces should be viewed only as an elaboration of its conventional quality.

[37]The link between Austin and Searle's insistence on convention and illocution and Grice's concept of 'non-natural' meaning may permit the enticing integration of the two.

Illocution and communication

Hornsby's enthusiasm to exclude, or play down the attempts by Austin to define the illocutionary act is motivated by the need to explain the illocution's function in the larger scheme of things: that is to say, to explain first how it fits into language as communication, and secondly, and implicitly, how there is a symmetry at its heart (involving the hearer too)[38]. The rough categorisations which Austin provides will not do. I am not here concerned with the social or political consequences of Hornsby's argument which she develops in the last part of her paper, but I agree that a closer look at the illocutionary role in communication might be rewarding.

Certainly, when you issue a warning the addressee may or may not be warned. If he stands warned, he may or may not act as expected. If he runs for the fence then this effect is the perlocutionary one of the warning. However, even if he does not react thus, the illocutionary effect may have been achieved. What is meant by the addressee's being warned is that he has understood the utterance as such. "The speaker relies on a certain receptiveness on her audience's part for her utterance to work for her as illocutionarily meant", writes Hornsby. Where this warning is concerned the receptiveness might rely on the speaker's high-pitched shout or simply his/her belief that bulls are dangerous[39]. Hornsby calls this feature of communication which is enabled by receptiveness to utterances, and which distinguishes illocution, 'reciprocity'. Her formal characterisation of it is as follows:

> " φ-ing is an illocutionary act iff (if and only if) a sufficient condition of a person's φ-ing that p is that an attempt on her part at φ-ing that p causes an audience to take her to be φ-ing that p"

This formulation is close in its essentials with the original Gricean expression of 'non-natural meaning. As Searle (1969) has it,

---

[38]As Dascal above, however, points out: *pure* SA theory is 'monological'.

[39]The pitch of the voice could more correctly be said to render the warning more effective, while it is the belief about the bull that *constitutes*, at least in part, the warning.

> "To say that a speaker S meant something by X is to say that S intended the utterance of X to produce some effect in a hearer H by means of the recognition of this intention".

Grice's concept of 'non-natural' meaning is not, of course, equivalent to conventional meaning (where language is composed of tokens), but suggests an ordered relationship resting on *agreed values and objectives* between the utterance and the thought communicated, which puts it in the same general region as convention. For Grice, it is contrasted with 'natural' meaning, as in, 'Those clouds mean rain'. One can see here another contrast, i.e., that, unlike illocutionary acts, this 'meaning' signifies something which normally follows something else: there is a causal connection. What is intended then by the attribution of convention, by Austin and Searle, might be more than simply derivative of institutional formulae, such as christening ships or other standard commissives, but indicates a special event which is primarily not to be accounted for causally. It is, however, an event for which the speaker is responsible, and falls into the category of actions like that of intentionally moving one's arm, i.e., unlike that of accidentally knocking a coffee cup over with one's arm.

The significance of Hornsby's refinement of the illocutionary act in terms of 'reciprocity' is that the introduction of both parties to the communication permits the possibility of the idea of *consensus on agreed values and goals*, and this is the force of convention – agreement through custom or joint stipulation. In this way, we can, in principle, fuse the monological and the dialogical perspectives on convention, without resorting to a simple idea of language meaning which rests on arbitrary tokens.

A second consequence of this analysis might appear to be that Suchman's allegation (see previous chapter) was correct: that a plan-based account of language confuses the plan and the action – that plans are then "synonymous with purposive action". However, although I took a defensive position in that chapter, it may now be possible, in the light of the above, to adopt a more positive stance: that plans *can* be viewed, in her words *but in contradistinction to her standpoint*, as "the generative mechanism of action", as I shall try to show.

Planning and Uncaused events

As I observed above, planning has traditionally been associated with problem-solving, and, within the AI community at least, has been adopted as a tool to interact with the world. Preconditions, actions and effects have been the constituents of plan, where effects are the goals. I have noted that P-BSA theory comprises 'plans', i.e., speech act schemas, which do not allow of these distinctions, and as I hope I have shown, writers such as Bruce (1975)[40] and Cohen & Perrault (1979) have detected 'peculiar' effects (the 'side effects) which they have not questioned further. Ramsay (1990), however, concluded that there were only preconditions and the act itself, but no effect. I commented that another way of looking at the speech act was to consider that there was an outcome. After all, a communication has taken place but the act and the outcome (comprehension of the speech act) were indistinguishable qua cause and effect. It is widely accepted that we *do* things, and that it is sometimes intentional and sometimes not: I might stand on your toe because I lost my balance as the train lurched, but when I come to a seminar I usually do so for a reason, and am not normally caused to come to the seminar. My point is not to attempt to reduce what some may interpret as the interaction of physical systems – some view of communication as the real exchange of information content, or, alternatively, a chain of mental events related causally – to that of mental communion of a mysterious kind. Instead, it is to indicate that there is a difficulty with the application of a standard view of planning when we talk of a plan-based speech act representation of dialogue.

Putting the different approaches adopted by Cohen & Perrault and Ramsay (1990) side by side provokes the recognition that capturing meaning is trickier than was first thought. Indeed, according to Ramsay, we cannot achieve it without a proper system of representing 'epistemic planning', which he does not see how to construct. I think that what he implicitly confuses is the task of providing a kind of representational language for the purpose of designing NLD devices for specific domains with that of representing natural language as a whole. What Ramsay is taken aback at is similar to realising that language is 'emergent'. But what exactly is meant here? To say it is

---

[40] Bruce emphasises that the plan formulation stage and that of communication via conventions are "inextricably linked since the language conventions for expressing intentions make implicit references to plans

emergent is to say something about its (i.e., language-as-a-whole's) relationship 'with the world': that it does not map directly ab initio (and as a whole) but is founded in our activity and interactivity. This is what it means to be intentional agents. However, the intentionality referred to is not essentially unstable. It is largely common to us agents. In other words, I suggest, language might be *both* emergent and systematic[41], but under different aspects.

If it can be both then Ramsay's incredulity is understandable since though it is systematic the structure of the system is not discernible in the syntax. It is systematic with respect to domains, at least; and, in any case, we are like-minded (and 'like-bodied'). As it were, the domain over all application domains (the domain of 'conversation') is defined by our human requirements. That is how and why we are, in general, mutually comprehensible. If we do not consider planning as conforming to the *causal* planning paradigm, then we do not need to conclude as Ramsay does that a paradoxical situation arises with the extrapolation of his 'epistemic planning'. (He confuses the 'epistemic planning' exercise with the causal paradigm of planning which derives from AI work (Hobbs, 1995 – "a plan is essentially a representation of causal structure").) However, we do have to re-express the *kind* of 'plan-based' interaction which is taking place. Though I want to persist in the claim that language is action, I do not want to commit myself to the position that it is intrinsically the same as any other kind of action, e.g. molecular or mechanical action. I also want to maintain that planning underlies action, but this planning associated with communication is a joint (or kind of shared) planning which rests on a consensus of goal-oriented activity which defines us animals, humans, members of the same culture, members of the same group, and so on down the scale.

The notion of shared plans is to be found elsewhere, for example, in the writings of Cohen & Levesque (1980). It is possible that the authors opposition to the primitive notion of illocutionary acts is on the right track, deriving these acts from shared plans,

---

of both the speaker and hearer".

[41] I mentioned, in the introductory chapter, how Simon's puzzling statement about a lawlike account of language which was the *consequence* of our "collective artifice", would re-surface as a characterisation of the emergent property of natural language.

but their view of a shared plan is one that is simply mutually believed by the agents, defined formally. Their work is an extension of that of Cohen & Perrault (1979), and they are often concerned with, and thus constrained by, matters "from a computational perspective". It may be, however, that by demonstrating the redundancy of illocutionary operators they are really avoiding the issue of how the illocutionary act might be analysed and how it therefore plays its central role in communication; and for the purpose of Part 1 of the thesis, this account would take us away from the intimate connection between the communication in the dialogue and the work (and its subtask) which constitute the object of that dialogue. In another paper by the same authors (Cohen & Levesque, 1990) they move even further from the idiom of planning which expresses so well that intimate connection. This paper, however, is very much more sophisticated than Cohen & Perrault (1979), and they make several new and important observations: they emphasise that language use is not essentially "consciously thought out and planned"; that their theory "suggests that the theory of illocutionary acts is not explanatory but descriptive", which I believe is not adequately appreciated; and, that "the heuristic value of illocutionary-act recognition...remains to be seen. Our main point here is that actual identification adds nothing in principle"[42].

Likewise, Grosz & Sidner (1990) have done work on specifying the operation 'SharePlan' and theirs too is largely motivated by the aim of computer implementation. Though both pairs of authors go beyond what Cohen & Perrault (1979) had to offer by considering the formal description of *cooperation* in dialogue, and though both recognise that some form of cooperation or sharing of plans is a *prerequisite* of communication they do not identify cooperation (as the sharing of plans) and communication closely enough. This is not surprising because their goal is the development of models within the context of current computational specifications, whereas that of my thesis is the widest possible applicable framework (for the design of effective systems) which provides such an account, irrespective of implementation.

---

[42] This point is perhaps the same one as made by Ramsay (1990), and analysed above.

Conclusions

My argument has been that language can be described or best understood as made up of actions, and these actions have propositional components; that one way of so describing language is as composed of speech acts; that they can be described constitutively so that to be what they are claimed to be requires that they are constituted or 'made up' in a certain way and that people using them, in order to use them correctly, must follow constitutive rules derived from the conditions of their use; that these rules, however, are vacuous if we want to do design based on the elementary items of speech acts. To be useful, i.e., to generate regulative rules, they must be placed in contexts (among which are domains of related tasks) and that one way is to embed them in planning systems. Planning serves two essential roles in the process of making SAs useful and intelligible: firstly, it accommodates intentions thus admitting the agent's motives; and, secondly, it realises tasks which must have goals and means to those goals, and tasks make up the stuff of work. In other words, it mediates the agent and the world of work in and through linguistic behaviour, permitting specification for design which is partly the identification of the means to satisfying a set of required task goals.

The relationship of planning and speech acts was not as plain as it seemed and the ideas of both have had to be refined, but I hope to have shown both how this refinement permits alliances which were natural but problematic (how to have a strategy for unity); as well as to show how one can start with ideas which are supposed to be heuristic, and therefore regulative, but which turn out to be constitutive and vacuous as resources for design specification – resources which, however, when recognised as such can be provided with a bridge (planning) to the world of action (tasks and work). I have tried to demonstrate that planning for cognitive engineering is necessarily different when communication is examined at an elementary level. This is so far a relatively negative exercise but shows itself to be consistent with Power's ideas, and different from Cohen & Perrault's[43]; but perhaps

[43] But consistent with Bruce's who writes of his study (Bruce, 1975) that "(T)he principal point of this paper is that generation needs to be understood as an action *in a social context*" (my italics).

something like these two views can be accommodated within a framework for NLD design and evaluation.

Before such an accommodation can be attempted, I want to try to complete the description of the framework by taking, as my point of departure, the inadequacy of Cohen & Perrault's model, and then examining what might be, arguably, a theory which would answer the requirements for these unaccounted-for features of dialogue suggested by Power's work: Relevance theory. It will, I hope, become apparent that this proposed theory is not, however, adequate to the task, and I shall criticise it as yet another resource which fails to measure up to the need to be regulative with respect to the design purpose of NLD. The analysis will serve the purpose of demonstrating the usefulness of the criteria for resources for design, as well as concentrating the mind on how to develop a basis for dialogue coherence (sufficient conditions) on top of Cohen & Perrault's model for dialogue consistency (necessary conditions). We found that in spite of the introduction of planning to overcome the limitations of SA theory, at least for the purposes of NLD design, Cohen & Perrault's model was not, as evidenced by my analysis (and with the help of Ramsay and Hornsby) adequate to the task. If it was right to flesh out SA theory to account for (i) the monological limitation of SA theory, as claimed by Dascal, and (ii) to provide a framework for NLD design, might the problem not lie in the concept of planning (or its application), as I have suggested above? It looks as though Power's emphasis on the primacy of agents' plans, and their interaction, might have been the best research orientation to adopt.

CHAPTER 6

In the last chapter I attempted to show that the partial operationalisation of SA theory in the form of a plan-based version (P-BSA theory of Cohen & Perrault) revealed the need for a conceptual revision. Since Cohen & Perrault's theory was not known to be other than of limited value as an account or model of NLD we might express the extent of its coverage as follows,

P-BSA theory		**?**
consistency	→	coherence
↓		↓
necessary conditions		sufficient conditions

We need to complement the plan-based linguistic model as exemplified in P-BSA theory with a view which answers the need for some kind of coherence. There is, therefore, the problem of extending the NLD framework to cover coherence and, at the same time, the recognition that there may be some difficulty doing so in terms of P-BSA theory. In this chapter, my intention is to consider, in general categorical terms, what kind of theory P-BSA is and consequently what kind of coherence theory it is necessary to seek. In the course of the attempt to place P-BSA theory, I shall analyse and re-describe some common divisions of theories within linguistics.

P-BSA Theory is not enough

Consistency and Coherence

In the last chapter, I started my critical examination of P-BSA theory with the assumption that it was the way to proceed with a design model derived from an analytic linguistic base but 'stiffened' by the introduction of planning for the practical purposes of design. I attempted to show that in spite of what I claimed were its promising features it fell short of fulfilling its promise. I want now to step back and consider how such a theory as Cohen & Perrault's should be viewed and if it is not adequate, what kind of theory or model would be; or in what way it could be supplemented. I take account of what kind of model Cohen & Perrault think that it is, as well as arguing that it can be classified with respect to some general and well used categories. Further, that if it is inadequate, is there a theory or model otherwise available which could compensate for this inadequacy? And if there is a theory or model which falls into a complementary category, can some more general conclusions be drawn about knowledge which will allow us to assess NLD's conformity with an engineering view of cognitive matters?

A further less specific reason for this need to gain some perspective is that much of the argument deployed so far has a bearing on linguistic theory but I have not addressed, except obliquely, some of these points of difference with the linguistic theories which abound. Different linguistic theories may well bear on the issues I raise at many points but I cannot deal with them individually. I propose, therefore, by concentrating on these more general properties and categories of linguistic theories, referred to above, firstly, to individuate my viewpoint with respect to others; and, secondly, at the same time, to focus on just those features which I need to exploit for the more general NLD framework and its relationship with Human-Computer Interaction as an engineering discipline (HCIe (Dowell & Long, 1989)), as well as for the foundational framework which I intend to develop, on the basis of the terms of the NLD design framework, in the second half of the thesis.

It is clear, then, that the presuppositions and assumptions of P-BSA theory, after the analysis of the last two chapters, are not going to instantiate *the* framework, but only a stage on the way to that framework. Cohen & Perrault themselves acknowledge that not only is their work incomplete but that it is by its very nature partial with respect to dialogue: that their theory is only a 'competence' theory (that they cannot account for 'performance' or 'process'), and, in addition to this partiality, they concede half the dialogue ground to the work of Allen and Perrault (see Allen (1979), and Allen & Perrault (1980)), whose work concentrates on the other half of the dialogue – plan-based *recognition* of the speech act. Even taken together, their work will not account for 'coherence'. One way of describing this partial treatment by Cohen & Perrault and Allen & Perrault might be to say that they deal only with that component of the dialogue model which allows us to judge for consistency – the necessary conditions of the dialogue. The missing component is the one which would allow us to assess the sufficiency conditions of the dialogue: a step closer to an explanation of what 'coherence' might be. I want first to examine what sort of theory the P-BSA account is claimed to be by positioning it with respect to some others, and ask how its status or type might reflect on what is needed to complement it, for my purposes. I would like, then, to see how it might fit other more general categories to suit my more general aims.

Competence and Performance

What do Cohen & Perrault mean by saying that they are setting out to construct a P-BSA theory which fulfils the requirements of a 'competence' theory? They write that it is “analogous to competence theories of grammar” and refer to Chomsky (1965) as their source. It appears to be analogous in the sense that their theory “describes the *set of possible plans* underlying the use of particular kinds of speech act, and thus states the conditions under which speech acts of those types are appropriate”. At the end of the paper they describe it as follows:

> “a plan-based competence theory, given configurations of beliefs and goals, speech act operators, and plan construction inferences should generate plans for *all and only* those speech acts that are appropriate in those configurations” (my italics).

Chomsky (1965) writes that he is concerned “with the syntactic component of a generative grammar, that is, with the rules that specify the well-formed strings of minimally syntactically functioning units”. In the context of discussing the changes Chomsky made to refine the relationship of syntax with the sounds and meanings of language, Garnham (1985) writes, “Chomsky intends to relate sounds and meanings by generating in the syntax *all and only* the sentences of the language, and then associating with each sentence a sound and a meaning” (my italics). One difference between Cohen & Perrault and Chomsky is the relative completeness of the scope of application of their respective theories. Cohen & Perrault are only concerned with dialogue; Chomsky addresses competence for language as a whole. A second important difference is what kind of status is being attributed to this ‘competence’, since Chomsky believes that his grammar springs from an innate and mentalistic source. This difference has implications for what goes beyond the minimal claims of such as the P-BSA theory, but the common ground between Cohen & Perrault and Chomsky is that, with respect to the realm of the application of the theories or models in question, the two theories aim to account for all possible expressions without paying attention to the causes or rationalisations for the actual (or *performed*) dialogue or general linguistic expression. Both Cohen & Perrault and Chomsky contrast their approach with a fuller empirical one: in Chomsky’s case, that wider scope is covered by ‘performance’ theory; in Cohen & Perrault’s, the reference is to ‘process’ theory. In both cases, the narrowing of the scope of their attention is determined by the purposes of their projects.

Cohen & Perrault's contrast with 'competence' is 'process', which presumably lies somewhere between 'competence' and 'performance', since it is defined as follows,

> "a *process* theory...concerns how an **ideal** speaker/hearer chooses one (or perhaps more than one) plan out of the set of possible plans. Such a theory would characterize how a speaker decides what speech act to perform and how a hearer identifies what speech act was performed by recognising the plan(s) in which that utterance was to play a part." (my bold)

The mention of the 'ideal speaker/hearer' appears to preclude it from being a true 'performance' theory[44], while a competence theory goes no further than presenting (but not selecting from) a set of possible plans. At any rate, they seem to see this 'process' type of theory as going some substantial way towards a 'performance' theory. Further, they identify it with the work on plan-based speech act recognition by Allen & Perrault, who deal in addition with the phenomenon of the indirect speech act (ISA). A distinction which Cohen & Perrault do make clear is that whatever they, or Allen & Perrault, are doing it is not contributing to 'cognitive process' theory, which "claims would require empirical validation", and Cohen & Perrault comment, "it is unclear whether there could be just one process theory of intentional behaviour since each individual might *use* a different method" (my italics). In sum, this view is close to what Garnham (1985) calls

> "an alternative formulation of the competence/performance distinction, originally proposed by Stanley Peters (see Johnson-Laird, 1983, p167).... In technical terms, the linguistic theory specifies the *function*, in the mathematical sense, to be computed. Psycholinguistics has to determine which of the indefinitely many possible procedures the human language understanding system uses to compute the function".

It is difficult to square this formulation exactly with what Chomsky characterises as performance theory – a linguistic theory which deals with the "false starts, deviations from rules, changes of plan in mid-course, and so on" (Chomsky, 1965). He says that performance would be "a direct reflection of competence" if it were

---

[44]"Psycholinguistics can thus explain *deviations from the ideal* in what Chomsky calls **linguistic performance** ('the *actual* use of language in *concrete* situations' (Chomsky, '65))" (my italics) (Garnham, 85).

> "concerned primarily with an ideal speaker-listener, in a completely homogeneous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance",

and then he adds that this is, of course, impossible. The implication is that the serious study of linguistics *must* concern itself with linguistic competence, which, for Chomsky, is absolute and knowable; that any other knowledge of language is relatively ad hoc, and deals with the chaotic and unpredictable results of error and imperfect capacity. There is no indication that Cohen & Perrault, by contrast, are holding to the view that their model is a representation of an innate mentalistic ability, nor that nothing serious or rigorous can be done in the wider field of, what they call, process or cognitive process theory. What, for them, makes the study rigorous is the assumption of ideality, and one can distinguish between the ideality of a competence theory and that of a performance theory (which they carefully distinguish as a 'process theory' insofar as its *application* is not studied empirically)[45]. (It is the more normally acceptable view that all theory is essentially ideal: it postulates components and regularities, tentatively, and in its applications to real-world situations operates on the qualifying principle of 'ceteris paribus'.) Perhaps, however, the categories 'competence', 'process' and 'performance' are employed too differently in different contexts to have much general import. Can one, therefore, employ more generally used categories to bring these theories, or kinds of knowledge, into line with the wider world of linguistics and design?

Semantic/Pragmatic

Origins and Rationale

In 1938, Charles Morris introduced the trichotomy – semantics, syntactics and pragmatics – in order to refer exhaustively to the three parts of the theory of signs, semiotics (Garnham, 1985). Semantics was concerned with the relations of signs to what they signified; syntactics, with their relations with each other; and pragmatics

---

[45]However, in "Rules and Representations" (BBS 'target article', v3, 1980), Chomsky writes "that it is possible in principle for a person to have full grammatical competence and no *pragmatic competence*"(my italics). Given his technical definition of 'competence' this quote implies that a performance theory could be idealised away from actual performance, and looks closer than his original notion (see Chomsky (1965) quote in the text above) to Cohen & Perrault's 'process' theory.

with the relationship of signs to the user (deictic terms such as 'I', 'here' and 'now'). The salient contrasts which arose were semantic/syntactic and semantic/pragmatic. 'Semantic', therefore, developed a dual aspect. Unfortunately, the two dichotomies were both used in the field of linguistics, though the latter – semantic/pragmatic – is applied more widely (see below). As a consequence, some confusion can arise. A theory, for example, which, unlike Chomsky's might not emphasise the primacy of syntactic rules, could be assumed to be a semantic theory, i.e., one which emphasises the primacy of meaning. However, though concerned with meaning, it might be so but in a pragmatic manner. Further, how this semantic/pragmatic distinction is made varies from writer to writer. For example, Garnham (1985) points out that, according to Montague Grammar, both formal semantics and formal pragmatics deals with truth conditions. But if Gazdar's suggestion that any theory of language which deals in truth conditions is a semantic theory (Gazdar, 1979) then Montague's pragmatics would, contrary to his own convictions, fall within the purview of semantics.

At the same time, syntax (the form of language) and semantics (the content of language) are inextricably linked; and the particular position which Chomsky adopts rests, I think, on a particular assumption: that to understand language in a *scientific* way one must restrict the scope of the theory to what 'competence' covers, and only the form or syntax can be expressed explicitly. However, since what one intuitively accepts as a sentence (with a given syntax) must be an expression which can be applied meaningfully in some circumstances or other, then, in those circumstances it could be *meaningfully* true or false, and, therefore, according to Gazdar – a semantic theory. Chomsky's concentration on syntax is partly methodological and epistemological in motive, but the overall goal is the understanding of the meaningful and *demonstrably* valid use of language. Perhaps we should accept, as Ramsay (1990) does, that syntax and semantics are inseparable.

In the beginning, then, Morris' categories drew attention to aspects of linguistic study, on more or less equal terms with each other. As the discipline developed, the context or world played a critical role in the validation of linguistic knowledge. The boundary between the semantic and the pragmatic became epistemologically problematic. This dichotomy took on a more than merely categorial significance, and this more general distinction, I believe, is that which exists between theories which are, in some sense, circumscribed and complete and those which are essentially open-ended. Although the dividing line between them is disputed and/or fuzzy, most people agree that these categories are nevertheless distinguishable; and have a crucial relevance to the holder's world-view.

Implications for Competence/Performance Dichotomy

Therefore, if performance is understood as adequately empirical, the distinction between competence and performance should perhaps best be seen as defining the boundary between the *autonomous* core of linguistic knowledge (or dispositions), on the one hand, and the relatively ad hoc understanding of the manifold diversity of actual linguistic behaviour, on the other; *leaving aside the validation of any ontology.* The general aim of the P-BSA theory is to concentrate on what can be defined rigorously, accepting that there are no direct empirical cognitive implications. And although Cohen & Perrault's undertaking does not appear to carry with it the kind of ontological presuppositions which Chomsky's does, they seem to agree that they are dealing with linguistic/conceptual issues unsullied by foreign factors such as memory limitation, or other more general cognitive influences.

Putting the details of the process/performance distinction on one side (and also sidelining ontological issues), I want to suggest that the semantic/pragmatic dichotomy used with respect to the division within linguistic theory is a better categorisation of what are referred to, respectively, as 'competence' theory on the one hand, and 'performance' theory on the other; but it comes without the implication that one part is superior to the other. Rather, as I have said, it indicates different realms of application or function. What is characteristic of Cohen & Perrault's or Chomsky's theories is that they tend to the more or less formal. Chomsky has separated his theorising from any empirical consequences. For Chomsky, as Garnham (1985) puts it,

> "if psycholinguistic data did not confirm the predictions of a performance theory, it was not the competence theory that was wrong, but the account of how linguistic knowledge is used".

This is so because Chomsky has arrived at his theory as a rationalist. Although Cohen & Perrault have adopted an analytical approach they have not argued that it is apodictic in the sense which is claimed for Chomsky's. Their theory, therefore, arguably falls within the category of semantic theories in that it attempts to impose rigour, and adopt a version of the principle of compositionality, as I have already noted: features which, if achievable, should guarantee consistency. Cohen & Perrault do not, as I said above, and unlike Chomsky, make any cognitive reality the object of their theory.

It is characteristic of pragmatics, however, in contrast with semantics, that the rules for those linguistic phenomena with which it is concerned are defeasible, i.e., they may always be overridden by higher principles (or the application of the principle at one remove). Consider the maxims of Grice, a writer thought of as dealing with pragmatics: they are guides which are supposed to indicate the conditions for certain utterances, but the conditions under which they themselves operate cannot be specified unless by reference to the principle of cooperation, and then only given the circumstances. But when Cohen & Perrault specify the conditions, they do so with the intention of making them precisely implementable.

They are, indeed, embodying the conditions in a formal setting, using the possible world's model, and are attempting to assess how to make the above specifications in a consistent fashion; they are ignoring those factors which determine how a topic might arise in a dialogue – either in terms of its generation or its comprehension. When they allude to the work (by Perrault & Allen, 1980) which might be done on indirect speech acts by employing similar techniques to the ones they have developed, they do not claim that it is anything other than a competence theory like theirs (though Perrault & Allen do factor in heuristics to support the search through the problem space – a pragmatic device). In the terms in which I am operating, I would suggest that theirs too is an endeavour in semantics, since, as Cohen & Perrault's also does, it adheres largely to the standard principle of compositionality. Pragmatics is, therefore, what is left over in accordance with the Levinson's (1983) catchphrase 'pragmatics is meaning minus semantics' when semantics is concentrated on truth-conditionally determined meaning, conforming fairly closely to Gazdar's view (Gazdar, 1979) – at least with respect to a model. The remaining ambivalence rests on that to which truth conditions apply. As Garnham (1985) points out, both Montague and Gazdar would agree on the importance of fulfilling truth conditions but, for adherents to Montague grammar these conditions apply to a model, not our ordinary everyday reality.

My Revised Attitude to Semantic/Pragmatic Division

There is, then, in *absolute* terms, no clear and univocal divide between the areas of language concerned with semantics and pragmatics. Semantics and syntax should be considered equally (inextricably) determinative of meaning, even if it can be accepted that the knowledge of one is considered more important or more epistemologically secure. Ramsay (1990) gives as one of the reasons for adopting the principle of compositionality that "it shows the distinction between syntax and semantics to be largely illusory". I shall, therefore, like Ramsay, conflate syntax with semantics and call this conflation semantics, considered as one term of the dichotomy semantic/pragmatic. Ramsay, however, makes no mention of pragmatics in his scheme of things, since he appears to believe that 'epistemic reasoning' may handle the phenomena usually associated with this term, but that they will somehow participate in a language representation which follows the principle of compositionality. (However, it is an unhappy assumption since, as I described in Chapter 5, he can find no way of rendering speech acts comprehensible and yet remain faithful to that principle.)

Essentially, my reasons for adhering to the more broadly based distinction (semantic/pragmatic) is to cope with such problems as Ramsay encounters in the potentially unrestricted field of language design, as well as, generally, with the open-ended and creative aspect of language use. Just as repairing/rebuilding Neurath's boat[46] is a good representation of the pragmatic framework of the world, so it is of the employment of language: there are some key components at any time which cannot be pulled out and put somewhere else, but there are others more peripheral

[46]"We are like sailors who are forced to reconstruct totally their boat on the open sea with beams they carry along, by replacing beam for beam and thus changing the form of the whole. Since they cannot land they are never able to pull apart the ship entirely in order to build it anew. The new ship emerges from the old through a process of continuous transformation." quoted by Cartwright et al (1996)

which can be adapted and changed safely. This position is close to what Levinson (1983) considers a strong argument: that just as Chomsky can claim phonology is non-autonomous with respect to syntax ("phonological descriptions require reference to syntactic categories"), so we can ask with Levinson,

> "Is it possible to argue that there is some accepted component of grammar which is non-autonomous with respect to pragmatics (i.e., some component requiring pragmatic input)? If so pragmatics must be logically prior to that component, and so must be included in an *overall theory*...." (my italics)?

It seems, Levinson argues, that semantics requires this deus ex machina. Of course, this emphasis on the importance of pragmatics is expressed in a very strong form in certain performance theories. Sperber & Wilson's book (1986), has as its self-confessed goal a theory of pragmatics with a universal flavour. The distinction, however, should remain, like the competence/performance dichotomy, one defined *relative to a given position.*

At the other extreme, when Ramsay (1990) acknowledged that epistemic planning was a 'cul de sac' he did so because he could discern no semantic marker in the schemata underlying ostensibly different speech acts which would allow him to explain them being understood differently. He was prepared to admit these schemata only as structural features of language, and not, for example, as a consequence of the interaction of plans – a pragmatic device. For Ramsay, language mirrors the world in its structure (see also Situation Semantics (Winograd, 1985)).

What counts, then, as adhering to a semantic, pragmatic, or a semantic *and* pragmatic view may depend on other ambitions or agenda of a more metaphysical kind. Thus Chomsky appears to believe that those structures which underpin linguistic competence are fixed and independent of application, an absolute ontological decision. Others, such as Quine, do not hold with any fixity. I shall, by contrast, employ semantic and pragmatic relativistically. That is to say, I do not believe that they represent, or define, basic and fixed ontological or epistemological features of the world, but that they apply to aspects of knowledge; and knowledge changes both with respect to itself and with respect to its function. The pertinence of the issue to science and technology is illustrated in the following quotation from Devlin (1997):

> "The semantic-pragmatic approach is analogous to the way scientists work. For example, physicists first study frictionless motion of perfectly shaped objects and then try to make use of their results by taking into account the effects of friction and shape. Likewise, the linguist might try to study the semantics of grammatical sentences, free of context, and then add in the various pragmatic effects." [47]

There has, however, been a lot of unease about the use of the application of the term 'semantic', and its demarcation from that of 'pragmatic' – within linguistics. As I pointed out, Gazdar's definition encompassed Montague's version of pragmatics; and Stalnaker (1972), for example, acknowledges that, in a scheme in which "syntax studies sentences, semantics studies propositions" and "pragmatics is the study of linguistic acts and the contexts in which they are performed", "*both contexts and possible worlds are partial determinants of the truth value* of what is expressed by a given sentence" (my italics). Finally, what argues for a detachment of semantic from a narrow definition based on meaning and truth conditions, is the movement which is at the root of the linguistic theory exploited in this thesis: speech as action. What Austin called 'felicity' is the quality which defines the aptness of the use of a sentence. Truth is a component but it is not the be-all-and-end-all. Indeed, representation which is restricted in this way is inadequate to comprehend all these activities which we value across the gamut from science to engineering – as I shall try to show.

My conclusion, then, is that the distinction between the semantic and the pragmatic is a relative one, but that one could postulate that the semantic covered those theories or models of representation which are rigorous, consistent and possessed of some kind of formal basis. The assumption is that there are some underlying rules or structures which explain or justify this rigour; that pragmatic theories, by implication, deal with what is not covered by semantic theory, but about which some general principles might be adduced. Levinson (1983) thinks that the value of pragmatic theories lies in their ability to simplify semantic theories by an appropriate 'reading in' of these pragmatic principles to semantic theory. Needless to say, semantic theorists such as

---

[47] Woods & Roth (1988) refer to this same dual approach, but in the context of 'cognitive engineering'.

the adherents of Montague Grammar[48] would not agree with this view. At least part of the problem about defining the boundary between the semantic and the pragmatic is that semanticists and pragmaticists have different ideas about (i) whether the 'competing' group/s need exist and (ii) if they do, then they contest where the border lies between them. Thus, from the point of view of someone on the sidelines, the best approach seems to be to define them, in the first instance, relative to rigour and completeness; but perhaps we can go further.

Relative but Objective

There need not be instability or vagueness of application, but it depends on the recognition – a reasonable one – that linguistic theorists do not only express a technical vocabulary and sets of procedures on how to use this vocabulary, but they also have a view of what kind of relationship exists between language and what language maps to, or rests on: that it is not only the internal features of the theory which are important. For example, grammar-theoretic/model-theoretic linguists believe in possible worlds, situational semanticists believe in situational features (Barwise & Perry '83), and Chomskyans might believe in the existence of innate ideas. Pragmaticists, by contrast, live in a less predictable world (in principle) and their ontologies are of an emergent kind.

Let us, therefore, take the concepts which we applied relatively but this time, instead of applying them relative to one another and with respect to their formality, apply them relatively with respect to what they are explicating. This move leaves the groups more or less where we would expect to find them, but the new arrangement gives us some leverage which we did not have before. That is to say, what they are explicating is their *domain of interest*. So, although the application of the terms is relative we are not free to postulate the criteria for its relative application. Semanticists might believe that language as a whole (or the important part of it) is circumscribable by their theory (language and the world are co-extensive in a definite way[49]), and

[48] Montague employed the term pragmatics but his pragmatics was formal, as Garnham (1985) points out.

[49] A good example of someone who took both a semantic and then a pragmatic view is Wittgenstein, with the semantic period being that of the "Tractatus", and the pragmatic – "The Philosophical Investigations".

according to rules which maintain truth functionality. Pragmaticists whether they feel they can account for all linguistic phenomena, as Sperber & Wilson appear to do, recognise their principles are not like the semanticist's rules (in the strict sense), since they may govern meaning with no dependence on truth functions, the factors determining the meaning not appearing to be internal to the linguistic elements, nor determined by what they stand for in some straightforward way. Their positions are, therefore, defined relative to the idea of a 'domain': the semanticists relate their *propositions*, say, to some real world or model reference (the domain) and when not direct there is a regular and specifiable relationship between *propositions* (truth-functional as well as truth-conditional); the pragmaticists try to establish principles which accommodate marginal linguistic behaviour, where the *context* of *sentences* and *utterances* plays a role or where these *sentences* and *utterances* are a determining factor of that real reference[50].

Further Illustration of the Case

Polson, Miller & Kintsch (1984) compare and contrast the work of Givon (1984) and that of Bresnan and Kaplan (1984): the latter

> "assert certain language phenomena as a starting point and then develop an explicit formal structure to account for them. There is nothing fuzzy or open-ended about that: they are dealing with a rule system – a formalism – in much the same way in which linguistic and AI have done."

Givon's "main point is quite simple. Natural language is essentially pragmatic, open-ended, inductive, and fuzzy". Under the section heading "Performance versus Competence", the authors write,

> "suppose we accept the lexical functional grammar (Bresnan and Kaplan) as an adequate theory of grammatical competence. For cognitive science, the problem would not end here. We also need a performance component to show how this competence is used by people or machines".

[50] Gazdar (1979) suggests that utterances can change the context (Levinson, 1983).

They go on to suggest that Kaplan's parser is best employed in conjunction with strategies, "where some of Givon's observations could prove helpful"[51]. Here the emphasis is turning from the purely explanatory function of linguistic theory to how they might be used in *engineering* projects. They point out, by contrast, that one can see the origins of the two approaches in

> "the dual origins of cognitive science. Bresnan and Kaplan's work is best understood via the information transmission metaphor: a source and a receiver, with information flowing between them and language as the permanent record of information transfer. That record can be studied for itself".

They write of the alternative approach adopted by Givon:

> "that from the inside. What is important for him are the means that one conscious mind uses to change the consciousness of another, through the use of language. Such a view predisposes one to ask questions about the function of language and to attend to the subtle nuances of language use. There is no corpus to be studied, but an ever shifting communication situation".

These two approaches typify what I am calling, respectively, the semantic and the pragmatic, and as Polson, Miller & Kintsch suggest, they are deeply rooted categories. In their paper, and in Clark & Malt's (1984), it is clear that these authors all believe that the disputes are, in a sense, *academic*. That is to say, treated purely as linguistic theories and not cognitive ones, it is difficult to see how they can be reconciled. Once, however, regarded in a cognitive light, i.e., as part of whole functioning cognitive agent, a way of bringing them together can be discerned. However, even this reconciliation is insecure because the justification for the theories has implications for what counts as explanation and writers differ on that, as Clark & Malt point out: i.e., just what counts as explanation in cognitive science is unclear.

Semantic/Pragmatic Distinction and Cognitive Engineering

A corollary of the 'domain-oriented' view (an extrapolation of the Polson, Miller & Kintsch's account down to the particular case) would be that the object of semantics so defined is semantic knowledge, and that this is the core knowledge of projects,

---

[51] "...pragmatic principles of language can be shown systematically to 'read in' to utterances more than they conventionally or literally mean." (Levinson, 1983) Levinson sees this as the way pragmatic knowledge may

tasks and dialogues, that is to say, where these are bounded by requirements of a more or less precise nature, a job of work; otherwise, the knowledge in question would be pragmatic knowledge.

So, in tune with the above distinctions and their rationale, I want to suggest that we might model NLD as the behaviour of agents under certain cognitive and physical constraints whose behaviour is undertaken, within a certain context, to fulfil certain goals. The work/application goals are what determine the domain of the dialogue; and both the behaviour of the agents and the domain of the dialogue are affected by, and affect, the context. When we refer to the content of the domain we mean its semantics, and they can be more or less (relative to the requirements) carefully specified with respect to the tasks and goals of the work/application. When well specified, these 'requirements' ensure a stable set of relationships of the semantics, but it is part of a complete and careful specification that, at the limit, qualifications are made when other things are in an exceptional state ('non ceteris paribus'). These factors are introduced with respect to the context and affect the semantics of the domain. They are, as it were, the pragmatic 'input'. Thus, the domain is mediated by both the semantic and the pragmatic features, but the domain is potentially definable. Analogous with Stalnaker's scheme, as described above, "both contexts and possible worlds are partial determinants of the truth value of what is expressed by a given sentence". We need not enter into the question of the mechanics of pragmatic devices such as those governed by something like the principle of relevance. This has to be an open question which depends on the availability of technical facilities and requirements.

Finally, Levinson (1983) recognises that relating semantic and pragmatic concerns is highly problematic, but has not despaired of some resolution. He writes,

> "there remains the hope that with two components, a semantics and a pragmatics working in tandem, each can be built on relatively homogeneous and systematic lines. Such a hybrid theory will almost certainly be simpler and more principled that a single amorphous and heterogeneous theory of semantics".

---

simplify semantic theory, as referred to already in the text above.

Perhaps, in the light of the above argument, the aim should be, not a hybrid linguistic theory, but an *accommodation* of the two components via the mediation of design (or study) of particular 'domains', treating the components as aspects of engineering practice and knowledge. This would not amount to a compromise (or hybrid) since it would be rendered legitimate and unified within the design discipline.

Conclusions

There are several terms used to describe linguistic theories: to flag their focus or realm of study, or their status with respect to each other and to the world. In this thesis, which is concerned mainly with knowledge, either cognitive knowledge and its interaction with language or general epistemology, it is important to distinguish differences which appear to run into one another or to separate out categories which look similar but hide differences. I have tried to point out some of these problem categories: semantic-syntactic-pragmatic; semantic-pragmatic; competence-performance; competence-process-performance; and it is my intention to distil out what I need for the purposes of the thesis, while suggesting that those purposes (entailed by the project of understanding language design requirements) could give some sense and stability to terms which appear unstable and difficult to define. In other words, I am cutting a more direct route through the undergrowth of linguistics with the goals of my project as a justification for what might appear to be an oversimplification. It is, however, in the context of the overall aim of the thesis, perhaps an indictment that these confusing categorisations in linguistics exist; confusions which can, nevertheless, be resolved in a operationalised setting.

Now that I have asserted what I mean by the semantic/pragmatic dichotomy, and have placed P-BSA theory in the semantic category, it is time to see what might be required of a candidate for the pragmatic category. The candidate to be considered is Sperber & Wilson's Relevance theory, and the next chapter is devoted to its consideration.

CHAPTER 7

I have tried to clarify the different categories into which theories under discussion might fall; and to provide a stable rationale for the terms 'semantic' and 'pragmatic'. Consistent with the view that P-BSA can, at best, only be a semantic theory, there remains a gap to be filled with some pragmatic model, in order to complete the framework for NLD design or evaluation. In this chapter, I attempt to demonstrate that the candidate, the theory of Relevance (Sperber & Wilson, 1982; 1986; and, 1995), suffers from the same weakness as Grice's CP (Cooperative Principle) when considered as a design resource.

In addition, Relevance theory exhibits properties which exemplify the difficulties of capturing, and *describing*, the *general* features of cognition, thereby presenting an opportunity to resolve those difficulties in terms of my developing frameworks. There are troubling symptoms in the difficulty with which the theory tries to blend the terms of cognitive science and computing; and these symptoms suggest the strain of Sperber & Wilson's project.

Relevance Theory and Cognitive Knowledge

Sperber & Wilson subtitle their book "Communication and Cognition", and suggest strongly that they can close the gap between pragmatism, as, for example, practised by Grice, and cognitive science[52]. The express aim, then, of Sperber & Wilson's theory is what might make it a good example, for my purposes, of how *not* to go about the search for that component (i.e., as part of a cognitive theory) which supplements the NLD framework and provides the coherence. It could be argued, by contrast, that Grice and his followers, whose ideas appear to offer hope to the NLD designer, might be on the track of other prey: that he and they, by contrast, cannot be criticised in quite the same way for not being up to the mark.

Sperber & Wilson's Idea of Relevance

I undertake an examination of the ideas in Sperber & Wilson's book for two main reasons: as a natural answer to the needs of the NLD framework; and, in the event, as a vehicle for reinforcing my own conceptual tools: among others, to provide another illustration of the utility of the ideas of constitutive and regulative knowledge as helpful in the discrimination of those resources which are useful relative to the project of design. This task is a good deal more difficult than that of indicating the weakness of Grice's Cooperative Principle (CP). The complexity of Sperber & Wilson's claims make any real examination beyond the scope of the thesis, but I hope to do enough to

---

[52]Theirs is "an attempt to ground models of human communication squarely in cognitive psychology" (Sperber & Wilson, 1986, p170)

expose the presence of some epistemic confusion. Whether my critique hits the mark precisely or not, I believe it will be close enough to endorse the moves which I make. Sperber & Wilson's work is, I think, a good example of the problems which arise in cognitive science and are to do with the place of explanation – and its limitations. Part of the point of the thesis is to allow these limitations to be defined better, and part is to provide, in the foundational framework, a means of overcoming those limitations. The aim, then, is not to offer a radical criticism of their theory as such (i.e., that their views cannot be saved by qualification of their implied scope and ambition), but to undermine their formulation and its apparent scope (a) to illustrate the utility of some of the developing concepts of the foundational framework, and (b) to underline the distinction between kinds of knowledge appropriate for different purposes.

For me, Sperber & Wilson are unclear on four issues which are symptomatic of their ambivalent position: firstly, their criticism, but partial adoption, of the code theory of communication by shifting it from a central to a marginal position; secondly, their criticism of mutual knowledge and their adoption of mutual *manifestness*, which appears to reject the former, for epistemological reasons, and replace it with the second, for operational reasons; thirdly, the uneasy cohabitation of deduction and the broader, more pragmatic, notion of the inferential process, instead of a full-blown pragmatic analysis – analogous with their treatment of code theory; and finally, the stark opposition between the Principle of Relevance (PR) as what communicators attempt to achieve and the claim that it PR is an ineluctable feature of cognition, i.e., the conflation of descriptive and prescriptive knowledge (I shall make several page references since the sources of the imputed inconsistencies should be identified exactly).

I have discussed SA theory and its relative position with respect to pragmatic theories like Grice's. Bird (1994), whose views I agree with, points out certain ways in which Sperber & Wilson's attack on SA theory misses the target. The Relevance theorists do want to adopt SA theory in some form, but Bird concludes that their attacks on it, as an autonomous pragmatic theory, are weak, and consequently, their useful absorption of it into Relevance theory depends on the cogency of that latter theory.

Some of his views on the success of Relevance theory, drawn on to support my own perspective, are cited below. So, though I shall consider the general reasons for Sperber & Wilson's adoption of a degenerate form of SA theory, I shall not deal with it in particular. However, their attacks on SA theory reduce it, *in their view*, to less than the status of a pragmatic theory. It, therefore, becomes in this attenuated form little more than a code theory, about which I shall have something to say.

I shall also make some general remarks on Sperber & Wilson's rejection of the idea of Mutual Knowledge (MK), and whether this rejection can be supported or if it is appropriate.

Qualification of Some Relevance Theory Criticism

Sperber & Wilson's project has been, I believe, misunderstood, and this fact should be acknowledged. For example, McCarthy & Monk (1994) and Haslett (1987) both fail to appreciate the authors' intentions.

McCarthy & Monk wrongly believe that Sperber & Wilson's work is a simple elaboration of one of Grice's maxims, and then compound this error with the mistaken view of Grice's theory itself as an expression of the principle of cooperation as "an implicit contract to be cooperative". Firstly, Sperber & Wilson (1982) strongly disavow the claim that their theory is a mere extension of Grice's maxim of relation – an assumption which Clark & Carlson (1982) also make: the Principle of Relevance replaces the Cooperative Principle. Secondly, although it is true that Grice can be accused of giving his principle a normative flavour, he talks of it as a principle which "all or most do IN FACT follow" (Grice's caps.) (Grice, 1975,), and in general his formulation of the principle is tentative. Indeed, having considered it, he explicitly rejects thinking of it as "quasi-contractual" (Grice ibid., pp48-49).

Haslett introduces Sperber & Wilson's ideas more fully, and though I agree there is a problem for Sperber & Wilson in the transition from the general cognitive idea of relevance to its 'application' to communication, I think she is wrong to claim that Sperber & Wilson "ignore the fact that humans are not *random* information-seeking machines". This comment is at least unclear since Sperber & Wilson are insisting on the property of relevance as a factor which is adaptively evolved because it is an advantageous property. Haslett writes, that "humans seek information *for specific purposes*". However, Sperber & Wilson are claiming that the cognitive property of relevance is deeper than those express purposes, and they would, I believe, argue that a narrow idea of human purpose as the driving force would not result in adaptive selection. As I point out, Sperber & Wilson exhibit some equivocation on the extent or significance of the intentional role, but *it is not ignored.*

The following section should illustrate the kind of project they have in mind, and leads us to their exposition of the 'mechanism' which, for them, explains and resolves the apparent difficulties – the general concept of relevance and its principle.

Context & Mutual Knowledge

A key feature of Sperber & Wilson's method of address, when compared with that of others concerned with the cognitive phenomenon of language and communication, derives from the authors' attitude to the relationship between context and retrieval of meaning. They, I think correctly, recognise that there is a holistic solution required to answer the problem of the understanding of utterances. I mean by this what I have characterised as an elaboration of the principle of compositionality: not only bottom-up but also top-down. It is this solution which is indicated by the intrusion of 'side effects' in the P-BSA theory, and the consolidating critique of Ramsay's analysis of 'epistemic planning'. There seems to be something like an event at the heart of understanding and communication, but one which is not articulated in time; and consequently one which is not causal.

Given this approach, the difficulty of their book is, in large part, due to a conflict between the aims of satisfying the criteria of cognitive science as well as those of anthropology and linguistics; and this tension is evident in the attempt to retain, as I suggested above, technical devices such as code theory and deductive mechanisms, more at home in computationally-oriented cognitive science; and at the same time to have a unified theory as the goal. Several authors comment on these tensions in their work: Good (1990) suggests that taking Grice's concept of 'cooperativeness' in a relatively narrow sense puts Sperber & Wilson's theory at the other end of the pragmatic spectrum – the former notion more naturally linked to the interests of the dialogue participants, while the idea of relevance is "that the only common purpose necessary for a genuine communicator and willing audience is to have the communicator's informative intention recognised by the audience"; likewise Walker (1989) writes, "relevance in their sense is a strictly cognitive notion, and not in any direct way a function of anyone's interest…". I shall argue that Grice's view is not so obviously narrow or prescriptive; nor is it at all clear that Sperber & Wilson are able to avoid the accusation of prescriptivism.

There are, therefore, signs of two broad categories of difficulty; and they generate a dilemma. On the one hand, Sperber & Wilson are striving for generality. After all, they are attempting to erect a theory of cognitive behaviour, concerned in their case with communicative behaviour. They have detected the difficulty in Grice's view that the maxims are defeasible and perhaps the class of such maxims is problematic in its even-ness as well as in its extent. However, in striving for generality they risk attenuating the substance of the idea of relevance. On the other hand, to base communication on the intentional goals of the dialogue participants would be to lose the generality required by a unified theory with claims to belong to cognitive science. And there are two related problems which they have to deal with pari passu: (a) the object of their concern is in part intentional behaviour and, therefore, normative, and (b) as I have indicated, they are determined to retain the idea of code as one kind of communication, employed at some *subsidiary* level in human communication. It is relatively easy to illustrate the tensions caused by (a), though not easy to dispose of their arguments. Problem (b) is more difficult but, I believe, important. I shall deal first, therefore, with the elements implicated, or underlying, (a): context and content of utterance. I shall then come back to (b).

Context and Content

In one of Sperber & Wilson's first Relevance papers (Sperber & Wilson, 1982) a crucial point in the development of Relevance theory is their judgement of the inadequate role of mutual knowledge: its problematic relationship with the stable background against which the communication occurs, or, more particularly, against which the apprehension of the meaning of the utterance takes place. (It is, perhaps, Sperber & Wilson's view of the impossibility of MK (mutual knowledge) which bars their explanation of pragmatic 'behaviour' as a plausible basis for application in specific domains, i.e., for its systematic exploitation – a position reached at the expense of eliding the problems of reality.[53])

Sperber & Wilson (1982) state that their area of interest is pragmatics:

> "The main aim of pragmatic theory is to explain how successful communication is possible, and, in particular, how utterances are understood."

The distinction I want to make, with respect to their theory, is between those issues which are epistemological and/or general and which, therefore, have no leverage on the retrieval of meaning in any particular situation; and those which deal with the 'mechanics' which are claimed to allow this retrieval. It is not that these two issues are not related. They undoubtedly are. But it is important to maintain the distinction clearly and consistently throughout, and this is something which I think Sperber & Wilson fail to do.

One of the principal motives for Sperber & Wilson's Relevance theory resides in their uneasiness with the idea of MK as the basis on which communication between human agents could be explained. The specific object of their criticisms was the kind of analysis which Clark[54] had been (and still is) doing. Sperber & Wilson summarise the then current view of the "mechanisms" of understanding as follows:

Step 1: to determine context, then,

Step 2: to determine content on the basis of context & and linguistic properties of the utterance, then,

---

[53] "Our point of view here is cognitive rather than epistemological." (Sperber & Wilson '86, p39); and it might more accurately be described as phenomenological than cognitive.

[54] For example, Clark & Carlson (1982), Clark & Marshall (1981) and Clark (1996)

Step 3: to draw the intended inferences on the basis of the content and context

Sperber & Wilson assume that this mechanism rests on a concept of MK which is not helpful to the problem of how utterances are grasped. They use three main arguments against this view of MK as the key to mutual comprehension:

(a) *The identification of MK presents problems which do not give rise to corresponding problems of comprehension.* But why should an epistemological issue determine another (the comprehension issue) which need not be of the same order? Knowing that the cat is on the mat is something we might agree to, while holding that there is a persuasive philosophical argument that we can be certain of nothing. That kind of knowledge is evident in the impossibility of saying, 'I know he was here last year, but he might not have been'. You could say, of course, 'I know he was here last year (pause)…but I could be mistaken'. And, in any case, in a perfectly ordinary sense of MK, lack of it *does* lead to consequences for comprehension (see example below).

(b) *MK is not a sufficient condition for a proposition to belong to a context. Understood as a general foundation it is not clear that anyone has said that it should be anything other than necessary.* I do not think, in any case, that this is Clark's or Clark et al's view. Though they have considered what a shared environment might consist in, they have not gone as far as (b).

(c) *Nor is it a necessary condition – a proposition may belong to a context without being mutually known.* But this can only be true if comprehension or communication is not implicated. It is a weaker but still coherent claim that MK is possible, i.e., *does* take place, although in any instance may not have. In other words, it *is* a necessary condition for an *utterance* (not a proposition) to belong to a context (compare to (b))

They, by contrast, are proposing a "single principle which simultaneously determines context, content, and intended inferences with no appeal to mutual knowledge".

Comments on Sperber & Wilson's Idea of Context, Content etc.

It is fundamental to the question of the relationship of utterance to context just exactly what kind of relationship it can be. There are, at least, two distinct ways of viewing the relationship: the context can be seen as the background for the remark/utterance; or the context can be conceived as inextricably and intimately bound up with the remark or utterance, to the extent that, under certain circumstances, we can understand the remark or utterance as selecting the context. That the latter is possible might suggest that it is always true, and that therefore Sperber & Wilson's view of the matter is correct: "that there is a single principle which *simultaneously* determines context, content and intended inference, with no appeal to mutual knowledge" (my italics). What I want to propose is that the former view is also tenable but, of course, in different terms. I have suggested in the course of criticising the plan-based approach to SA theory that we have to think of planning as non-causal: that there is a sense in which *meaningful reasoning* is non-causal, something which is unsurprising treated as an individual process but seems more difficult to incorporate in communicated thought treated formally, especially when language is considered as action. In Sperber & Wilson's above quotation it is, I believe, out of place to use the word 'simultaneously', since what they are targeting is non-temporal, and so the contrast cannot be made. Nevertheless, whatever the import of their view of context, content etc., I want to maintain that it does not bear on the approach they are criticising (incarnated as Clark and Clark et al, specifically with reference to Clark & Marshall, 1981) because, for Clark and others, communication (one instance of what they call 'joint activity') is behaviour which has to be justified by evidence – that has to be 'authorised': the rationale being that a context thus understood is, in fact, the *ground* for the remark or utterance – a context which also, in some sense, gives rise to the remark or utterance.

Comments on Mutual Knowledge

A second key feature of Sperber & Wilson's theory which they believe sets them apart from the prevalent position of communication is their attitude to MK. Clark and others, who are to an extent straw men (and women) for Sperber & Wilson, talk importantly not only of the 'ground' but of the 'common ground', and identify this common ground with MK. Clark & Marshall (1981) acknowledge that there are philosophical arguments which undermine such an idea by creating a paradox, but they have argued that, equally, there are other arguments which support it. Further, they have developed three categories which encapsulate it: co-presence (physical), co-presence (linguistic) and common community membership. As with the arguments about the relationship of context, content etc., there is, I think, an analogous confusion behind the positions taken by Sperber & Wilson and Clark et al. The analogy rests on the sense one might attribute to the 'priority' of mutual knowledge, just as it did with that of context. Sperber & Wilson's objection assumes that Clark and others must *prove* MK, and only then can they carry on with an account of the process of communication. Their own view inextricably associates comprehension and mutual awareness[55]. The conflict can be resolved analogously if we take Clark and others to be proposing grounds for communication, i.e., '*common* ground' for comprehension – a *public* world which we as participants in communication can access to justify or authorise[56] the remarks we make as well as to comprehend the remarks others make.

Sperber & Wilson, however, have a difficulty which Clark and others need not address, since they (Sperber & Wilson) do not believe in MK, i.e., they have no 'real' ground for *knowledge*. Their common reality is essentially suspect. For Clark and others, the public world exists and we can study ways in which that common world can be explored to progress and justify our joint activity. Sperber & Wilson are closeted in a world of cognitive impressions (not unlike that of the closeted world of sense data theorists) with no access to 'real' knowledge. It is, after all, arguable that all knowledge, to be such, has to be potentially public or common, i.e., implies mutuality.

Code Theory, Pragmatics and Mutual Knowledge

As part of their introduction to the detailed arguments in favour of a principle of relevance, they attempt to found these arguments on a proof that such a pragmatic

---

55 What they call 'mutual manifestness'.

56 This common public setting is a prerequisite for the justification or authorisation basic to the "accumulation of common ground" (Clark, 1996)

theory is necessary, in the first place,. The theory against which they contrast such a pragmatic theory is what they call a 'code theory' of communication. They claim that up until very recently there was no other kind, and that only with Grice did some distinct alternative arrive. Grice argued that communicating with a dialogue partner did not depend on knowledge of a code which each participant required, but rather that we needed only to understand the particular communicative intention of the speaker. Grice's view was part and parcel of the modern perspective that we lived in a world of intention and task, and it was that common understanding which underlay communication – not simple representation and coding/decoding of representation. There are, and Sperber & Wilson cite them, many instances of comprehension which obviously cannot rest on code knowledge, e.g. the unusual use of words in a certain setting, and the subsequent comprehension by a hearer.

Sperber and Wilson go further, and explicitly rule out the possibility of amalgamating code theory and 'inferential' theory (Grice's type of theory). *However*, they write (Sperber & Wilson, 1986), "the code model and the inferential model are not incompatible". This apparent contradiction is removed once the kind of compatibility is exposed. They thus also write from the same source, "a coding/decoding process is *subservient* to a Gricean inferential process" (my italics). So, if a hierarchical relationship is established they are compatible, otherwise not. I want to question this view. It is, of course, true that we use codes but only because we can communicate without using those codes; that is to say, in no sense are they primitive. The use of a code is a conventional means of conveying that which is fundamentally 'inferential', e.g., morse code. However, it is wrong to set code and inferential type together, juxtaposed in *the chain of communication*, when the latter is not fundamental to the former. Given the second quote above, it is surprising that Sperber & Wilson appear to support this juxtaposition. I shall take two instances.

The first is cited as contained in a linguistic module, dealing with basic syntactic form, and referred to as deriving from Fodor's modular model of cognition . They agree with Fodor (1983) that linguistic coding/decoding takes place ("like him we see linguistic decoding as modular" Sperber & Wilson (1987)). The authors appear to treat the linguistic module as in some sense irreducible. It serves a marginal but necessary role in communication, and, by its nature, is not replaceable by (reducible to) some inferential process. It is, therefore, my view (and, I think, *implicitly* theirs) that it cannot, any more than anything else in the communication chain, be encoded. It can be treated as causally implicated, but, of course, this raises new problems of compatibility. The only sense in which they can mean that this use of code is "subservient" to the inferential processes is in the sense that the latter are more central, not that such codes are reducible to such inferential processes, i.e., used in place of the inferential processes.

The second example, although not explicitly described as such in the book (Sperber & Wilson, 1986), is treated as follows in Sperber & Wilson (1987):

> "Our book questions some of the basic assumptions of current speech-act theory, and sketches an alternative approach which puts a *much greater load on inference than on decoding* in the *identification* of illocutionary force",

and is embodied in Sperber & Wilson's use of 'generic speech acts'.

In addition, once it has been argued that non-demonstrative inference is at the heart of communication it is difficult to understand how code can play *any* part in basic *communication.* Sperber & Wilson question MK as essential to communication, and they argue that code theory requires the prior proof of the MK thesis. But if MK just cannot be asserted how can code theory be admitted at any stage in communication, i.e., as an element in that communication? It is difficult to square the admission of the so-called 'generic speech acts' ('saying', 'telling' and 'asking') as such elements with a thoroughgoing pragmatic theory, and in Sperber & Wilson's theory this move seems to assume an arbitrary line between pragmatic and non-pragmatic elements.

Sperber & Wilson's position rests on the impossibility of securing MK, where the MK is of the ground of comprehension. There may be some merit in this view, but it is rather more likely that securing MK, but not *demonstrably* securing it, is what is required, and I believe that this is what Clark and Clark et al attempt to do.

Universal Principles and Their Explanatory Power

Grice versus Sperber & Wilson

The difference between Grice and Sperber & Wilson (1986) is that the latter emphasise that the roots of relevance are in our 'biological imperative' to pay attention to what is relevant ("*The fact* is that the development of the human language was made possible by a specialised biological endowment." (my italics) ("Relevance", p53); as if it were an impersonal force. Grice (1975), however, writes, in the context a famous coded message 'Peccavi' ('I have sinned') sent by a British general after the capture of the town of Sind,

> "Whether or not the straightforward interpretant ('I have sinned') is being conveyed, it seems that the nonstraightforward interpretant must be. There might be stylistic reasons for conveying by a sentence merely its nonstraightforward interpretant, but it would *be pointless*, and perhaps also stylistically objectionable, *to go to the trouble* of finding an expression that nonstraighforwardly conveys that p, thus *imposing* on an audience *the effort* involved in finding this interpretant, if this interpretant were otiose so far as communication was concerned." (my italics).

He uses the word 'pointless', demonstrating a reasoned and goal-oriented stance towards conversation. Sperber & Wilson, however, because of a commitment of an anthropological kind to a deep cognitive drive to Relevance, find themselves expressing contradictory observations such as, "Communicators do not 'follow' the principle of relevance; and they could not violate it even if they wanted to. The principle of relevance (PR) applies without exception: every act of ostensive communication communicates a presumption of relevance." p162, and,

> "The principle of relevance is a generalisation about ostensive-inferential communication. Communicators and audience need no more know the PR to communicate than they need to know the principles of genetics to reproduce."

However, they write, "Our claim is that all human beings automatically *aim* at the most efficient information processing possible." In this quote, there is the uneasy proximity of the word 'aim' which suggests purpose and will, governed by the adverb 'automatically', which normally denies those attributes. There is, as Bird points out, also reference to "being guided in conversation by what they 'should' do according to the principle" referring to p48-49 of "Relevance"[57], and intermediate positions are hinted at by the use of unclear expressions such as "relevance-oriented" (Sperber & Wilson, p152). Thus the claim above that implies a "descriptive, and exceptionless, generalisation" (Bird, 1994) is compromised, and their intended contrast with Grice's normative stance is less than complete. This conflict of formulations of the PR is a good example of the distinction which was made more explicit by Searle (1969), between the constitutive rules of, say, promising and (what was implied) the regulative rules governing the actual institution of promising. When Sperber & Wilson treat the PR as normative they are acknowledging the regulative aspect, but in fact their development of this aspect is nugatory[58].

Grice's principle and maxims may not, indeed, as they stand, assist us directly in our project of supporting design or evaluation, but his position does not imply the claim of a scientific kind, nor does he offer some calculus of Relevance 'effect' over Relevance 'processing effort' to retrieve that 'effect', as Sperber & Wilson do. Underlying the operation of the Gricean maxims, i.e., their proper use and the comprehension of their proper use, is the Cooperative Principle (CP). The principle is vacuous, with respect to design, because (as I have already pointed out) we do not know in what cooperativeness consists, except to say that it appears to indicate that an essential feature of conversation is its cooperativeness. But this definition just means that, negatively, if it were not cooperative it would not be conversation, and this is what is meant by calling it a constitutive notion. In what manner it is cooperative is not specified. We know only that appeal to the principle overrules other maxims, or guides.

---

[57] In fact, Sperber & Wilson's say, of paying attention to ostension, that a recipient of information "*should* do so, that is, if she is *aiming at* cognitive efficiency" (p49) (my italics). This last example comes from a section entitled Relevance and Ostension and although not an example of conversational dialogue they write, "Showing someone something is a case of ostension. So too we will argue is human intentional communication". Later they write, "Inferential communication and ostension are one and the same process".

[58] There are several problems. Essentially they try to create a functional relationship between that of 'contextual effects' produced by the speaker and 'processing effort' made by the hearer. The arguments against their position are many and varied, cf. notably papers by Gazdar, and Wilks in "Mutual Knowledge" ed. Smith (1982), and among comments on their peer- review paper in the BBS (Sperber & Wilson, 1987)). The importance of the operationalisation and testing of concepts is dramatically highlighted by the seemingly endless dispute over what it might mean for Relevance to be normatively employed.

Grice's insight, which has been subsequently maintained and successfully adapted by other philosophers and linguists, is that thinking of conversation along the lines of a simple exchange of information is a dead-end and explains nothing. The problem with Grice's insight is that it is subtle, intuitive and part of wider pragmatic theory of meaning. Sperber & Wilson's work is a development of Grice's insight and whereas the latter's principle is, as part of philosophy, acceptable with all its intangibility, Sperber & Wilson's extension which claims to be a theory of cognition is not acceptable. As noted above and as observed by Sperber & Wilson, Grice's view of the CP is normative and involves, when the occasion arises, the intention to maximise relevance (among other aims), a position which is only confused by the desire to establish a quasi-scientific basis for the theory. Sperber & Wilson confound the necessary feature of exceptionless generalisation with the necessary and sufficient joint properties of generality and operationalisability which are the stuff of scientific laws.

Sperber & Wilson attack the Gricean model for being of only post hoc value, but it is not clear that their own principle is much less so. Relevance is, in part, by definition, a component of understanding, communication and meaning; and these three are, in part, dependent, in their turn, on Relevance. It looks like an ideal candidate for the constitutive category. The mistake (if it is such), for them, that Grice has made is to suggest maxims which communicators should aim for, or 'flout' in a systematic way. Sperber & Wilson are right that, having started out on the road, it is correct to ask if the maxim set is complete. *Had it been Grice's aim to provide a cognitive model*, the lack of a complete framework within which to express the maxims – some kind of conceptual operationalisation, at least – would have been fatal. Sperber & Wilson appear pointedly to avoid the maxims but, while claiming a theory capable of empirical 'extension', they fail where Grice had no need to try. As Bird (1994), therefore, points out, if all they can provide is an exceptionless generalisation, then for it to offer a helpful model (for whatever purpose – explanation or prescription) they will need to make available many ad hoc devices to justify the PR's variable operation, as suggested below, thereby weakening fatally the impact of their principle.

Relevance and Scientific Explanation

If we think of the PR as analogous to the law of gravity it might make things a little clearer. The claim of Newtonian Mechanics is that gravity is a universal attraction between bodies and varies in proportion to the mass of the bodies and the distance of these bodies. The exact proportions are known. Of course we know of instances in which bodies do not 'apparently' fall towards one another. Normally functioning aeroplanes do not appear to obey the law, but then we understand why and when they do not (apparently) obey it. Thus Sperber & Wilson may think of the PR in this way, as Bird (1994) describes it, "PR is an exceptionless generalisation, which we cannot choose to violate even if we wanted to". But Bird goes on, however, to write,

> "It gains its exceptionless character, fraudulently, by *failing to mark any real distinction* between the cases where it does operate and those where it does not" (my italics).

By contrast, the law of gravity has explanatory value because it is independent, and evidently so, of the effect of any countervailing force such as the powered lift of the aeroplane (and here the real distinctions *are* clearly marked). The difference between Sperber & Wilson's position and that of the law of gravity is not simply one of quantification, which they also avoid by stipulating that the 'cost benefit analysis' is non-quantitative but 'comparative'. The difficulty they have is that the terms of the comparative assessment cannot be pre-specified in any generally recognisable way. As Bird remarks, "For even if PR is conceived as an 'ideal' principle, if the ideal circumstances are not generally applicable the specific explanations of utterance interpretation might have to focus more on the divergent factors than on PR itself", *and they cannot be generally applicable if there is no means of identifying the extent of their influence*. The criticism is not that something like the PR does not hold. In other words, as suggested above, the *constitutive* cogency of the idea is strong, but exactly how this kind of cogency can issue in an operationalisable representation (or representations) is not at all clear, and so both its powers of explanation or its virtue as a model for design are as good as worthless, i.e., its regulative force is absent or scarcely more than sketched.

The law of gravity, then, is similar in this respect to a putative PR: that it is a constitutive rule, i.e., it is part of the meaning of the relationship between massive objects; and relevance is part of the meaningful relationship between components of a dialogue; but it is not *merely* constitutive with respect to some application. The difference is crucial: that there are practices which accompany the law of gravity linked to an understanding of how it operates, which allow us to instantiate the general rule, systematically. These means are absent in the case of the PR.

Descriptive and Normative Accounts, and Cognition

'Relevance', in spite of Sperber & Wilson's attraction to its biological and anthropological significance, has the flavour of 'meaningful relationship' (its mundane sense), just as Grice's cooperativeness is misleadingly connected with the idea of helpfulness. 'Relevance', in this sense, is at odds with an anthropologically grounded concept which detracts from the intentions of the agent. It is true that Sperber & Wilson also describe agents striving after norms, but it is not clear how they can, on the one hand, unavoidably follow the principle of relevance, and yet, on the other, decide how much they should follow it.

The pervasive difficulty, it seems, with comprehensive views of linguistics such as Relevance theory, or indeed of any cognitively based behaviour, is in drawing the line between the rules we, as agents, have to follow and the rules we make. Another way of putting this is to say that we seem, in order to have a comprehensive view, to have to adopt two different points of view and then accept that they must sit slightly awkwardly together. Sperber & Wilson clearly exhibit this ambivalence. That is not to say that these putative inconsistencies are entirely wrong. There appears to be a problem at the heart of grasping cognition generally, and communication as a particular and complicated instance of cognition. And I believe that this difficulty at least partly explains the confusion generated by the idea of relevance, which the authors themselves appear to acknowledge (Sperber & Wilson, 2nd edition, 1995, p266). They, however, believe that what they called, in the original publication, the principle of relevance (and which, in the 2nd edition is called the second principle of relevance, since the general cognitive property of relevance is now called the 1st principle of relevance) is a more precise expression of the idea of relevance. The problem that remains is twofold: we cannot assume that the second principle can be safely separated from the first, since they say that it " is grounded in the First Principle" (p263, 2nd ed.) and this means that some of that dangerous generality of the non-specifiable kind may continue to adhere; but, secondly, there is evidence that the problem of avoiding a conflation of the principle as a descriptive one with that of a normative one is still there.

The quandary appears to be that if Sperber & Wilson want to have a universal principle it may be vacuous as an explanatory rule, and yet if it is qualified by choice and intention on the part of the agents involved its continuity with such a general principle is undermined.

The result is either a 'post hoc' explanation "that there must have been such a balance" (of benefit and cost) "simply because, after all, the person did pay some attention to the phenomenon" or an 'ad hoc' explanation relative to the context or circumstances of the utterance, and independent of any principle. This 'pervasive difficulty' which I mentioned rests on a 'chicken and egg' conception, and corresponds to the relationship between context and utterance, which I examined before concentrating on Sperber & Wilson's notion of relevance.

Harris (1996), whose book is largely devoted to what others have found to be this 'pervasive difficulty', distinguishes between what he calls segregationist and integrationists in the field of communication. For example, with regard to context, he writes,

> "context, in short, for the integrationist, is the product of contextualisation. Signification and contextualisation are not two independent elements but facets of the same creative activity".

Unlike Sperber & Wilson, however, he is not tempted to provide a principle which accounts for the act of signification, and he ends his book 'in the middle' with an epilogue to chapter 15 and a suggestion that the reader "supply the remaining chapters of the book....Because everyone...needs a personal communication survival kit". The implication appears to be that as soon as we attempt to dismantle linguistic behaviour in the form of communication we seem only to be able to account for it inadequately (or in an ad hoc manner), or the account is general and principled but vacuous because, for example, "an assessment of contextual effect and processing effort...is difficult to determine operationally" (Bird, 1994).

What Sperber & Wilson identify with the general notion of relevance (the First principle) is, in my terminology, knowledge of a *constitutive* kind, and the misunderstandings which arise do so because they at least appear to treat it as if it were *regulative*, or as if the relationship between this general idea and some regulative view of relevance – in the form of the Second principle – is clear and straightforward. It is not my claim that some connection cannot be articulated between the two forms of knowledge; only that it has not happened in this case, and that an acknowledgement of these distinct forms of knowledge is required for this connection to be made properly. A more detailed examination of these forms of knowledge and their significance for the *use* of knowledge must await the development of the foundational framework.

Conclusions

What I have wanted to do is assimilate Relevance theory's difficulties to those general problems with representation which appear to attach to descriptive knowledge employed as a resource for design or evaluation. It has not been my principal aim to refute their theory. To do that properly, would require a very lengthy critique, because their book covers a multitude of difficult but central areas of interest. I have attempted to take the criticisms far enough to disturb the cogency of Sperber & Wilson's key arguments and use the consequent exposition as a vehicle for developing my own ideas within the confines of the thesis aims.

However, what, I believe, Sperber & Wilson are aiming for is a deeper (or more remote) role for Relevance. Aside from the confusions in their expression of this meaning, this deeper source is not graspable in the form in which it is presented by them for several reasons, some of which I have given above. I have tried to show that there are inconsistencies (in some cases quite blatant, and in others it has needed some analysis to uncover them), and it is difficult to know whether the inconsistencies can be ironed out, or if another unified theory of cognitive science could be erected in the place of Relevance theory.

My claim is not that no account can be given of these cognitive phenomena, only that the kind and status of such an account may be critical to its success. I want, therefore, to treat Sperber & Wilson's theory as an exemplary attempt to provide a quasi-scientific account, and to conclude that because of the difficulties it encounters it is not unreasonable to speculate that those difficulties might symptomise a radical problem with respect to *this* kind of theory about *those* sorts of phenomena. If I can suggest other ways in which some of the inconsistencies observed and commented on might be accommodated and if, moreover, these other ways derived from an alternative conceptual framework, then the claims of such a theory might be further weakened.

Let me summarise what I have concluded and suggest some ways forward. Firstly, the generality of a principle of relevance appears to be of a kind to which there is no exception, and that, unlike the generality of a law of nature where other forces are observed and measured in their opposing effects, there are none such observable and measurable effects where relevance is concerned. This is what makes it 'constitutive' and with no 'regulative' potential – potential which endows science with utility (at least as far as its experimental practices are applicable). This lack also impacts on its immediate value for design or engineering. Secondly, it is evident from the text of Sperber & Wilson's book that there is some equivocation over the 'mood' of the principle of relevance – whether it is descriptive or prescriptive, and this equivocation appears to be related to the extent that the principle operates as a force of (cognitive) nature or personal motivation or interest. It appears to be the content or area of study which results in this equivocal stance, i.e., human cognitive behaviour is essentially intentional and volitional. And thirdly, the way in which the details of the theory have been articulated seems to cause difficulties: the ambiguous perspective on Mutual Knowledge; the ambivalent attitude to the relative status of code and inferential theories of comprehension and the way they are supposed to interact (likewise, deduction and non-demonstrative inference).

The last chapter emphasised the importance of the idea of the 'domain' for (a) mediating the semantic and the pragmatic. In this chapter the domain can be employed again to reconcile difficult incompatibilities in Relevance theory, integrating the divergent themes of that project: for example, code theory versus the deeper forces of relevance, (plus deductive versus inferential process of reasoning). Without some such notion as that of the 'domain' the best that can be hoped for these themes is that they are juxtaposed in the overall theory. For example, perhaps the awkward pairing of deduction and 'non-demonstrable inference' might be best understood as part of modelling round domains of interest (or work).

Likewise, rather than reifying 'modules' and 'devices', it might be preferable to define the representations generally with respect to the domain: code and deduction are part of the semantic *representation* of that domain. Inferential processes and relevance are pragmatic representations which bear on the domain, but at the periphery. Neither is the more fundamental, but then this is not a question concerning 'real' entities, but is conceptualisation relative to function or purpose. And analogous with Sperber & Wilson's juxtaposition of code and inferential components, the P-BSA representation, on the one hand, can be interpreted as a coded version of the domain defined in terms of the tasks and goals, themselves defined by, for example, the requirements of some cognitive design problem, and, on the other, the underlying values, goals etc of the human agents could be cast in terms of pragmatic reasoning (Sperber & Wilson's inferential processes). Thus, the codes are the concern of the semantic characterisation, and the inferences, which are not simply deductive, are the concern of the pragmatic characterisation.

The difference between Clark & Carlson (1982) and Sperber & Wilson is, therefore, now easier to discern, and reconcile. Clark & Carlson's work attempts to disentangle the *semantics* of cooperative utterances, which arguably cannot be done without a given, the context; and Sperber & Wilson's project of trying to uncover the process of identifying and (simultaneously) using the context, via a grand and unified view of context and content, i.e., a *pragmatic* endeavour. Once we have a framework of the semantic and pragmatic components of cognition determined via a domain of concern, we have no need to raise the kind of questions of an epistemological nature such as mutual knowledge, nor thereby deny Clark et al's right to analyse the meaningful relationship of a dialogue to its ground, however potentially complex and dynamic that might be.

It may be that Sperber & Wilson's mistake is to assume that the route to proper knowledge of *cognition* can be that of science (or, at least, pure science); and as we have seen the 'domain' plays an essential role in resolution of certain difficulties in their theory. It is, however, not only the relationship of the particular and the general which is at issue; nor simply the need for constitutive and regulative components in the theoretical model. As can be seen repeatedly throughout Sperber & Wilson's book, another issue of central importance is the shifting representation of the descriptive and normative content of cognitive behaviour such as effective communication. The implication may be that such content cannot be characterised in scientific terms, but the reason why this may not be possible cannot be made clear at this stage, but must await the development of the foundational framework and the conclusion of the thesis.

CHAPTER 8

In this chapter I abstract the main components of the NLD framework and map them to the model of the conception of HCI as an engineering discipline (HCIe), and take those features of the HCIe conception of Dowell & Long (1989) as common to a cognitive engineering framework. In that context, I further illustrate the idea of the 'domain', as the solution to the issue of the demarcation of the semantic and the pragmatic, having exposed possible confusions in Relevance theory which were due to this unresolved problem. One of the other central components to which I allude is that of 'structure'. I try to show how this term, which was developed as part of the conception of HCI as an engineering discipline, may be extended in an epistemological direction, as I have tried to do with the term 'domain'. This chapter, therefore, stands at the midpoint of the thesis, looking back at the NLD framework and forward to the foundational framework, for which the terms will be further elaborated

I propose that this reassessment of HCIe is more consistent with a view of cognitive engineering as a unitary than a multi-disciplinary field of research.

NLD Framework and HCIe/Cognitive Engineering

Descriptive to Prescriptive Exemplar

I would like now to draw together what can be concluded from the constructive proposals, critiques and analyses of the project undertaken in Part 1, of which this is the last chapter. The aim was to arrive at a framework for NLD evaluation via a plausible exemplar of the transition from a scientific stance to one of engineering with respect to language.

I addressed the project by considering how language might be characterised. I set the scene in Chapter 1, firstly, introducing the root problem of evaluating (and, indeed, designing) so-called 'natural' artefacts, and sketching my views on design and evaluation in the context of research in the fields of cognitive science and HCI. I settled, in Chapter 2, on a descriptive/explanatory theory of language after drawing on the reflections of a notable contemporary philosopher, Dummett, concerned with that phenomenon – a thinker whose position appears to consolidate trends in the history of attitudes to language. I, then, examined an analysis of 'applied' versions of the more formal views of how language should be represented, and concluded that a realistic candidate for a descriptive/explanatory theory of language was Speech Act (SA) theory. It answered the requirements of the dual aspect of natural language: bearing thoughts and being a species of action.

The design project SUNDIAL had met difficulties in employing ideas such as SA theory, Grice's Cooperative Principle (CP) and Conversation Analysis (CA) for the purpose of evaluating the output of their work – effective *natural* dialogue. I decided to adopt that part of Searle's work (not directly inherited by linguistics) which had wider conceptual implications – his use of 'rules' – in order to develop criteria for identifying useful knowledge for design. I set these criteria in the context of a scale of the representation of the world which went from the purely descriptive to the purely prescriptive (the 'constitutive' to the 'regulative'). These criteria provided a possible account of why pure SA theory, Grice's CP etc were inadequate for design purposes.

Having identified the range of 'objects' for which representation was possible, and found wanting those representations associated with theories such as SA theory, Grice's CP etc.; and, further, having alighted on action as a convincing feature of language, partly to permit its role in a world of work (the effort of achieving goals, the aim of design being to do this effectively as possible), I suggested that planning as a general attribute of goal-oriented behaviour might be incorporated into SA theory (given its dual aspect of meaning *and* action). This incorporation would be according to Cohen & Perrault's Plan-based Speech Act (P-BSA) theory, and in order to remedy what pure SA theory lacked. The next step was to look closely at their theory in order to assess its completeness as a model for NLD which combined the descriptive/explanatory and the prescriptive aspects of language in the form, respectively, of SA theory and planning.

I concluded that Cohen & Perrault's planning schema needed revision, and suggested a way in which the illocutionary act might be conceived which maintained its connection with planning *and* served the purpose of encapsulating communication. Cohen & Perrault's theory appeared inadequate, and in fact their own view of it was as a limited picture of dialogue behaviour. I saw it as fulfilling the role of offering only an account of the *necessary conditions* of dialogue, and it was accordingly necessary to look for some further theory which would complete (and complement) something like a P-BSA theory. This theory might account for the elusive attribute of natural language behaviour – *coherence*. At the level of abstract reflection on language, SA theory itself was considered by some to be an inadequate basis for dialogue or, more generally, communication. Searle (1969) had adapted Grice's ideas but Grice considered this reformulation a distortion. Dascal (1994) considered that *neither* theory could be allowed to adapt the other without losing what was essential to each. My attempt at reconfiguring the illocutionary act might permit the fusing of the two, but I had already suggested that Grice's theory was 'constitutive', so I had to look further – for a theory with 'regulative' qualities appropriate for design.

I spent some time positioning P-BSA theory with respect to other theories in general terms. As I remarked above, Cohen & Perrault had seen their theory as of limited scope and of a particular type – a 'competence' theory. I examined what this characterisation might mean and how it would be related to other kinds of theories, concluding that, in terms wider than simply linguistics, theirs was a semantic theory with a strict scope of application, and that it needed a pragmatic theory to fulfil its intended purpose of providing a complete model for design purposes; and that this fusion of the semantic and the pragmatic would be enabled by the concept of domain.

Relevance theory was a general theory of the pragmatic type I was seeking. Indeed, the authors claimed that it was a refinement of Grice's theory which they had found flawed. It was my conclusion, however, that Sperber & Wilson's theory also suffered from a fundamental problem, highlighted by their stated aim of delivering a cognitive theory. This weakness provided the opportunity to show how the concepts of 'constitutive' and 'regulative' knowledge might again be helpful in allowing me to express, on this occasion, the inadequacy of Relevance theory. As with my critique of P-BSA, I recruited others to support and endorse my analysis.

The assumption that the SA theory skeleton of language had the potential to allow the translation from a descriptive/explanatory account of language to one potentially of a prescriptive/diagnostic nature has proved itself. However, I have had to modify some of the presumptions of how planning is integrated with speech acts.

Since it was not the intention of the thesis to provide a detailed model of language for design nor to offer a methodology for NLD evaluation the fact that the criticism has been largely negative should not be a problem. I have, however, suggested an alternative approach and hinted at ways in which it might solve other problems of design and its compatibility with more holistic accounts; enough, I hope, to justify maintaining planning at the framework level. In this chapter, my aim is to show the congruence of the NLD framework with one of Cognitive Engineering (CE) which underpins a conception of HCI as an engineering discipline (HCIe).

Framework: Models and Knowledge

The criticisms of the candidate theories plus reflection on the import of planning in the make-up of utterances have led me to some conclusions about the conditions pertaining to the evaluation and design of dialogue systems: the NLD framework. A framework is the context for a model or models. It should describe the conditions which support a model. In this case, the model is of the exemplar, i.e., SA theory plus planning to permit the specification of a domain of interest to design, i.e., a semantic theory, reinforced with pragmatic principles. The features of the framework have been exposed in the course of the arguments from Chapter 3 to the present chapter. They assume the correctness of the view of representation which I have adopted, including firstly the rules, 'constitutive' and 'regulative', and subsequently the corresponding kinds of knowledge relative to a given purpose. I have tried to show how these terms might be extended, beyond that of their purpose as a 'litmus test'.

However, in my examination of rules (Chapter 3), I tried to show how I differed from Searle in the manner of their application: I claimed that, though illocutions could be accounted for formally by constitutive rules, they derived their real force from values which transcended the 'institution' of which they formed a part. My view, then, is that it is important to distinguish between constitutive and regulative representations, but not so that they become isolated from one another. The acknowledgement of the necessary coexistence of the two distinct kinds of knowledge in the utterance or basic unit of language is a paradigm for their coexistence in the design model: where the domain of interest is specified relative to the requirements of, say, an artefact; but principles from the wider context of the world might be 'read in', signifying

> "the possibility that pragmatics can effect a radical simplification of semantics....By unburdening semantics of phenomena that are resistant to semantic treatment but tractable to pragmatic explanation, there is considerable hope that pragmatics can simplify semantic theories." (my parenthesis) (Levinson, 1983)

I shall continue to employ the terms 'semantic' and 'pragmatic' in the context of the model of NLD as I have developed those terms in the chapter before last. However, I want to point out that the terms 'constitutive' and 'regulative' are applicable to what is, respectively, internal and external to the domain. Previously they have been employed to discriminate knowledge with respect to design – thus, 'external' to design. Now, it is possible, once the discrimination is clearer, to employ them 'internal' to design; and I introduce them in this way to open a connection with the goal of Part 2, the remainder of the thesis: the foundational framework. The terms have the potential for a higher level application.

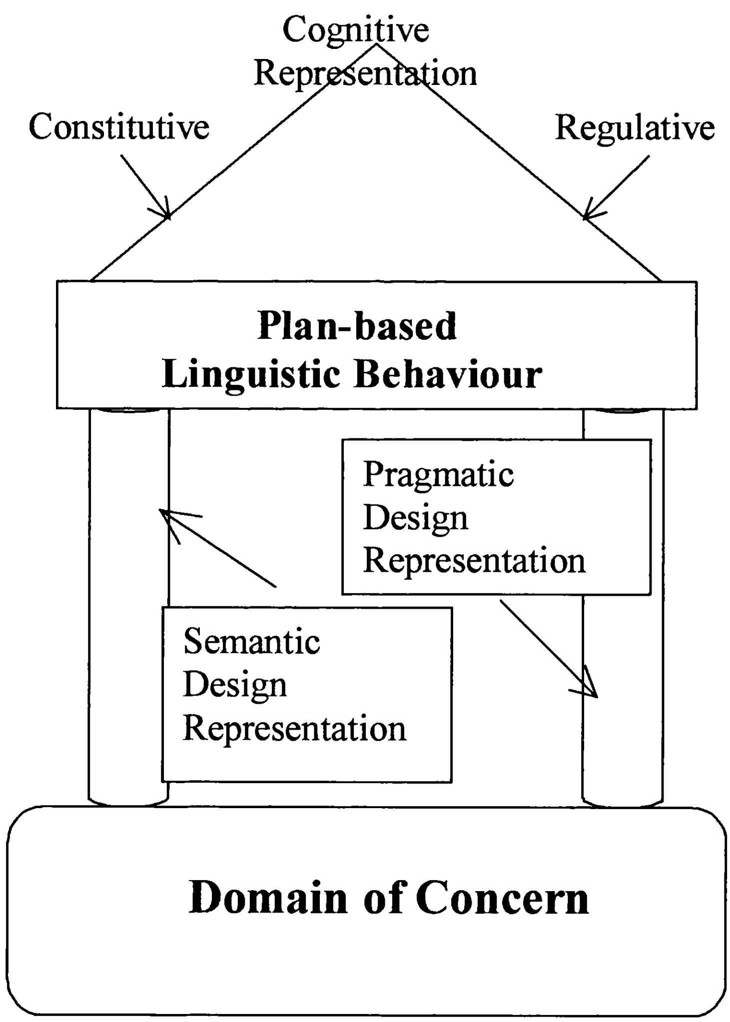


Figure 5: NLD Framework Model

This link between the way in which language does its job of representation in and of the world, and the manner in which language itself is represented as the object of design or evaluation is shown in Figure 5, which is represented as a building. We can think of the domain *supporting* the behaviour, and to a certain extent, the linguistic medium of the dialogue interaction. There are two principal 'pillars' (the semantic and the pragmatic) support the roof of dialogue behaviour, in its turn deriving from natural language representation (carried forward from Figure 4, in Chapter 3). These pillars stand, respectively, for the domain-content and the domain-context determining factors of meaning. The design problem specifies the relationship between the agents and their linguistic behaviour, on the one hand, and the domain of concern, on the other. The semantics is derived from the description and analysis of the goals, sub-goals etc. of the domain of concern; and, analogous with Levinson's 'reading-in' of pragmatic principles of a linguistic kind, would be the adoption of broad general principles which affect cognate domains where lawlike regularities are lacking.

Prior to the design specification is the characterisation of natural language representation. The most comprehensive view of language is as both the bearer of meaning and the effector of change, which is captured by the terms constitutive and regulative, as I have defined them. The dichotomy is continuous with that of the semantic and pragmatic and there is a dynamic interaction from the peak of the roof, representation, to the domain of concern.

In the chapter before last, I argued that language theory benefited from the introduction of the notion of domain, and so, in a sense, the figurative depiction of the design problem is only a formal version of the natural state of affairs. This way of putting it shows the intimate conceptual connection between the descriptive and the prescriptive state of affairs. Though intimate, they are, however, not identical.

Framework; and Planning

I have dissociated planning from intention narrowly interpreted, and it is true that problems arise such as those pointed out by Suchman (1987); problems which she tried to encapsulate by asserting that plans were not the 'generative mechanism' of action. It might seem that I have come round to her way of thinking, but the distinction which I have been attempting to develop in the course of my criticism of the P-BSA theory has been tempered by the introduction of the concept of the domain.

Suchman does not run the risk which Sperber & Wilson face because she does not believe in the propriety of a complete description/explanation of cognition, and it has been (relatively) easier to demonstrate the domain as a solution to the alleged problems in Sperber & Wilson's work. Suchman takes the extreme pragmatic, holistic and emergent perspective and, quite consistently, does not attempt to 'apply' it. She may not be wrong except when arguing that others may *not* adopt a representation of cognitive behaviour to solve cognitive design problems. The concept of the domain allows one to agree with her about cognition in general *and* address cognition in particular. As was implicit in one of my defences of planning, part of the problem seems to stem from her assumption of planning as a global specification applied to the world in all its potential detail. As I have been emphasising, however, any 'application' is limited, by the nature of design, to domains.

It was clear that the Cohen & Perrault view of planning closely corresponded to that of intention – even conscious intention, and I contrasted their view with that of Power (1979) whose starting point for the protagonists of his dialogue was to contrive them as copies of one program, and then individuate them by changing key variables. In Power's view the process of planning was a private procedure and conversational procedures were consequent on, and called by the planning procedures. In Cohen & Perrault's case, the plans were incorporated in the speech act syntax.

It is an essential feature, therefore, of the NLD framework that the concept of planning as a specification tool is not limited to intention, understood narrowly, but covers any accessible goal which might be discerned with respect to the domain of interest, and governed by more general principles. As I shall develop the argument in this chapter, it should be noted that this characterisation of the plans supporting the behaviour in the Interactive Work System (IWS)[59] (Dowell & Long, 1989) means that it can correspond with a deeper understanding of the work (which is the target domain of a design project), its goals, and the sub-tasks and their goals.

The loosening of the notion of planning which I tried to effect in the chapter on the examination of the P-BSA theory's use of it should, therefore, make way for these general goals and values to account for meaning globally, i.e., expressed as common concerns; and understanding the design or evaluation process will improve with a greater link between these global concerns and the local ones within the domain of work, as expressed by task goals.

Planning and Convention

In Chapter 5, what I have referred to above as the "loosening of the notion of planning" calls for a reformulation of the speech act schema. I argued, alluding to Ramsay (1990) (whose work undermines Cohen & Perrault's view of the utterance's syntax determined by a planning schema), that planning in relation to speech acts was to be distinguished from commonly accepted views of it as a causal tripartite event. This reformulation was intended to maintain the view of the speech act when mediating communication as involving 'convention' for its success. It was intended that this use of 'convention' should be compatible both with Searle's and Grice's attitude. Grice would not have accepted the sense of 'convention' as an arbitrary relationship between tokens and their referents, since for him intention[60] and its reciprocal understanding[61] were fundamental to communication. At least one solution to the problem drawn attention to by Ramsay and by the 'oddity' of 'side effects' noticed by Cohen & Perrault was to accept, *not* that there was no effect of the speech act – which seems manifestly unreasonable – but that the preconditions form part of

---

[59] The behavioural system of the cognitive agents doing work in a domain.

[60] Grice's view might be expressed as giving priority to what *people mean*, rather than what their sayings (and writings) *mean* as a function of their given meaning.

[61] In Chapter 6, I cited Hornsby's definition of reciprocity, comparing it with Grice's version. Habermas writes, "Thus the illocutionary force of an acceptable speech act consists in the fact that it can move a hearer to rely on the speech-act-typical commitments of the speaker". (Habermas, 1979)

the act. The 'illocutionary uptake' by the hearer of the warning that the bull is coming is not some simple coded relationship which is agreed by custom or formal agreement. It is, however, *conventionally* meaningful because we are, most of us, afraid of bulls charging. And it is 'conventional' because it is a *consensus* built into us, born of basic urges and experience of a direct or indirect nature.

> "Whether or not they have an explicitly linguistic form, communicative actions are related to a context of action norms and values. Without the normative background of routines, roles, forms of life – in short, *conventions* – the individual action would remain indeterminate. All communicative actions satisfy or violate normative expectations or conventions. (my italics) (Habermas,1979)

This relationship between those "routines, roles, forms of life", which are intimately tied up with values and goals also provides one general answer (at the framework level) to what is referred to as the 'integration problem' (Novick, 1988), i.e., bringing together the sense of a particular speech act and its immediate context with that of the global situation.

Habermas (1979) near the end of his paper, "What is Universal Pragmatics", which is concerned largely with an interpretation and extension of Austin's speech act work, gives some "provisional results", which he summarises in his last section. Given that the "speaker and hearer can reciprocally motivate one another to recognise validity claims", these claims are (a) "truth for a stated propositional content or for the existential presuppositions of a mentioned propositional content", (b) "rightness (or appropriateness) for norms (or values), which in a given context, justify an interpersonal relation that is to be performatively established", and, finally, (c) "truthfulness for the intentions expressed". These claims create obligations, Habermas says. Without being more specific in the application of these conditions for specifying or describing dialogue, it is clear that there is a context of normative constraints, both motivating action via the "routines, roles, forms of life" referred to above[62], and providing warranty for the 'contractual' ground of any dialogue – (a), (b) and (c) above.

These goals are only, in small part, the sort which would count as purposes, or intentional goals. Which is to say that agents are, in a sense, used by as well as use, what I have called the context to drive the dialogue. Thus, for the purposes of a framework for NLD, we are extending the idea of plans to cover *all retrievable teleological schemata* including the desire for pleasure, freedom from pain, consumption etc. which fall under the above human, cultural, personal etc. values, *in relation to* some project or other, i.e., work – the purpose of the design requirements being to carry out this work effectively. The simple infrastructure of the P-BSA is too 'near-sighted'; and the general assumption of goals as intentions or personal aims in pragmatic frameworks such as Haslett's[63] and McCarthy & Monk's is wrong for the same reason. The schemata which I have in mind are instead the implicit features of Sperber & Wilson's theory but their account is, I believe, inexpressible as a scientific endeavour.

Consistent with a broader view of planning (and of intention), which includes social and cultural as well as individual values and goals, my conclusion was that, as between a plan-based view of speech acts of the kind offered by Cohen & Perrault and one which embodied the plans in the 'life' projects of the agents, of common stock (copies of the same program) which is suggested by Power's (1979) work, a reasonable view was to concur with the latter. I have found then that Habermas, a writer whose work has been applauded by Winograd & Flores (1986), provides a general schema for speech acts, within which the revised view of planning might flourish. But the NLD framework cannot be considered complete until it is re-united with a conception of cognitive engineering.

---

[62] "After Schutz, Heritage (1984) suggests that individuals can know one another in four distinct ways: as human beings, as members of the same culture, as specific persons, and as specific persons now-in-this-immediate-situation" (Haslett, 1987)

[63] Haslett (1987) writes, "The only goal speakers and hearers *must* share is that of cooperating so that interaction may take place." (my italics)

HCIe, NLD and Cognitive Engineering

Following my general description of the NLD framework, I want to show how it might be seen as a transform of the engineering conception of HCI (HCIe). For the time being, my discussion will be in terms of HCIe (IWS = Interactive Work System; Domain = Domain of Work), and the mapping can be seen graphically in Figure 6.

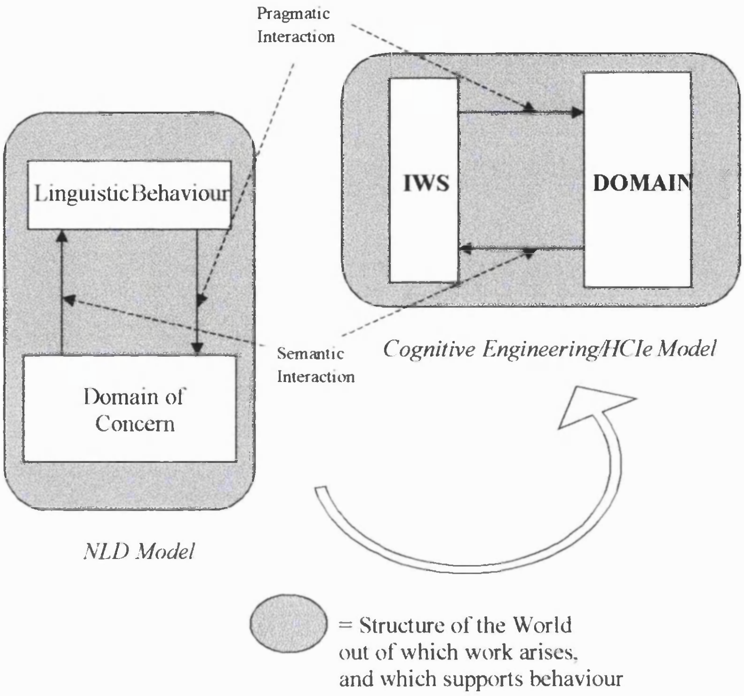


Figure 6: NLD to HCIe Mapping

The essential components are the same for cognitive engineering. The terms referred to in this illustration will be amplified, and their inter-relationship discussed, in the course of this chapter.

In order to complete the integration of the NLD framework and HCIe, I have to revisit some of my initial concerns about the adequacy of the HCIe model to accommodate the design and evaluation of NLD, and then to illustrate how the elements of my framework for NLD design and evaluation can be adapted easily to the existing properties of the HCIe framework if those properties are provided with a rationale for reinterpretation .

Human Computer Interaction as an Engineering Discipline (HCIe)

The HCIe conception is a high-level description of the components general to engineering design solutions in HCI. It was developed by Dowell & Long (1989) to provide a framework for the Human Factors (HF) side of HCI, complementing the Software Engineering (SE) side. In its more general form it is also a model for Cognitive Engineering (CE), and was employed as such by Dowell (1993).

Long & Dowell (1989) and Dowell & Long (1989) have addressed the issue of HCI as a discipline and detailed its constituents and practices. Their approach has been to describe and relate the components and concepts of the General Design Problem of HCI (its ontology) and propose an engineering means of solving design problems using those components and concepts. My project has addressed the epistemology of design in the area of NLD, in particular the relationship between descriptive, explanatory linguistics and systematic design knowledge of language. After the development of certain ideas over the last few chapters it may be possible to unite some of those ideas with some key concepts of HCIe. It is part of my intention, in this thesis, to reinforce the HCIe and CE projects, which I believe suffer from two principal and related flaws.

Firstly, Dowell & Long (1989), Long & Dowell (1989) and Dowell (1993) see the science and technology (in particular, engineering) as subject to separate development. They took this stance to get away from the approximate and uncertain use of psychology to answer design problems. I have suggested that they have gone too far and this poses problems for understanding how *as a matter of fact* most design is reasonably successful. The second and related difficulty is the problem they have of operating with engineering knowledge as an *autonomous* discipline – epistemically self-sufficient.

An argument for both of these difficulties will be developed in the wider context of pure science and technology, but I need here to address them roughly in the setting of cognitive science and cognitive engineering. (I am taking the HCIe model as adequate to CE, as Dowell & Long (1998) do, and the NLD model as representative of a successful transition from cognitive science to CE.) I shall broach the second first because I can deal with it more pointedly, and then conclude with a brief comment on the consequences for the first, concentrating more on this issue in the context of the foundational framework.

Because Dowell and Long individually and severally promote the view that engineering knowledge is to be derived only from engineering practices[64] this autonomy poses the danger of epistemological impotence. Their approach is to work outwards from the solution of particular problems to that of more general ones. The question is, On what basis can such a generalising move be made? In cognitive science, hypotheses are made which are of a general nature, in the first place, and ideally particular well-designed experiments test these general hypotheses. So a comparable problem does not exist for cognitive *science*. It is difficult to see how, without further conceptual qualification, generalisation can take place unless against a background of scientific generality, and, therefore, the 'strict' practice of CE advocated by Dowell and Long is suspect. If a foundational framework can retain CE's crucial independence, without the necessity of generalising from the particular design solution, then it will no longer be necessary to take this 'strict' view, and the problem is avoided.

Apart from my claim that the latter approach is epistemologically correct, it is from a practical point of view more appropriate with applications which are (at the moment) immeasurably complex, such as linguistic behaviour. This more general argument about the communicative relationship of science and technology is only mentioned as the solution without being argued for, but the doubts about the 'strict' autonomy of engineering knowledge lends cogency to its potential.

[64] Both Long and Dowell accept that there is an interaction between science and engineering disciplines, but it is not a specified interaction (see also Long (1996)). At a certain level of description I, however, am offering a specification.

Domain

The mechanism which permits the move from the particular to the general, and vice versa, is that component of the ontology which Dowell & Long use for defining the work of agents and their relation to it. I am extending its use, while retaining their sense of it. For me, its boundaries are defined with respect to the work, on the one hand, and the cognitive dispositions of the agents, on the other, but also with respect to general knowledge. The particularity of the domain is what poses the design problem. It is what distinguishes knowledge of it from general knowledge, but it is reasonable to assume that between them there is a systematic connection; and, as I have argued in the case of NLD (probably more widely in cognitive science) the introduction of planning to knowledge of a constitutive nature may permit one solution of a systematic kind. In general terms, there are regulative rules which must be discovered which are peculiar to the particular problem. This account is, of course, a 'description', using the extended concept of the domain, of what the designer does. It is no more a 'method' than Dowell & Long's use of it. It is a brief account, in HCIe terms, which obviates an epistemological problem while reinforcing the conception. The argument, however, needs to be elaborated and set in the context of view of science and technology – something which the investigations of Part 2 will tackle.

It is my intention, therefore, to assert that the domain (i.e., the domain of application) mediates not only between the semantic and the pragmatic, more generally than with respect to NLD, or, indeed, linguistics, but also between the relevant general knowledge (that provided by, for example, cognitive science), on the one hand, and the particular knowledge required to solve a design problem, on the other.

Structure

In Dowell & Long (1989) and Dowell (1993), 'structure' is postulated to *support* behaviour, in order to complete the ontology of the high-level model of the design problem. It is, therefore, the assumed reality which underlies the observable behaviour. Interaction, they state, is not structural but only behavioural. Their view that the structures of the agent and the device (taken to mean entities which are physical and abstract) do not interact – only the behaviours do, cannot imply that they do not directly affect one another (viz. Ashby's example of cars crashing in Dowell (1993)). It is equally indisputable that a person's disposition to some piece of behaviour can be degraded if not destroyed through interaction, as Dowell & Long note.

So what can be meant by this hermetic division of the structures of the device and user, both sealed off from each other and from the IWS? *I believe that this requirement can be seen as a constraint which reflects the necessity of positing stability with respect to both the IWS and the domain (and further, the rest of the world), and within the IWS between the cognitive agents.* It is a means of stating unequivocally that there is a ground in both the cognitive and the material realms, without which 'context' vis à vis the cognitive, and the 'situation' vis à vis the mental and physical objects, respectively, will play an unpredictable role; and that, further, they must be kept distinct otherwise, and for the same reasons, the force of a change of state will be indeterminate. That is to say, structures conceptualise that which underpins prescriptive and descriptive characterisations of behaviour. Thus, any 'coupling' within such a system would be behavioural but *founded on* an underlying structure. The structures need not represent 'real' physical things (they will encompass those) which are unaffected by events taking place within the IWS, but they, qua such structures, must be allowed their roles independent of changes within the IWS and the domain. It does not follow that anything which is designated a structure will always remain one. They will have to be reconceived from time to time.

Domain and Structure

Can we bring the concepts of structure and domain together to enable a better expression of the relationship between the particular and the general? As can be seen from Figure 6, I have stated that it is structure which underlies both the IWS and the domain. But we have only seen above how it might be used in the context of an account of IWS behaviour – a background against which agents could be individuated and their behaviour understood. However, something akin is required of the domain objects. I shall only sketch how it can be achieved, since as I have said it will be dealt with more fully later.

I want first to introduce and adapt another key concept from the HCIe framework (Dowell & Long, 1989): that of 'affordance',

> "Conception of the domain then, is of objects, characterized by their attributes, and exhibiting an affordance arising from the potential changes of state of those attributes."

Elsewhere they write, "A domain of application may be conceptualized as: 'a class of affordance of a class of *objects*'" (my italics).

Firstly, I think there is a problem with the emphasis on the object. It is true of an object that it can be conceived as potentially changeable, but I believe that it is so only as a consequence of possible changes in 'states of affairs'. It is into states of affairs that the domain is more appropriately divided because of its task structure, and this division makes the state of affairs more aptly conceived as the object of plans. The domain then, as the semantic realm, is that set of states of affairs which is meaningful because it is composed of these affordances, where the affordances can be systematically[65] described. There are also pragmatic forces which determine affordance but which derive from outside the domain and whose impact is less predictable. The interface between the semantic and the pragmatic determinants of affordance is the boundary of the domain: this is the converse expression of the domain as the mediator between the semantic and the pragmatic which I argued for in Chapter 6.

Thus the structures, in the IWS, are those stable entities which bear dispositional characteristics, interact behaviourally, and are coordinate (owing to the mapping of plans and tasks) over against those things in the world which are stable but have their own version of dispositional attributes – their 'affordances'[66]. Therefore, in a fashion corresponding to the argument about structure and the IWS, I am suggesting that

[65] 'Systematic' representation would encompass component or feature based and truth-conditional kinds as well as something like that which is specified by rules for what Austin calls 'felicity' – of natural language (or dialogue behaviour of a more general kind if we are considering an HCIe or CE model).

[66] "If the worksystem is well adapted to its domain, it will reflect the goals, regularities and complexities in the domain." (Dowell & Long, 1998)

what underlies the variations which may be represented in domains, illustrated above, is what can be thought of as structure. Particular domains are like circles or ellipses of light cast by a lamp at various angles and showing different views depending on where and how the lamp is positioned. The structure is the continuous background which is highlighted here and there. Like the function of structure with respect to the IWS it serves as a ground for generality, continuity and reliability. Because HCIe, with the help of such concepts as domain and structure, can be re-integrated with general knowledge and cognitive scientific knowledge, it has greater power to reject subjective and irrational approaches which might arise from epistemic isolation. Likewise it can argue for knowledge with a greater guarantee, but the fuller picture awaits the development of Part 2.

Finally, the IWS and the domain can be definite in their distinction not only between each other but also between themselves and the background against which they come into focus as the design problem is articulated. Just as the domain boundary interface is defined by the relative coverage of precise states of affairs and their connections with the wider context, so the IWS is defined by the specific representation which corresponds with specifiable tasks congruent with the affordances of the relevant states of affairs and these representations connection with more implicit ones: the less specifiable but related tasks will correspond with the less specific representation and their dependence on more remote or high-level goals – cultural values such as loyalty to family and national groupings or more remote and instinctive goals such as self-preservation. There is, therefore, a pragmatic penumbra around the IWS just as there is around the domain. The centrality of the domain rests on the connection with a shared reality.

In conclusion, apart from the arguments for the correctness of the move from science to engineering, which I shall develop in the following chapters – arguments illustrated by means of the constitutive/regulative classification, my aim is to employ the concept of the 'domain' as described in the HCIe conception as the element which permits two key moves: firstly, enabling the generalisations of scientific knowledge, and, pertinent to the purposes of this chapter, secondly, the relationship between the semantics of an application and its governing pragmatics. These are, as it were, coordinate moves on separate but related levels of the framework. The reasoning for them takes place in the foundational work of the NLD edifice. I have discussed the second move in the context of NLD. An analogous interaction occurs via the domain in HCIe and cognitive engineering, more generally. The roles the domain fills will be better understood with the perspective provided by the foundational framework. The first move is making the point, as Long & Dowell (1989) do when they consider Hammond's transforming predictive propositions into guidelines, that it may be prescriptive in some general sense but cannot have a bearing on design, because it does not take into account the particular design problems in question. However, my emphasis is different, as I believe that the crucial difference is not that the design knowledge must be "conceptualised, operationalised, tested and *generalised* with respect to the design for effective performance", since this approach commits us to working from the bottom up and having to face the obstacle of rationally justifying moving from the particular to the universal – a notoriously insurmountable problem for radical empiricists. My position is rather one of arguing for a systematic move from the scientific (general by its nature) to particular applications whether the job is predictive or prescriptive[67], and that the means for doing so systematically is the specification of the domain.

Is Cognitive Engineering Multi- or Uni-disciplinary?

The framework has other than the technical/operational implications of a common epistemological basis of cognitive science (in the form of linguistic) and cognitive engineering (in the form of NLD design). It is also intended to provide a better rationale for the cogency of the idea of cognitive engineering as one discipline, *not a descriptive method with a multi-disciplinary input*. I want to claim that the multidisciplinary approach is both weak and unnecessary: weak because it inevitably generates demarcation disputes which are the result of arguments over whether a particular method, for example, is being employed 'correctly'; unnecessary because it can be argued, I claim, (a) that 'rational' design is not at odds with, for example, the more metaphysical presuppositions of approaches such as Conversational Analysis

[67] Of course, if there is no appropriate scientific model one might assume one. The foundational framework simply supports the move – real or ideal.

(the negative defence), and (b) that scientific or systematic descriptive knowledge has a demonstrably common root with engineering or prescriptive knowledge (the positive defence). This link between science and technology allows the adaptation of descriptive/explanatory knowledge, but engineering criteria (i.e., those of a *radically* distinct type of discipline from science) ensure its unity.

McCarthy & Monk (1994), on the one hand, and Haslett (1987), on the other, both promote the idea of integrating the work of various schools by attempting to provide a framework to accommodate them all. However, neither can deny that these schools do not, in the main, want to collaborate as they have different agendas. As with Levinson's (1983) attempts to bring conflicting or competing strands of linguistics together, these projects depend ultimately for any real success on the protagonists' charitable cooperation. Neither McCarthy & Monk nor Haslett provide an alternative schema which will both accommodate *and provide a rationale* for the integration; the key to achieving that integration is the recognition that what we are engaged in, with NLD as the example – but more generally cognition, is *engineering*.

In their paper, McCarthy & Monk attempt to come to terms with the diverse approaches to the understanding of communication offered by engineers, psychologists, social scientists, linguists and philosophers. They write, "it is very difficult to get any kind of integrated picture". Their point of view is a scientific one. They recognise that although HCI workers can employ any ideas they choose, this licence will be "most upsetting to the disciplinary purists" because "much of the methodological background to these insights will be lost". However, the problem is not simply a difference of methodology. For the "discipline purists" it is a question of truth and reality, and unless their proposed "common psycho-linguistic framework" is going to reconcile these 'purists' on the conceptual level as well, the 'purists' will certainly resist the blandishments of any such framework as is proposed by McCarthy & Monk. This is the same problem facing Levinson and Haslett in their different ways. However, Levinson, as a linguistics expert, wants to make some telling points on the way to his view of how he thinks language should be *understood.* Levinson is in the business of competing with, or feels himself competent to, improve these views which undergo his critique: but his project, as well as those which undergo his critique are descriptive and explanatory projects. Consequently, there is consistency of approach.

This is not the case with either McCarthy & Monk or Haslett. These authors are concerned to comprehend descriptive and explanatory accounts within a *design* framework If the distinction between science and engineering is not taken account of, the outcome will be unclear, and ultimately ineffectual. But without some conceptual basis for integrating different kinds of knowledge and practices, it is difficult to see how their aim could be achieved. As soon as we can distinguish the two activities, and mark the distinction conceptually, we can separate what we need into an explicitly engineering pool; the knowledge employed by the cognitive engineering discipline is defined and unified by the project of systematic design. The unity of the knowledge is guaranteed by the foundational framework and the identifiably distinct activity of engineering. What we need then is a way of re-expressing what is scientific knowledge as engineering knowledge. I have attempted to provide this re-expression for NLD, and I shall devote the rest of the thesis to providing the more general framework for science and engineering.

Likewise, liberal linguists who take a catholic or eclectic view of discourse and linguistic mechanisms or structures, while they provide a rationale or a choice of rationales, do not answer the need for doing other than 'pick'n mix' design.

> "Both broadly focused and narrowly focused analyses of communication are needed: I have chosen to do the former, while many scholars choose to do the latter. I believe there is an imbalance between these two alternatives; too much attention has been directed at fully developing views on a particular aspect of communication, and not enough attention paid to how various aspects of communication fit together in the process of humans communicating." (from the preface of Haslett (1987)).

Haslett's practical approach ("Through this book, I wish to present a general view of how verbal communication works...") also has the weakness that it lets in the multidisciplinary scramble for control; and this 'control' is of a scientific kind. We ought to adopt an approach, in the context of design issues, which is principled and apt for the purpose. If we start by taking the view that knowledge is for a purpose, and this purpose is design, then the purpose should be the guiding light showing us the limits of our methods and representational schemata. If we want to engineer, then we have to start with firm foundations.

Since my view is that engineering is not derivative of science, but that they are aspects of the same knowledge (i.e., they are representationally inextricable), then in the face of a descriptive body of knowledge which does not exhibit the uniformity of scientific description (disciplines contributing to the understanding of language, for instance), my job is to argue for uniformity through its 'application' in design, rather than from first principles. Conversely, and since I hold this view of the inextricability of engineering and science, through engineering research of linguistic system design we may arrive at an integrated conception of language, obviating some of the uneven knowledge in the area. In any event, the uniformity of science derives more from its method than from any demonstrable common features that individual scientific assertions possess. Likewise, engineering knowledge of language or communication may derive from the methods and purposes of engineering, not from isolated solutions to engineering problems or anything they might have in common. Thus, it seems that although Haslett is clearly interested in practical matters, probably involving design, her stance is essentially a scientific one, as is that of McCarthy & Monk. The value of their surveys and critical comments, which *might* be considerable, can only be of craft (rather then explicit design) value. To say this is not to demean them but only to point out that, viewed in the long term, their presuppositions about scientific and design knowledge could be counter-productive. If disciplines can be distinguished *but* channels of communication maintained with its related science then CE or HCI could be uni- rather than multi-disciplinary and defined by its function or overall aim as engineering. CE and HCI would then be sub-disciplines of engineering; just as the various branches of science are sub-disciplines of the discipline of science.

Conclusions of Part 1

In Part 1, then, I began by questioning the difficulty of treating the design of 'natural language' if we were constrained to deal with it as a pure design problem, i.e., without having recourse to knowledge of a descriptive and explanatory character. This address would be the perspective of a strict engineering approach, sceptical of the value of scientific results of language study. I believed this scepticism to be appropriate, but there is clearly some traffic between science and technology (including engineering). In order to understand on what basis this traffic could rest I decided to consider the means of making scientific (linguistic) knowledge usable for the design of an NLD project. This was the 'plausible' exemplar mentioned in the thesis abstract. It is only offered as such because it is relatively specific, and doubts might remain about whether, for example, linguistics is science or NLD design is engineering. A more comprehensive argument is needed and it will lend weight and cogency to that of the exemplar and its surrounding arguments.

The framework which results, as I have tried to show, is orthogonal with the model of General Design Problem (GDP) of HCIe, as depicted by Dowell & Long (1989), and I have adapted some of the terms of HCIe such as 'domain' and 'structure' to be consistent with the terms of the NLD framework showing how they might play an epistemological role. The broader based argument of the remainder of the thesis will make the application of these terms yet more general, and the outcome will also reinforce the case for CE as proposed by Dowell & Long (1989) , Long & Dowell (1989) and Dowell (1993).

I discussed an important implication of what I think is a re-integration of CE with cognitive science and general knowledge: that there can be one discipline of CE, which is inclusive but which really imposes only one directive – to make design knowledge as explicit as possible. I have tried to suggest that the NLD framework and more generally the adaptation I have made to the CE framework means that any knowledge from the most formal and rigorous to the least explicit and precise is admissible; that this continuum is accepted with the only proviso that we must attempt to make a difference between its extremes, and that this difference should lie in the specification, insofar as it is possible, of the domain.

The first step has, therefore, been taken towards a reintegration of the grounds for both the scientific and the engineering disciplines with this 'plausible' exemplar. The second part of the thesis will broaden the argument as described, and take on some of the familiar general objections to the assimilation of science and engineering as well as to what I argue are mistaken attempts to conflate the two; which will result, I shall claim, in a better balanced view of their relationship with each other and what their common denominators are, while maintaining their distinguishing features.

PART 2

"Truth therefore and utility are here the very same things[68], and works themselves are of greater value as pledges of truth than as contributing to the comforts of life."

(Francis Bacon, quoted by Hacking (1983))

CHAPTER 9

Since this thesis emphasises the epistemological rather than the ontological features of any discipline of cognitive engineering, and, since one of its particular aims is to understand better the relationship between science and engineering in order to improve the support of evaluation and design, this chapter continues the generalisation from NLD to HCI and cognitive engineering with an enlargement of the idea of the domain – to satisfy those ends. I attempt to characterise the kind of enquiry I have undertaken, and distinguish it from that of Dowell & Long's (1989) conception, and the chapter will serve as the ground for the development of the foundational framework. The character of the domain, as I have and will develop it, is intended to encompass and elaborate Dowell & Long's concept in order to fashion it for more general epistemological employment. I show, for example, how it might be used to separate out discipline features as empirical or meta-theoretical, and hint how it may illuminate the relationship of theory to practice (to be examined further). In the latter portion of the chapter, I try to demonstrate the concept's broad utility by illustrating the more mundane epistemological role it plays in determining the question of the ownership of knowledge, trying to improve on at least one attempt to find a criterion for identifying technological and scientific products.

Domain

Introduction

In the last chapter I attempted to make more precise what I meant by domain through a critical examination of the HCIe framework or 'conception'. I had developed a view of it in the course of reaching a framework for NLD, and this critical examination of HCIe, including adaptation of some of its principal components, allowed me to bring it into focus in a very particular but appropriate setting in order to complete the task of Part 1 and bring it home to CE.

It is time now to broaden its application so that I can construct a general foundational framework whose purpose is to give cogency to the more specific NLD framework which I have referred to as 'plausible' because it as yet lacks this general *conceptual* validation. I shall start by trying to locate the beginnings of the idea of the domain, a novel concept as used in Dowell & Long's (1989) and Dowell's (1993) HCIe framework, raising some questions straightaway, which I have answered in part in the

---

[68]Rossi (1996) doubts this translation by Spedding et al (1857-1858): his alternative disputes the claimed *identity* of truth and utility but does not disagree with the import of the *convergence* of truth and utility.

context of the last chapter. I shall then proceed to outline the differences in my approach. How another similar approach of a famous thinker might endorse the basic quality of the domain as an idea.

I have alluded to the domain's role in the science-type role of linguistics. I want to widen this role, and also link the idea to broad categories of knowledge including science, and with particular reference to its conceptual properties, I would like to give an instance of how it might resolve a metaphysical controversy at the same time as ruling out another common allegation of the limits of a discipline such as CE. I shall then sum up with reference to the beginning of the chapter's critique.

Finally, I present a illustration of how the idea of the domain, while still concerned with the epistemic distinctions associated with the general argument, might help focus attention on a practical problem of policy administration: whether and how basic research and technology research might be distinguished. The term's use in a wider currency might lend it some authority in its more abstract role.

Origins

In John Dowell (1995), the author makes, among others, the following comments on the origins and properties of domains:

1 The "idea of a domain" is related to the 'ecological' view of behaviour (Dowell is referring here to, e.g., Gibson (1977)), in which the human is understood to be inseparably coupled to the environment: "it makes no sense to study one in isolation of the other" (Dowell, 1995).

2 "More accurately, the ecological view sees the environment as many different environments or *domains*, each of whose affordances and constraints are organised around specific goals. (Dowell, 1995)"

3 Dowell draws attention to Herbert Simon's reflections on an ant's behaviour as a reflection of the environment's complexity rather than its own: inner and outer

consonance is the fulfilment of the ant's 'design', but Dowell infers that the understanding should be of the domain, "minimising our assumptions about the inner environment of the system" (Dowell, 1995).

Some implications of the above (and corresponding with 1,2 & 3) are:

(i) The idea of the domain derives the plausibility for its adoption as a psychological view. That is to say, although Dowell goes on to justify its utility in terms of its ability to elucidate and support the design process, he does not address adequately the analytic grounds on which this plausibility is founded.

(ii) It raises the question of how these many domains are related.

(iii) That the appeal of Simon's ant example renders it plausible to dispense with the inner environment of other more complex cognitive beings. Dowell (1993), however, himself writes, "The vexing experience of watching controllers at work is the small repertoire of overt behaviours by which controllers express their very rich cognitive behaviours." Presumably these complex behaviours are not a mere by-product of the domain and, indeed, there is some incompatibility with Dowell's first comment, since "coupling" assumes criteria of 'fit' along with cognate 'structures' in the domain and the Interactive Work System (IWS), or an agent within the IWS. This fit must be achieved between the inner "rich cognitive behaviour" and the domain. Analysing the domain will not stand in for understanding, for example, linguistic behaviour.

It is worth noting also that the notion of the domain is meaningful in the investigation of areas other than those concerned with design. Dowell's appeal to Gibson evidently involves a recognition that the domain plays a role in behaviour that is being *described* or *explained*, as well as figuring in a high-level characterisation of the framework for a discipline of *design*, which is what he goes on to give an account of in the paper.

I would like to try to address some of these observations and implications in the course of this chapter.

Domain and HCIe

Dowell & Long (1989)[69] write that engineering disciplines require an epistemological enquiry, and they refer to this enquiry by reference to a paper of Van Gisch & Pipino's, in which the latter authors conceive of the paradigm as the product of such an enquiry. In Dowell & Long's case, they aim to produce the design equivalent of a paradigm for science – the HCIe conception, and they note that like a paradigm it is "open to rejection and replacement". Given that one of the project's aims is to provide guarantees that go with the knowledge of the new discipline, this defeasibility would, I think, be, at least prima facie, at odds with one of the fundamental goals of the HCIe project. Indeed, in the section of the previous chapter, in which I recount my version of the role that the structures play in the conception, I emphasised my belief that a structure need not stand for something 'real', in any absolute way; but that they, in general, serve as an explanatory background against which the behaviour can be understood, with the same sense as the following:

> "Whenever we are concerned with explaining some regularity, as opposed to merely describing it, it is essential that we view the regularity against a background of what might have occurred given different circumstances", (Pylyshyn, 1991a).

Engineering employs the notion of 'parameter variation' as a systematic method (see for example Vincenti (1990)), which only makes sense against some background reality. It is my view, then, that it is any *particular* structure "which is open to rejection and replacement", in the course of scientific and engineering development, not the HCIe conception model. The conception, by contrast, must be made secure.

It seems to me that Dowell & Long's favourable reference to the above version of a paradigm is not helpful. The authors compare what they are doing to what they call the epistemological enquiry at the beginning of Newton's "Principia" which issues in,

---

[69]The conception criticised here has been re-presented as Dowell & Long (1998), and any detailed critique might be somewhat different. However, I believe that my general points stand.

for example, the concept of inertia and 'the Second Law'. However, this is a *directly* testable hypothesis which has been arrived at in a similar fashion to that of the 'thought experiments' of Einstein. What Dowell & Long are embarked on is surely a foundational, pre-theoretical or meta-theoretical, exercise. What Newton was engaged in was not an epistemological enquiry. There may be some blurring of the distinction but defined by the enquiry's outcome it was rather a stage of a scientific investigation.

That brings me to the other comment that can be made of this use of Van Gisch & Pipino's 'paradigm'. It is a *scientific* idea and Dowell & Long's project was not going to be dependent on science but redefine a design discipline as engineering.

According to Kuhn (1970a):

> "...somehow, the practice of astronomy, physics, chemistry, or biology normally fails to evoke those controversies over fundamentals that today often seem endemic among, say, psychologists or sociologists. Attempting to discover the source of that *difference* led me to recognise the role in scientific research of what I have since called 'paradigms'" (my italics).

Psychology is the discipline most closely associated with the Human Factors side of HCI, so introducing the notion of the paradigm, though it is used differently by Van Gisch & Pipino, is likely to draw HCI back to a science base. If not then, at least for Dowell & Long, the question again arises: How to account for what Kuhn is referring to as the science's unitary quality – the failure "to evoke controversies over fundamentals", which is certainly not a feature of HCI. This question is fundamental to the thesis. It is partly answered by Part 1, but my adaptation of the idea of the domain, in conjunction with structure and rules, will overcome the general problem.

To distinguish the discipline of Cognitive Engineering (and, therefore, Cognitive Ergonomics and its young offspring, HCI) from those fields such as Psychology, Sociology etc. we should, at one and the same time, distinguish it as a discipline separate from science, i.e., as an engineering activity rather than a scientific activity, and yet show its epistemological rapport with science. It is partly thus that we derive the guarantee, i.e., both by recognising engineering's joint epistemology with science, and the systematic methods which are peculiar to engineering. In any case,

> "Neither the initial acceptance of a paradigm nor the normal scientific practice following on acceptance is rule-governed; so there can be no question of justifying scientific claims (ultimately) by showing that they are licensed by methodological rules. Rather, the cognitive authority of a scientific claim rests finally on the judgement of the scientific community that it is worthy of acceptance." (Gutting, 1984)

It is doubtful whether Van Gisch & Pipino's view of a paradigm, and its 'derivation'(?) is really helpful, since it is, arguably, only loosely related to the Kuhnian idea, which is properly descriptive not prescriptive, and, possibly, even wholly post hoc. Likewise, the conception is not properly justified by the establishment of methodological rules, and, *to the extent that this is the thesis of HCIe*, the conception is inadequate for my purposes. For this 'proper' justification, we need to extend and qualify the conception as described by Dowell & Long, looking more closely at the fundamental notions employed therein: the interactive worksystem and, in this chapter, the domain, in particular. I looked at the IWS in the last chapter, and have been concerned with it throughout much of the thesis in the sense that I have been attempting to introduce, in a systematic way, general knowledge of the agent's linguistic structure and behaviour. Now I want to turn my attention to the domain, to see how this idea fits into a scheme of technology and science.

General Approach

What I am doing might be called phenomenology. By that I mean the determination of the possibilities. While for physics, "phenomenology would be the grammar of the description of those facts on which physics builds its theories" (Wittgenstein, 1975)[70], for engineering, my idea of phenomenology is the expression of the grammar of design which is a priori with respect to any method of design, in the sense that it tries to capture the epistemological conditions of design. Dowell & Long (1989) concern themselves primarily with the foundations of a method, i.e., without doubt something

---

[70] Cf. Newton's 'epistemological enquiry' in the last section.

more than a mere language, and the rationale lies in its projected employment. I, on the other hand, am more concerned with the kind of knowledge in which it consists and with its epistemological origins, in order to relate it to other forms of knowledge. There is, however, undoubtedly this nascent phenomenology in the Dowell & Long 1989 paper: "the conception for the (super-ordinate) engineering discipline of HCI asserts a fundamental distinction between behavioural systems which perform work, and a world in which *work originates, is performed and has its consequences*." This assertion has two important implications:

(i) It is not the world which is the domain. It is that area of interest which is defined by the work performed. The more informally and tacitly this work is performed the less defined is the domain. The world is the background.

and,

(ii) As it is described in the 1989 paper, we start with the observation that something (work) takes place in the world and fulfils particular functions. The actual project of design is as yet only implicit, but the initial delineation is clear.

In other words, although the paper is not developed as an investigation into the a priori conditions of HCI as an engineering discipline, it appears to recognise the validity of the first steps of this route.

I start with a 'rock-bottom' view of domain, which is adequate with respect to the two activities of science and technology. I want to assert, with regard to a descriptive context, that the domain is defined by the *sense* of its application, while a prescriptive context defines it in terms of *purpose*. In both contexts it is defined relative to a notion of behaviour and its underlying structure. *Representation and structure are, therefore, neutral terms in relation to both sense and goal* : both are essential intermediates between the objects and knowledge (in descriptive settings), and intentions (in prescriptive settings) Clark, Pylyshyn & Chomsky are all interested in structure, expressed in slightly different ways, but a crucial difference would be the part played by the domain in any account which passes into the evaluative area of design. If this is appreciated by them it is not made explicit enough.

The reason why such an analytic investigation is undertaken as a prerequisite to understanding cognitive engineering, in general, and HCI, in particular, is that it is at least one way to avoid painting oneself into a corner, as it were; and (to continue the metaphor) to provide a plan of the 'house'(the 'house' being the framework of conditions for design in HCI) and offer procedures for negotiating one's way around it. In other words, the justification of an enquiry into such a conception can either be in terms of the minimum means to an end or in terms of the underlying conditions of its status as one discipline among others with which it is assumed to have some conceptual relationship. Or again, such a framework can be forward- or rearward-looking: forward-looking to the clear specification of design procedures, making clear distinctions to enable this outcome (Dowell & Long, 1989 and Long & Dowell, 1988 have done this admirably well) ; rearward-looking to set the conception in its relationship to science (and ordinary notions of knowledge) by looking for the epistemological origins of the key concepts and the way in which they are configured. This latter approach offers the solution to certain problems which continue to dog the perception and the practice of HCI, and, more particularly, resolves the problem of the use of science in the practice (and theory) of design. Therefore, and finally, though Dowell & Long (1989) have begun the process which leads to an understanding of cognitive engineering's ontological features, it is at times unclear at what points and how it is an epistemological enquiry. In any event, it is my aim to take what I believe to be those points, and develop or refine them as an epistemological investigation, in order to understand the central question of the relationship between cognitive science and cognitive engineering, and between science and engineering generally.

Heidegger's Phenomenological Account

When Heidegger, who often uses the idea of a tool to illustrate the relationship of the human agent with the world, addresses the problem of how the theoretical sits with the practical, he elucidates it by considering the tool, as object, in terms of two aspects of the attribute 'heavy', where the term can mean 'too heavy for some purpose' or, alternatively, 'heavy as an expression of mass under the influence of the force of gravity':

"the entity in itself, as we now encounter it, gives us *nothing with relation to which* it could be 'found' too heavy or too light" (Heidegger, 1962),

and bringing the notion of the domain to bear on both the theoretical/practical and the design/description intersections, he writes,

> "In the 'physical' assertions that 'the hammer is heavy' we *overlook* not only the tool-character of the entity we encounter, but also something that belongs to any ready-to-hand equipment: its place. Its place becomes a matter of indifference. This does not mean that what is present-at-hand loses its 'location' altogether. But its place becomes a spatio-temporal position, a 'world-point', which is in no way distinguished from any other" (ibid.).

The object is "released from such *confinement*" (my italics) (ibid.). The "confinement" here is the restriction to a domain. and, I would argue, something like this change in mode, i.e., from the 'present-to-hand' to the 'ready-to-hand' must be managed with discrimination. In some instances, it may not be feasible. We have seen that the notion of Relevance is used after the fashion of pure science, but with some uncertainty as to whether it is also 'ready-to-hand', and confusion thus arises.

Domains, however, on this view, and in contrast perhaps even with that of Gibson, are in no danger of becoming fixed and cut off from one another: that is to say, for Heidegger, there is a schema which accommodates and rationalises the many domains[71]. So what is it that might allow us, in consideration of the part domain plays in technology (and science), to integrate them, or pass from one to another: to say, for example, that we are dealing with the same objects in different domains? The objects in the world, their affordances, and the way in which these objects are related through likely tasks, are first revealed then constrained by the structures in the agent and the

---

[71] Heidegger's is an ontological schema, but because he is trying to solve epistemological problems through ontology his concerns may support mine.

requirements of the project. Thus, it is the pragmatic which permits the blending of domains, in conjunction with the stability and commonality of *structure*. I have said something about the general conditions of this integration of the pragmatic and the semantic in the context of ideas of science and technology and I shall have more to say on this topic, employing the concepts of the constitutive and the regulative. But, for the time being, I want to continue with this idea of the domain and how it affects the relative status of theory and practice.

Domains and the Relationship of Science, Applied Science and Engineering

One of the concerns revealed by tackling the problem of the evaluation and design of NLD is how we should understand what I claim is the necessary relationship between the descriptive knowledge of cognitive behaviour in the form of natural language and the problem of specifying spoken language's best use by computers to the end of natural and unconstrained interactions with human users. The reason is that the very project of attempting to simulate natural dialogue presupposes general characteristics of language use, as do other simulations of cognitive behaviour such as Virtual Reality systems.

These general characteristics cannot be derived from the solution of particular design problems, but only though an amalgamation of the descriptive and explanatory with the prescriptive. I believe this development and refinement of the notion of the engineering (i.e., its application in the cognitive field), which is inevitable in the face of advancing technology, is necessitating a radical review of the more general question of how science and engineering are related. In the hard sciences and traditional engineering, there is a growing recognition of some convergence from distinct starting points, which is consequent on the criticised assumptions of engineering as merely applied science (the one-stream view (Dowell, 1993)), but the conceptual ground of this convergence is still unclear[72]. However, one could argue that the difference, with respect to the hard sciences, between technology and science

[72] Vincenti (1990) recognises a certain rapprochement and quotes Polanyi's (1962) listing of what he (Polanyi) calls 'systematic technology' – fluid mechanics, heat transfer and solid body elasticity which is difficult to distinguish from science. Vincenti attempts only to make epistemological distinctions, and does not relate their epistemologies, except as a simple mapping exercise.

has been generally accepted. As far as cognitive science (with its heavy reliance on the apparatus of computing) and cognitive engineering are concerned, the distinction remains unclear, and has led, as I shall try to show, to some unhelpful ideas about the status of the discipline of HCI. Thus, although I believe that convergence is taking place, I do not hold that science and engineering are or will be indistinguishable, nor that cognitive science and cognitive engineering are or will be. I would claim that keeping the two classes of activity distinct while articulating the epistemological connection is the best way of enhancing them both. The concept of the domain will be the means.

Theoretical and Practical Knowledge, and the Domain

The laws, which it is the ambition of a science to establish, render comprehensible only that which is simple and general, or well circumscribed and particular. Take electricity, for example: the concepts appear complete. That is to say, what constitutes the phenomena, the units, and their relationships more or less covers the electrical phenomena: that electricity is, in some sense, well understood. However, of its manifestation in the form of electrical storms, very little is understood of their origins and development, and even less can be predicted. Or consider the physics of middle-sized objects (i.e., accelerating nowhere near the speed of light): it is perfectly well understood and precisely expressed in Newtonian principles, but, of real things, only if the material characteristics of the objects are exactly known, which, of course, is not the case. Indeed, it is not only a question of the lack of data but of inappropriate modelling.

In the absence of the ideal precise knowledge other means than the techniques of pure science, the theoretical discipline, are employed: reasonable assumptions are made or models constructed of what is happening in the particular case or mathematical theories are applied based on past experience of like phenomena: all employed in a *pragmatic* fashion. The job is still the job of *understanding* physical behaviour, and the object of study may be artificial or natural phenomena; and yet these applied sciences are not design disciplines. Their primary concern is the description and explanation of the physical world. However, in one sense, they are a step closer to the design disciplines and engineering than the theoretical sciences, and this domain connection may account for the apparent 'leakage' of knowledge from the one to the other, as well as the confusion of one with the other. This sense in which they are close lies in the kind of *engagement with the world* which is common to them both. The question the meteorologist might ask is, "What is happening in this thunderstorm or this sort of thunderstorm?", and the designer might be requested to provided a particular tool for a particular job, or a tool of a particular type, or, indeed, might be required to create or induce meteorological conditions of a particular kind, say, for a rocket launch. Both are thus engaging, in my sense, with a 'domain' of greater or lesser specificity.

The challenge is to understand this specificity, and in the case of design, to manage it. One can imagine situations where understanding and design might meet: for example, fortuitously, inducing a thunderstorm might fulfil some commercially viable project, or part thereof. The two domain-oriented projects of understanding particular instances of behaviour, on the one hand, and designing an artefact for a particular purpose, on the other, remain distinct activities, not to be confused, though they may offer one another mutual support.

> "Scientists, in their search for understanding, do not aim at rigidly specified goals. Engineers, to carry out their task of designing devices, must work to very concrete objectives; this requires that they devise relevant design criteria and specifications." (Vincenti, 1990).

Here, I understand "scientists" to mean pure scientists. Applied science shares those 'concrete objectives' with engineering, i.e., in the case of applied science, *specific* phenomena need to be explained or understood; but the job of engineering is to create artefacts to satisfy given criteria. Thus, even when pure science is engaged practically with the world, its goal is knowledge, while engineering's goal is the production of artefacts – along with which comes knowledge, in some way inseparable from the artefact. I hope I can make the basis for this distinction clearer.

Domain and the Value of Engineering as Knowledge

There are then two perspectives on descriptive and explanatory knowledge of the world which have validity, but one of them, pure science, has always had the epistemological edge. The assumption has been that it is only the absence of data which prevents the ideal knowledge of pure science solving all problems no matter their particularity. I believe that the introduction of the idea of the domain is another way of explaining their relationship which is more egalitarian and it allows a better integration of their distinct projects, which is problematic in itself but poses a real dilemma for the autonomy of systematic design knowledge.

It is not easy to mark the distinction between pure and applied science. People will want to ask why one cannot simply accept that the application of science is simply an approximation of the laws of basic or pure science. But the question is really how such an approximation is achieved. I would like to illustrate what I mean by referring to the work of a famous twentieth century philosopher, Wittgenstein, who faced a comparable conceptual problem in the evolution of this own thought; and he addressed this problem directly.

Wittgenstein went through two quite different periods of philosophy. His first was distilled in the "Tractatus Logico-philosophicus" (Wittgenstein, 1961) and was a system which he claimed was complete: that all the main problems of philosophy, he thought, were solved. He abandoned the academic life, but gradually realised not just that there were still some problems to be dealt with but that his whole approach had been wrong. He had treated philosophy as if it were science and could give a universal account of the world. The "Tractatus" was a analogue of scientific *theory*, and its completeness had led to problems. He summed up the tendency to approach philosophy this way as "the craving for generality" (Wittgenstein, 1958), which he saw as "a preoccupation with the method of science". Later in the same passage he writes, "instead of 'craving for generality', I could also have said 'the contemptuous attitude towards the particular case'". This view marked the change from his early to his later phase. The later Wittgenstein spurned the general, or at least believed that one could not re-integrate the general and the particular as features of one *system*. His first philosophy had been a defined and essentially complete structure, and his second was open-ended and incomplete with respect to the solution of particular problems. What he meant, I believe, by the phrase "contemptuous attitude towards the particular case" was the presumption involved in thinking that solving a particular problem would simply *follow* on understanding a general rule (or natural law). As Stern (1995) puts it, "On Wittgenstein's post-*Tractatus* conception of language, even formal rules must be understood in terms of their practical background, a change of view that emphasises practice over theory: 'rules leave loop-holes open, and the practice has to speak for itself'[73]. In another quotation of Wittgenstein's in the Stern's book: "Supposing there to be a certain general rule...I must recognise afresh that this rule may be applied *here*. No act of foresight can absolve me from this insight"[74] (Wittgenstein's italics). Elsewhere in Wittgenstein (1975) he writes, "This is the unbridgeable gulf between rule and application, or between law and special case"[75].

These comments are made in the context of the 'application' of mathematics, and I want to claim that the same sort of unbridgeable gulf exists between pure science and applied science, but although mathematics is one discipline it is uniform in its practices – unlike that of science and applied science. In the case of science there are two distinct sets of practices for the two categories of knowledge pursuit. I characterise them in the next chapter as species of institutions of practices; where I also reference Kuhn who has a similar view of 'application' to that of Wittgenstein.

To go back to my introduction of Wittgenstein, his two philosophical phases appear to represent an analogue of these two categories. "The world is all that is the case" (Wittgenstein, 1961). In my terms, the world is not one domain amongst many. Knowledge of the world is *constitutive*, because it represents the "totality of facts" and knowledge of it is not *regulated* with respect to anything else. It is the separation out of domains against the background of the world as well as against each other which introduces the costs and the rewards, and at the same time provides reasons for and against action – introducing value. Among the costs is the hard task of dealing with complex dynamic relationships, involving the difficulty of developing methods for holding variables steady against one another (an essential feature of an

---

[73] Wittgenstein (1969) para. 139

[74] Wittgenstein (1975) para. 149

[75] Wittgenstein (ibid. para. 164)

authoritative account) to maintain 'independence' (cf. Clark et Malt, 1984); or as Cummins (1984) puts it "the analyzed property should not reappear in the analysis", thus avoid circularity. Among the rewards is a growing understanding and familiarity with the world around us, through a concept of practice inextricably bound up with that of knowledge: a knowledge which is not provided directly by pure science.

This understanding of engineering might offer us the prospect of bolstering that notion of knowledge which we thought was the only real kind, i.e., pure scientific knowledge, but which we have been made suspicious of – by such thinkers as Kuhn and Popper. Pure theory's credibility has begun to grow thin. Engineering promises to be the way back to the origins of knowledge: a kind of historical process of editing out the putative and filling in the gaps to engender certainty, which is the basis of knowledge. Another way of looking at the introduction of domains is to see it as the first step in the re-instatement of the particular, and, thereby, opening up the possibility of a systematic move between the particular and the general.

As a device, therefore, for characterising and separating out kinds of knowledge, or the representation thereof, the domain mediates (as I put it in the last chapter) between the semantic and the pragmatic; and now, on another dimension, (or expressed in another way), between the theoretical and the practical (the practical includes the applied sciences and the engineering disciplines). The theoretical can provide well-defined semantics, but owing to uncertainties introduced by the particular and peculiar conditions of a given domain, pragmatic resources need to be employed. It is this interaction between the semantic and the pragmatic, or the theoretical and the practical which the domain mediates. It is interesting that Woods and Roth (1988), though they do not develop the idea, refer sympathetically to

> "Gibson (1979) and Dennett (1982), among others", who "have pointed out the need for a semantic and pragmatic analysis of environment-cognitive agent relationships with respect to the goals/resources of the agent and the demands/constraints in the environment."

Thus, one might say that there are common denominators between domains in applied science and engineering because certain fundamental knowledge of the domain and the cognate structures which support the behaviour in the IWS might be said to be held in common with applied science. As I said, earlier in the chapter, structures are neutral with respect to their employment, and, as I have emphasised, this commonality does not diminish the radical difference between engineering/design and science, that is to say, there is also design's own knowledge. However, with regard to diverse domains in engineering, "semantic approaches...are vulnerable to myopia" (Woods & Roth, 1988), and "to achieve relevance to specific worlds and generalizability across worlds, the cognitive language must be able to escape the language of particular worlds" – through pragmatic means. But we still need to account for this possibility.

Domain, Determinism and HCI

Wittgenstein adopted a dichotomy in his first philosophy which is in a sense carried through to his second philosophy: there was what could be 'said' and what could be 'shown'. Kant's revolution consisted in drawing a distinction between ideas which are 'real' or empirically derived, and those which are 'ideal' or transcendentally derived. I want to draw on some of Kant's ideas to illustrate another meaningful use for the domain. Determinism is an idea which is seen by some as a stick to beat the 'soft' sciences with and, by association, cognitive engineering or HCI. It is supposed that determinism is true for science, and, *given the traditional view of 'application'*, that it is also true for conventional technology. I would like to argue that this argument is impotent if determinism is considered closely and that this consideration demonstrates that the idea of the domain by highlighting the different manner of problem representation peculiar to science and technology is consistent with my developing conceptual scheme, the foundational framework.

There are two kinds of determinism: relative and in-principle.

(a) It is what is lacking when faced with complex systems, which may be rebarbative because they contain excessive detail beyond the extant tools and methods at our disposal; and might be said, in practical terms, to be indeterministic.

(b) The second is at the level of general attribution to science and engineering.

If a case can be made that it is incoherent at the general level then a fortiori it would be wholly inappropriate to apply it in order to distinguish the validity or authority of disciplines further down the hierarchy. What I want to claim is that, even on an assumption of determinism, indeterminacy is unavoidable. It is paradoxical and points up the concept of determinism as pre-scientific or metaphysical: in other words, irrelevant to the business of science and, I would argue, design in HCI.

The Laplacean argument was that God could, if he was given the position and qualities of all the matter in the universe and his knowledge of the laws of nature, calculate how it would be distributed at some time in the future. But the one indeterminate concept in this description of determinism is God. However, leaving God out since it introduces an unknown element: if one had knowledge of where all the matter of the universe were located and all the laws of physics (and a knowledge of what all this signified at the different levels of description!) we would still be unable to predict future states since the effect of the calculation of these future states would itself be unknown and, therefore, an unknown influence on these future states. This objection is principled, since it is not a problem of a contingent nature which we can get round.

In Wittgenstein's terminology, this concept of determinism cannot be 'said'; and in Kant's, it is not an empirical idea. The concept of determinism generates what Kant would call an antinomy. That is to say, it can be argued validly both that determinism is true and that it is false (as it probably is, as a matter of fact by the protagonists of opposing views), and therefore that it is a problematic idea and should be recategorised.

I think that the notion of the domain can account for this antinomy. If we think of particular problems of particular scope or domain, then we can control the calculations to satisfy what is required. If we try, as I pointed out, to predict without domain constraint, then it is impossible. We can, however, *assume* it to be true for science as a whole, but this determinism is not deduced or inferred. If it were it would be empirical. It is rather pre-theoretical. Kant, who is the origin of the constitutive and regulative rules employed by Searle, might have called the concept of determinism a 'regulative idea'[76] He applied the term 'constitutive' to the role his categories played in the attribution of objectivity to things; while 'regulative' was applied to that which could be not be determined by the categories, nor concluded from empirical study, but which governed the overall sense of reasoning.

According to Strawson's (1966) formulation, "human reason is inevitably led, in the search for systematic knowledge, to entertain ideas of an *absolute* character, for which no empirical condition of application can be specified but which may have a useful **regulative** role in the advancement of knowledge" (my bold). Scientific determinism is, I believe, such an idea, though it is not one which Kant identified.

Popper (1982), in the context of discussing determinism and free will, cites J B S Haldane's "brilliantly and clearly expressed" idea that

> "if materialism is true, it seems to me that we cannot know that it is true. If my opinions are the result of the chemical processes going on in my brain, they are obviously determined by the laws of chemistry, not those of logic"[77].

Here, in concise form, is the antinomy: if determinism is true, then it is false. Determinism is relative to knowledge, but, at the limit, it is asymptotic. "At the limit" means: if the world were to be treated as the one and only domain, per impossibile (because regulative rules can only be used relatively), determinism is demonstrably false. That it cannot be so treated is a proof of its status as of what Kant would call its 'transcendental' origin.

Determinism is simply regulative with regard to the domain: functionally dependent on the comparative preponderance of the semantic to the pragmatic. "Conception of the domain then, is of objects, characterised by their attributes, and exhibiting an

---

[76] From Kant (1964) para. A509/B537

[77] Popper (1982) writes, "It is obvious that what Haldane criticises here is not only the idea of materialism...but rather the idea of scientific determinism itself".

affordance arising from the potential changes of state of those attributes." (Dowell & Long, 1989). These affordances can be great or small, optional or not and, in some cases undecidable, but with every domain there is the possibility of greater refinement and definition of these affordances, and their representations or structures relative to the requirements which define the domain. Scientific determinism cannot be appealed to as the exclusive *property* of science. It is at best a presupposition of the scientific practice, and there is, therefore no reason why it should not be so for any systematic study or discipline. In addition, my argument implies that it makes greater sense to be systematic with respect to domain-constrained activity such as the 'application' of science and the activity of design. The fact that they are complex problems is in itself not an obstacle, in principle, whereas carrying through the pure scientific project, by contrast, does seem to pose one.

Reflections on the Domain: Summing Up

I set out to address the question of how the idea of the domain might be made more general, taking Dowell's (1995) paper as a starting point. Before embarking on the investigation, I tried to assess the status of the HCIe conception: whether it is epistemological or ontological. I concluded that it did not set out explicitly to be epistemological, though it appeared to take some steps in that direction, but overall it was an exercise in the ontology of the General Design Problem. When the authors considered what they called an "epistemological enquiry" it seemed a step which was poorly founded on scientific bases.

I had tried, in the last chapter, to illustrate how the domain made sense when taken in conjunction with notion of structure. In this chapter, given that my thesis is primarily addressing epistemological issues, I looked to the work of thinkers who appeared to be dealing with closely related matters to do with kinds of knowledge and the conditions of their validity. I considered the question, putting aside for the moment both the issue of design and the familiar subject of HCIe and CE.

I first alluded to Heidegger's manner of differentiating between things as objective entities and as entities relative to purpose or function and context. His expression was suggestive of the concept of scope and constraint which distinguishes these two modes. I then looked at the idea of the domain as not just an aspect of design but also of description and explanation, drawing on the ideas of the general and the particular, and the theoretical and the practical in the context of both Wittgenstein's philosophical development and his views on the knowledge and practice. Finally, I reflected on how the concept of the domain might account for the correct status of a certain fundamental idea which is assumed to be true and necessary for the proper execution of scientific and technological work: determinism. Re-expressing it in the terms of the domain and its associated properties strongly implies that it is not something which has to be a proven prerequisite for either science or technology; and it returns us aptly to the roots of the constitutive and regulative ideas: in Kant's work.

Domain and the Ownership of Knowledge

After a relatively abstruse foray, I want to consider the domain's more practical function: as a public policy instrument in the discrimination of science and engineering or design. It may be helpful to legitimise the conceptual terms of the foundational framework by showing how they might be used to reinterpret practical issues which turn on what science (i.e., pure science) and technology are. Public policy decisions which concern the funding of 'scientific' work, in the broad sense, are questions of import. It would be useful, therefore, if we had ways of talking about these matters which clarified the real problems. I believe, in particular, that the concept of 'domain' might offer that potential.

Polanyi, in his paper (Polanyi, 1956), which is written as a defence of the then western system of the organisation of scientific research, against that practised by the Soviet Union which was bureaucratic and authoritarian, begins by claiming that "nothing could have appeared more obvious, and indeed more trivial, in the past than the difference between pure and applied science", but he believes their discrimination requires more attention. His aim is less the purely analytic one of defining the difference between science and technology, more the proper power structure which is most beneficial given these differences. This requires some identification of what the two kinds of research are. Polanyi circles something like this idea but the paper is a complex concatenation of ideas which is not entirely cogent.

He begins by making the broad distinction between science and technology as coextensive with the difference between observation and invention, and notes that this difference is recognised by law. That is to say, "invention may be patented, observations not". However, his development of the argument leads him astray. Some little way on, he writes, "the value of a scientific observation cannot be affected by changes of value". This observation is probably true, but the implied converse – that if an invention has no utility it is not identifiable as such – is not. Whatever the immediate utility, one can easily imagine artefacts which might have utility in certain as yet unobtainable circumstances; or one can discuss meaningfully old and long overtaken technology. The qualities of technological knowledge are, thus, more fundamental than those determined by the question of value or utility.

The broad distinction of observation and invention is a fruitful one, and I shall come back to its intimate cousin – that of the theoretical and the practical – in later chapters. For the time being, however, I shall look at the implications of patentability as resting on a deeper schema. What I want to suggest is that the kind of requirements that are preconditions for the legitimate ownership of something are cognate with the criteria for something belonging to a domain, and that this is why the products or techniques of technology are, for example, patentable. In other words, real intellectual *property* has to have the quality of being predicated of a domain. As I observed in the context of Polanyi's paper cited above, it is not the case, of course, that any of the distinctions which I shall draw will demonstrate whether it is better to put money into basic science or technology. There may be other arguments for that. My concern is to show simply that there may be useful ways of discriminating one from the other if some previous policy for the funding has been decided on.

It is often said that science is public property, but is this just a contingent fact? Could it be private? In other words, does it just happen to be in the public domain, as they say? Tom Wilkie, in an article "Science is selling us out" (The Independent, Tuesday 28th May 1996), writes that "one cannot, for example, patent the law of gravity. Science is a 'public good' not only in the sense of something morally worthwhile but also in the sense of being public property", but he goes on, "modern science is expensive: it may be a public good but it is not a free good". It is clear that one cannot have the right to free access of all that is scientific knowledge, and yet it is not obvious why. Wilkie, in the same article, quotes a writer in the journal "Nature" who points out that the whole electronic industry depends "at a fundamental level on *applied* quantum physics" (my italics)[78] [79]. The writer in Nature goes on to assert that "quantum mechanics would not have been a good private investment" because, Wilkie writes, "no one company could have appropriated quantum mechanics as its own shareholders' intellectual property"; but no rationale for this view is offered. Likewise, Wilkie has earlier in the same article said that "knowledge is preferable to ignorance and that understanding is good in itself", and that he "was brought up in the post-war faith that science was both the disinterested pursuit of knowledge". All these observations are in an article which is ostensibly stimulated by the news that a certain Professor Goodfellow is about to leave academe for the industrial and commercial world, and that "the fruits of his fertile brain will become the private intellectual property of a commercial company". But just exactly what can be private property in this realm, or exactly what knowledge 'for its own sake' which is presumably the object of a "disinterested pursuit" is not even weakly broached.

There are many confusions in the above piece, and I would like to attempt, using it as a starting point, to clarify the issues by examining how the idea of domain might be usefully employed. Perhaps a profitable route to take would be that of property since it has been mentioned in the Wilkie article. I shall look at it from the point of view of the law. Wilkie cites the case of SmithKline Beechman (the company Professor Goodfellow was going to move to) having a major share in a company holding a database of human gene sequences. There is no claim here to the right of holding this information privately. It is simply owned because paid for and is kept secret as best it can be. The following comes from a publication called "Words and Phrases Legally Defined" (J B Saunders, 1969) and falls under the heading of 'property'. It is part of judgement made in an Australian court:

---

[78]It is arguably not *applied* at all, but certainly not at the *fundamental* level. However this misuse of terms demonstrates the need for the kind of foundational analysis which this thesis is engaged on.

[79]Pylyshyn ("Some Remarks on the Theory-Practice Gap"?) writes, "Engineers do design electrical circuits, machinery, and bridges with little recourse to basic physical principles, *let alone quantum mechanics*." (my italics)

> "Knowledge is valuable, but knowledge is neither real nor personal property....It is only in a loose, metaphorical sense that any knowledge as such can be said to be property. Either all knowledge is property, so that the teaching of, for example, mathematics, involves a transfer of property, or only some knowledge is property. If only some knowledge is property then it must be possible to state a criterion which will distinguish between that knowledge which is property and that knowledge which is not property."

In this country knowledge is thought of as property – intellectual property – and includes confidential information, patents and licensing. So the question is what is it that might be appealed to as a criterion to separate what, for example, can be patented or licensed from what cannot. In a general sense, know-how is potentially patentable because it provides the opportunity of exploitation and is distinguishable from the claims of right to that know-how from others. General knowledge is either too generally known to confer any possibility of exploitation, or it is too difficult to ascertain what exactly the right to knowledge is from which a patent, for example, would clearly exclude others. Wilkie mentions the law of gravity as a candidate for knowledge which is not patentable. However, it is not simply its generality which excludes it from the patentable. One can imagine artefacts which are universally used which do not thereby have their patentability reduced or diluted. What determines it is whether, in some sense, the knowledge is circumscribable. Gravity is a constitutive condition of being massive and, though particular techniques and inventions use components of particular masses in particular configurations to some useful end, the law of gravity consists of the important property that, *insofar as it is a law of nature, it is necessarily universal.* Universally employed artefacts would be only so *contingently.* Patented products, therefore, have to fulfil *particular* requirements and be distinguishable from others. This is what allows the patent examiner to say whether the application qualifies or not after studying the 'prior art'[80]. The 'inventor' negotiates with the patent office to obtain the patent, and the process clearly indicates the part 'domain' plays:

> "the draughtsman" (normally the inventor) " is usually trying to widen the *scope* of the claims, the examiner on the other hand will normally try to ensure that the *scope* of the claims is commensurate with what the examiner thinks the inventor

---

[80]"prior proposals in the same general area" (Gaythwaite, in press).

has invented, and that usually means narrowing the claims" (my italics) (Gaythwaite, in press)

To highlight 'scope' here is not to identify it with domain. The latter term has, as I have tried to show elsewhere, a more detailed and technical sense. Without going into this business of rights to ownership of knowledge it is clear that it is, nevertheless, a feature of technology, i.e., of an identifiably definite kind of knowledge that it is domain-constrained knowledge, and though this distinction is not deeply worked out it is better than that commonly used, and may serve to make discussion and decisions on public policy easier to manage[81]. At any rate, we now know that the kind of knowledge which is exploitable, i.e., not only 'for its own sake', is best characterised as *know-how*, happens to be of utility, and is distinguishable from potential competitors by virtue of its having a negotiable scope. The kind of *know-that* knowledge exemplified by a gene-sequencing database is valuable by enforced rarity, as the Australian court decision says, but as such confers no potential *right* to ownership. The company holding the database would only have the right not to have its property invaded *and* the gene-sequencing information stolen. If they were left in a public place and copied by a passer-by, the company would have no further rights over this material. Ownership can, therefore, be 'de facto' and 'de iure'. When we discuss patents we are talking of 'de iure' ownership. The ownership extends beyond physical possession: others may have the item in question in their possession but would be prevented from use without licence under pain of fine or punishment.

What I have tried to do is argue for features or properties of such items of technology which inhere in the scope or domain intimately associated with that item. It is this particular peculiarity which distinguishes the item as one of technology and not pure science; and allows it to be described and identified as particular and not universal – identifiable for the purposes of legal argument. To closely associate technology with

[81]Polanyi (1956) writes, "The difference between science and technology is broadly the same as found between observation and invention. This difference is recognized by law: invention may be patented, observations not. Empirical science is based on observation, while technology is a collection of inventions". This is a bit bald and I am, by contrast, trying to found the distinction conceptually with the help of the concept of the domain. It is also unsatisfactory to describe technology as "a collection of inventions". The rest of the paper makes it clear that he distinguishes between "empirical technology" and "systematic technology", which he identifies with applied science.

rarity value as Polanyi does is (a) to ignore this more fundamental feature, (b) to conflate its rarity value with other kinds of rarity value such as that imposed by concealment – a secret gene database, for example.

Conclusions

The general notion of domain is all pervasive: even philosophy, particularly in the twentieth century, has gone over to the resolution of puzzles and inconsistencies through the explicit examination of, for example, the way words are actually used, i.e., in their everyday specific contexts, e.g. Wittgenstein; or there has been recognition that the reality of the world is best grasped through particularly directed encounters and projects. It is part of the wide rejection of the thoughtless use of systems as media for the understanding of problems. Systems here play an analogous role to that of scientific theories. However, this rejection should not, in its turn, be carried out thoughtlessly. To take perhaps the most notable thinker who has adopted this approach and who has indicted the erection of systems as largely responsible for the creation of unreal difficulties: Witttgenstein. In his two periods of philosophical work, he espoused two very different philosophical views: one general and, one might say, system-oriented and dogmatic; the other, minimally systematic – highly Socratic and non-dogmatic. He did not indulge, however, in mere linguistic analysis, nor did it lead to reductionism and scepticism. Indeed, by reducing the importance of claims to knowledge and certainty of a global systematic kind, he put them on a surer foundation.

In his first phase, Wittgenstein's approach was formal, with the exception of the human self, ethics, aesthetics and religion. He could be described as seeing the 'objective' world as a domain: its substance expressed in elementary propositions representing what was known. He disassociated himself from this view as a complete account (he claimed that he had solved all the significant problems of philosophy, in this first phase); and, in his second phase, saw reality and knowledge as grounded in social acts or behaviour. "Giving grounds, however, justifying the evidence, comes to an end; – but the end is not certain propositions' striking us immediately as true, i.e., it is not a kind of *seeing* on our part, it is our *acting* which lies at the bottom of the language-game." (Wittgenstein (his italics), 1969), transliterated by Glock (1996a) as follows, "The ultimate foundations of our knowledge are not beliefs, but forms of behaviour."

The second phase is the practical point of view, the many-domain phase, where problems are always treated in all their particularity, i.e., regulative with respect to use. It is open-ended and pragmatic. The first phase seems to me to stand for the theoretical point of view: the one domain view or the constitutive view. I have indicated how the regulative plays a part. In the sense that theory and practice are interdependent, one should hope for some reconciliation between these two points of view. It may be that we can now discern a convergence of the two activities representing theory and praxis: science and engineering cooperating in more profound understanding of each other's role. And the fact that they cooperate means neither that their products amount to the same kind of knowledge (patentable or not), nor does it mean that the two activities merge into one another. In the next chapter, I shall turn to the question of what distinguishes them and how they might yet be connected, completing the foundational framework in the chapter after that.

I have not commented explicitly or, in particular, on Dowell's views in the main sections of the chapter, since they are set in different terms. However, I have tried to position myself with respect to them, and, at the same time, emphasise the analytic possibilities. Although I comment on Dowell's *apparently* inconsistent view that, on the one hand, the domain is complex and the principles governing the agent are relatively simple (resting on Simon's 'ant' metaphor) while, on the other, he notes the rich cognitive behaviour of the agent, I believe, nevertheless, that his intention was to emphasise the centrality of the domain as the authoritative basis for the guarantee and direction of the design.

I have felt it incumbent to widen the domain's terms of reference in order to pre-empt criticism from those hostile to the idea of cognitive engineering, by putting some of their weapons out of reach. In a nutshell, the business of CE and HCI is out in the open, a practical and systematic attempt to improve the understanding of how to determine that people or machines work better with one another, to some well defined end; and that this entails emphasis on and understanding of the domain. But it is important for the discipline's authority and unification to be clear about cognitive engineering's connection with other forms of knowledge and to see the part the domain (and structure) play in the comparative understanding of science and technology at large.

Domains are always in danger of reification and mutual exclusion. They are integrated only in the context of an understanding of the relationship of the general to the particular. Theories of perception, even ecological ones such as Gibson's, ultimately have a problem with this integration because they do not address the conceptual relationship between the more or less particular and the general. It is this conceptual investigation which allows us to zoom into small domains and out again to our whole world. It is equally the conceptual investigation which allows us to recognise the relevant changes which take place when we view the general or the particular, and take them into account, in the applied sciences or the engineering disciplines. I have tried to make clearer what I mean by domain, and how its understanding permits the discrimination of science and technology. In the next chapter I continue with the emphasis on the difference, and challenge attempts to integrate them in the wrong way, or at the wrong level.

CHAPTER 10

There are two distinct questions, which I address in this chapter: one, Is it possible to infer an 'ought' conclusion from an 'is' premiss? And two: Are the two worlds of facts and values separate (and if not how are they related)? These questions are posed as general expressions of the widespread scepticism that a prescriptive set of disciplines classified as engineering could be rationally linked to positive ones classified as science.

For the purposes of this thesis it is important also to contextualise the general questions which must be addressed by analysing typical arguments of those representing work in or close to the area of cognitive engineering. Simon might be said to be one of its fathers; Carroll one of his family members who, I believe, took a wrong turning in the development of the theory of Simon's 'design science' (while incidentally recommending many good practices consistent with what I take to be the correct view); and, finally, Dasgupta, a thinker in the closely related field of software engineering, who adopts an intriguingly extreme position.

Having looked at the consequences of these practitioners' reflections on ways across the rift 'is/ought', 'fact/value', explanation/diagnosis etc., and having found them unsatisfactory, I attempt to provide a schema which would fill the requirements of radically separating science from engineering while accounting for their interaction. This schema involves a fresh embellishment of the 'constitutive' and the 'regulative' in conjunction with another two of Searle's (1969) key concepts. They are ideas which I believe have richer implications than have been so far exploited and fit in well with the conceptual framework so far developed, but their employment is analogous to, rather than exactly as, Searle's.

Reason and Design

Introduction

In Part I of this chapter, I deal with three attempts at a rationalisation of the relationship of science to engineering (or systematic design). I believe these attempts hide typical conceptual misunderstandings of this relationship. In Part II, I argue for a view of practical reason which bridges the explanatory/prescriptive divide and could replace the problematic solutions offered by those three writers. I shall examine the epistemology of this view in the subsequent chapter.

I. Theoretical Reason & its Disciples

Even though the strong verificationist view of science has been overtaken with a more tentative falsifiability one, technology (including engineering) is widely seen as depending on scientific knowledge. It may be that an explicit application process is not understood, but in the absence of any other model it has been assumed to be so. Even if it is admitted that systems may occasionally be engineered or designed before they acquire a true scientific rationale (e.g. steam power), it is assumed that this is a fortuitous event and, more importantly, they get their authority only when the scientific rationale is established.

> "The specter of technology as a subordinate exercise, the tedious and unexciting result of applying the results of science to practical ends is hard to exorcise."

And, from the same preface,

> "...we need to confront the widely held belief that technology is applied science, and the corollary that *once we understand the discovery and justification of scientific knowledge*, nothing remains to be added about technological knowledge" (my italics) (Laudan, 1984).

One way in which this 'application' can be expressed as coherent is by considering the process of the transfer of knowledge as a rational one. It is not, then, a shoddy assumption that experts *do* believe in a kind of deduction from scientific knowledge to engineering. It is not a mere straw man in order to show off an argument: viz., "*deduction* from scientific principles has always played a minor role" ("The Task-Artifact Cycle", p75 "Designing Interaction") (my italics). In fact, it plays no role. No more do we *deduce* how to act from facts alone. They are two activities. If you could deduce one from the other they would arguably be one activity, in an important sense.

There has for some time, however, been the 'romantic' tendency to conclude, given the observation that deduction is misplaced, that design and engineering are autonomous *because* subjective and irrational. However, I believe that we do not need to deduce our engineering knowledge from science, nor do we need to imagine that because it is not deduced from science that it cannot be validated on its own terms, but must rely on the intermediacy or intervention of science.

To an extent, it is our conception of reason which is at fault. The stubborn persistence of the conception of reason as appropriate for theoretical argument has left the technologists with a sense of inferiority which has, 'faute de mieux', drawn them back to science for a proper justification of what they do, or has left them condemned to practise craft or (and some have preferred this) be seen as artists and properly creative unlike those who do what is merely normal science.

I shall examine the solutions provided by the three writers, who, in various ways do not believe in the model of engineering as the mere application of science, but who, in my opinion, fail to convince. They offer evidence of the strong desire to discover ways of giving engineering and systematic design some kind of independence or, at least, to attribute more to it than the status of "a subordinate exercise". I shall deal only briefly with Simon and Dasgupta, and concentrate my attention on Carroll. Simon's argument explicitly addresses the normative/positive divide, and I shall answer it specifically. Dasgupta's and Carroll's ideas are more intimately connected in their amalgamation of scientific and design knowledge: "designs may also be observed to constitute theories..." (Dasgupta, 1991). I shall, therefore, examine only one view, Carroll's, in any detail and point out what appears to me to be the grosser difficulties of the other's, Dasgupta's.

Quotations such as the following give us the flavour of the problem they face:

Simon (1969),

> "If natural phenomena have an air of 'necessity' about them in their subservience to natural law, artificial phenomena have an air of 'contingency' in their malleability by environment.
>
> The contingency of artificial phenomena has always created doubts as to whether they fall properly within the compass of science."

Dasgupta (1991),

> "If one examines the literature on design theory...one encounters a small number of recurrent and closely intertwined themes. Is design art or science? Can we construct a genuine logic of design? Should we try to formalise the design process? ...*What is the connection between design and science?* ...What is the nature of design knowledge?" (my italics)

and,

> "Engineering, medicine, business, architecture and painting are concerned not with the necessary but with the contingent – not with how things are but how they might be – in short, with design."

Carroll & Campbell (1989),

> "The emergence of HCI has occasioned considerable debate and perplexity about what sort of field it is, both about its relation to academic psychology and about the specific role that psychology can play in the design of new computer technology. The major positions derive from standard models of how science interfaces to practical applications in the world....
>
> "These positions all share the assumption that HCI can be defined in terms of traditional categories of inquiry."

The issue of how to see what we do in HCIe, or more generally, in cognitive engineering, or yet more generally, engineering/design disciplines, turns on distinctions which are dependent on what it means to operate in the theoretical and practical areas of enquiry. This dichotomy of the theoretical and the practical can be usefully dealt with in two moves: the bearing *scope* has on it; and how, with particular regard to engineering, the introduction of requirements, values and the associated prescriptions affects the plausibility of a practical conception of reason, and, in its turn, might affect our view of engineering knowledge relative to scientific knowledge.

Facts and Values

The thesis as a whole is the outcome of trying to answer the need for a framework for the evaluation or design of NLD. To rise above the contradictions and conceptual difficulties of the evaluation of NLD the 'semantic ascent' (Quine, 1960) gives rise to a framework for its evaluation with reference to linguistic science. This model of NLD design maps to the general HCI/HF (HF=Human Factors) or cognitive engineering model and is incorporated in an analogous framework with reference to cognitive science. The way in which the latter addresses the contradictions and conceptual difficulties encountered may resolve difficulties of relativism which arise from an unbalanced diet of knowledge as description, i.e., science. More generally, the systematic participation in the world, which is technology, takes on values of utility absent from science, epitomised by engineering, and reassuring us in its turn of the real basis of pure science, and bolstering its apparent fragility when considered through the abstractions of the philosophy of science.

The NLD framework comprised values, of a relative kind, in the form of goals and sub-goals, mapped to tasks. As part of the general justification, which is the foundational framework, we should consider the place of absolute values and the part they play in each discipline of science and engineering, and in the two of them jointly. That is to say, there may be connection between language description and language design by the introduction of relative values in the form of goals, but there is a more general relationship between descriptive/explanatory knowledge and systematic design knowledge mediated by absolute values which justify (a) realism, in science, and (b) confidence of guarantee, in engineering.

After the examination of the science/technology dichotomy in the light of a traditional conceptual dichotomy which separates facts and values, I shall continue with a critique of the three design thinkers' views, starting with one who faced the dichotomy squarely – Simon. In the second half of the chapter, I shall return to the dichotomy and a suggested solution in the terms which I have developed throughout the thesis – terms which will be enshrined in the foundational framework

Ethics and Design

In the previous chapter, as part of the argument to support the second or foundational framework, I have tried to deal with the reason for part of the temptation to misapply or misappropriate knowledge when faced with a set of design issues, in terms of an analysis of the concept of the domain. I would like now to deal with the other main reason for misunderstanding and misusing knowledge for these same purposes: the temptation to confuse the activity of design with the activity of explanation or description. I want to characterise this issue as the one addressing the assumed relationship between the descriptive and the prescriptive, or between "is" statements and "ought" statements (usually understood as an ethical problem – a version of the 'naturalistic fallacy'; its converse expression being 'the autonomy of values'). It is my contention that because this relationship has been only dimly appreciated, and not taken 'head-on', (and taking into account the inadequate appreciation of the role played by the notion of the domain) we have arrived at something of a dead-end. Design work is going on but unless the problem is faced it will be maintained only at the craft level with all the attendant problems of conflict and consequent misuse of resources. Since the issue has been raised both explicitly and implicitly by HCI writers, it needs to be looked at more closely.

Naturalistic Fallacy

Hume, the philosopher, was sceptical of the attribution of reason to matters of practicality such as morality (and more generally, evaluation and design, I would argue). For him reason dealt with ideas, i.e., analytic relations of ideas, or matters of fact. Beliefs, statements and argument beyond this pale were determined by passions, and as such did not merit the term rationality. He describes, in Part1/Section 1 of his "Treatise of Human Nature", how people discourse on religion and

> "observations concerning human affairs; when of a sudden I am surpriz'd to find, that instead of the usual copulations of propositions, *is*, and *is not*, I meet with no proposition that is not connected with an *ought*, or an *ought not*. This change is imperceptible; but is, however, of the last consequence. For as this *ought*, or *ought not*, expresses some new relation or affirmation, 'tis necessary that it should be observ'd and explain'd; and at the same time that a reason should be given, for what seems altogether inconceivable, <u>how this new relation can be a deduction from others, which are entirely different from it</u>." (my underlining)

This inference of which Hume was sceptical continues to have its resonance, and is part of what underlies one of the difficulties of the HCI discipline, and, indeed, more generally, of engineering. There are two distinct reasons for the respective difficulties. In engineering (say, aeronautical engineering), as the discipline has matured, particularly recently, so it has realised a knowledge which is not scientific knowledge because of both the source and manner of its acquisition. It has begun to look for a separate epistemological status from that of a craft or a practice derivative of a science. In the case of HCI/HF, practitioners, although at first appearing to exploit psychological knowledge as the natural way forward, began to doubt its appropriateness, and soon asked themselves whether what they were doing was not entirely different. There was a great deal of psychological knowledge but it was of little use, at least because of its generality, and yet clearly there was some connection between *what the HCI/HF workers were dealing with* and *the subject of the psychologists study – human cognitive behaviour.*

What some of these practitioners set out to do was define their area of work and propose a rigorous basis for their methods, a course of action which would generate a knowledge definitively for HCI (notably in Long & Dowell 1989 and Dowell & Long 1989). Thus, in the case of traditional engineering, questions were raised about its epistemological status through a growing awareness of its new and independently acquired knowledge; while, in the case of HCI/HF, there was a relatively rapid realisation of a lack of suitable knowledge which led to the search for independent means of acquiring that knowledge. Engineering provided a role model, but now both traditional engineering disciplines and HCI/HF needed to understand better what sort of knowledge they had, or needed to have, and how, though distinct that knowledge was related to the relevant science disciplines. Various attempts have been made, but they have led to distortions of either what science and engineering are, misconceptions of how they are related, or the creation of an unnecessary third type of discipline which is neither science (traditionally understood) nor engineering. The intentions of this chapter are to examine these attempts in order to clarify what the problem is, and to suggest a solution: both intentions will be expressed in terms of the relationship of descriptive knowledge to prescriptive knowledge, and are to be developed subsequently.

Three Views of Science and Engineering/Design

Simon

Perhaps the earliest person in the field of cognitive design to raise these questions in a closely related form was Simon who wrote in "The Sciences of the Artificial" (Simon, 1969)

> "*The natural sciences are concerned with how things are.* Ordinary systems of logic...serve these science well. Since the concern of standard logic is with declarative statements, it is well suited for assertions about the world and for inferences from those assertions."

Design, on the other hand, is concerned with how things ought to be, with devising artefacts to attain goals. We might question whether the forms of reasoning that are appropriate to natural sciences are suitable also for design. This formulation is clearly a very close approximation to the mode of expression adopted by Hume. The notable difference is that in the final analysis, Simon reveals that his approach will be to propose a new concept of reasoning to serve the purpose of design. Hume, on the other hand, implies a scepticism towards any kind of reason that could be applicable to such an activity as design (his concept of reason is a very limited affair which not only needs 'passion' to drive it, but has no reign over space and time, viz. induction and causality). My emphasis is going to be on how Simon's explanation of the nature of that newly developed kind of reasoning can contribute to design knowledge's status. It is, I think, this status which is his concern: that is to say, his temptation to endow his design reasoning with the stamp of scientific authority. However, it has to be acknowledged that Simon was among the first to clear the undergrowth in the general direction of the growing sophistication of engineering in general, and cognitive design knowledge in particular[82].

Some of the arguments Simon puts forward need examination, and one appears to involve 'begging the question', i.e., slipping the (desired) conclusion in with the premisses of an argument, thereby ensuring the preferred outcome. He says that the way we should find out what logic is required for design reasoning is by examining what logic designers use. But the question is precisely what logic they should use, or whether the one they do use is justifiably used. He goes on, sensing some circularity, "Now there would be no point in doing this if designers were always sloppy fellows who reasoned loosely, vaguely, and intuitively. Then we might say whatever logic they used was not the logic they should use." Here there is a further confusion between bad reasoning because "sloppy" and reasoning which is rigorous (good) but inappropriate.

He goes on to cite, as an example of exemplary reasoning, one of a number of methods of calculating optimum quantities for a purpose: a 'utility function'. A utility function expresses a relationship between what he calls the 'inner environment' and its constraints, and an 'outer environment' and its fixed parameters. His example concerns diet. The constraints (representing the inner environment) are daily vitamin, protein, etc. needs, plus what might be proscribed. The parameters (representing the outer environment) are the prices of foods, and the nutritional contents of the food.

[82]Nothing of what follows should detract from the importance of Simon's work and thought. In Simon (1980), he makes clear that he distinguishes the engineering and the scientific approach to cognition, *and* that they are, nevertheless, intimately related. However, in the 1980 paper at least, he makes no further attempt to say what this intimate relationship consists in, and the argument which follows is made to highlight what I take to be its inadequacy as an account of that intimate relationship. If we take this concession in conjunction with his assertion in Simon (1996) that he uses the word 'science' to distinguish systematic design from art or craft then we must acknowledge that his position is most sophisticated, does not confuse science and engineering, and only falls down in not providing a coherent account of their intimate relationship.

"The optimization problem is to find an admissible set of values of the command variables" (quantities of foods) compatible with the constraints. This is the utility function, i.e., the more or less complex functional relationship between the constraints (of the inner environment) and the parameters (of the outer environment). It is equally a more or less simple matter to optimise this utility function by computation. So this is the answer to the question about the special character of design reasoning , and Simon goes on to ask, "How does the formalism avoid making use of a special logic of imperatives? *It does so by dealing with sets of possible worlds...*" (my italics).

Simon sees the ground of the reasoning as deriving from some sort of extension of the notion of natural law to possible worlds. "We simply ask what values the command variables *would* have in a world meeting all these conditions and conclude that these are the values the command variables *should* have." (his italics). In effect, Simon is saying that owing in some way to the structure of reality, 'ought' is a kind of translation of a set of 'is'es, but as is clear from his last sentence we have come apparently, but not really, (or marginally, but not significantly) closer to an explanation, since we can now still ask what justifies the move from 'would' to 'should' (a plausible synonym for 'ought'); as Hume says, "(t)his change is imperceptible". In addition to this conclusion, which one can only describe as a dubious improvement on the *expression* of the original problem which he posed himself, he has introduced unnecessarily an ontological embellishment – possible worlds. Why does he not just talk about possibilities? Because to do so still leaves him a long way from science and its methods in which there is such faith.

Thus, Simon moves uncertainly towards an explanation of design, as a science, with all design's requirements, values and prescriptions, in effect merely *asserted* to exist in a world of requirements, values and prescriptions. Our understanding is not deepened. Remember that Simon himself wishes to uphold the difference between the descriptive and the normative:

> "I must say that I hold to the pristine positivist position of the irreducibility of 'ought' to 'is'..." (p9, footnote 3 of The Sciences of the Artificial")

However, he appears not to be able to bridge the gap (which is not the same as reducing them to identity), and so is forced by his own cogency to accept something like the radical difference between science and engineering, and to admitting that another explanation of the connection must be sought.

Dasgupta

We see a similar tendency in the work of the computer scientist, Dasgupta. That is to say, a tendency to assimilate the activity of design/engineering and science. It is not stated explicitly but the motive seems to be to account for the rigour which Dasgupta believes inheres in software *engineering*. He, like many others, feels his only hope for a principled approach to design/engineering is to draw on traditional science and its methods, but not just by analogy. The result of his reflections is a more thorough and lengthy treatment of the question, but, in the end, it is even more confused than, and as unnecessary as, Simon's.

Dasgupta's argument threads through a long book and it is not possible to refute it as such. The emphasis is, therefore, on the obvious difficulties. I believe that Dasgupta's mistake is of a fundamental kind and that the details of his argument are not enlightening. In addition, a crucial part of his argument rests on his choice of the "reference model of science", the authority of which I shall question in the next chapter.

Dasgupta starts by recognising that one can view the design process descriptively or prescriptively. His particular interest is software engineering design, and he asks,

> "Can we construct a theory of the design process – an explanatory model – that (a) can serve and enhance our understanding of how computer systems are, or can be, designed; and (b) consequently, provides a theoretical basis for building methods and computer-aided tools for the design of such systems?"

The flaw here is akin to the fallacy which Simon commits of looking at how designers reason (on what designers *as a matter of fact* do) in order to specify how they should reason (this is the *real* naturalistic fallacy). Long & Dowell (1989) draw attention to an analogous gap which exists between experimental work in psychology and the project of design. It is more than a question of domain dependence (which might be all that need distinguish the academic (pure) discipline of psychology from its possible application in HCI/HF) that accounts for this gap. Just as the mere 'application' of experimental knowledge may not (except by chance) answer the requirements of the design, so the extant methods of designers may not (except by chance) answer their requirements as designers. And even if the requirements of the design process were what drove the process of design, we could not thus specify the designers, without entering into an infinite regress. It is the practical reasoning of the artefact production which is central and endows thereby the designer with his/her rationale. This is the importance of the HCIe conception.

What Dasgupta should do is an epistemological analysis, but what he attempts is a kind of scientific investigation of the question. He refers to the Janus-faced quality of design/engineering, but appears to confuse two kinds of double-aspect: firstly, what doing design should be, and what doing design is; and, secondly, doing design, and, in doing design, following a method which is scientific in origin. The latter is more like a 'gestalt switch' underlying the mis-identification of design with applied science and is easier to understand as Janus-faced. The former is quite simply the difference between subjecting some behaviour to explanatory investigation as opposed to suggesting ways in which it might be improved.

After a review of different accounts of the process of design in software engineering, and a conclusion that these accounts are unsatisfactory, Dasgupta writes,

> " 'design problem solving is a special instance of (and is indistinguishable from) the process of scientific discovery'. I shall refer to this proposition as the *Design-as Scientific-Discovery* (DSD) hypothesis."

A little further on, he writes,

> "If the hypothesis is valid – if there are strong empirical grounds for believing it – it will shed some very useful light on (and perhaps dispel) a longstanding myth: that there is a fundamental difference between 'science' (meaning the natural sciences) and 'engineering' (that is, the artificial *sciences*)." (my italics)

Dasgupta goes on to admit that the "aims of the natural sciences and the sciences of the artificial differ", and that "one should not confuse differences in aims for differences in *methodology* or *process*" (his italics). He continues, "Thus, if the DSD hypothesis is accepted as valid it will signify that science and engineering share a common methodological basis: that is, science and design are methodologically indistinguishable."

Dasgupta refers to Simon on the second last page of the book. He points out that Simon claimed that science can be artificial too, and is not simply an extrapolation and application of pure science, which is how, he says, it is often seen. He, Dasgupta "seeks to go further, with the DSD hypothesis: that science and engineering, for example, "share a common methodology". This claim leads on, in the preface, to the stronger,

> "...from the perspective of methodology, one may actually conduct design in a manner that makes it indistinguishable from the activity we call science."

Dasgupta, again, "provided certain conditions are met, design problem solving is a special case of (and indistinguishable from) the process of scientific discovery." (It is true, conversely, that a theory is, *in a sense,* an artefact designed to meet certain requirements.) However, I would like to suggest that by "provided certain conditions are met" Dasgupta means that if we deduct what is distinctive about design and science then they, of course, become indistinguishable.

What they have in common is that they are both problems, and a problem would not be a problem if it did not have specifiable ways in which it had to be solved. Thus, in a way, Dasgupta's claim is a trivial one: problems *require* to be solved in particular ways. When they are solved they fulfil requirements. As I indicated above, if I play Dasgupta's game, I can assert that science can be seen as methodologically indistinguishable from engineering, as a species of design activity; which of course it is not. Scientists do not produce theories to answer a practical need (or even to complete a picture, qua picture). They do it to get nearer the truth: to have a better view of how the world is, or works – in short, to increase understanding. Engineers produce artefacts precisely to answer practical needs. When they increase our understanding of our environment they do so as a by-product of their goal-fulfilment – the production of artefacts.

Dasgupta is wrong. The aims of science and engineering characterise them, as he admits, but their methodologies cannot be separated from these aims, and insofar as they can, any parallels are as trivial as mentioned above, viz. that they are both problem-solving activities. Problems, however, are of all sorts, but *what best enables their solutions is a recognition of the sorts of problems that they are.*

As Laudan (1984) writes, in the preface to "The Nature of Technological Knowledge...",

> "One even still" (seven years before Dasgupta (1991)) "occasionally encounters the claim that technology is a form of science since its practitioners attempt to solve problems rationally and hence apply the scientific method. *This trivializes the issue by making the concept of scientific method so wide as to exclude nothing and explain little.*"(my italics)

Dasgupta's, indeed, raises the issue of the descriptive/prescriptive antithesis, and how the recognition of its existence must be addressed to understand better what engineering and, indeed, science are. One page (133) is confused in the light of what actually transpires:

> "A theory of design must at least in part be a cognitive theory and, therefore, descriptive in nature",

by contrast,

> "If design is indeed concerned with change it is obviously concerned with how things *ought* to be." p9, (cf. Hume and Simon)

and,

> "design is also concerned with *constructing* artefacts and the deliberate effecting of change. **A merely descriptive theory of design would be hopelessly inadequate in this regard.**" (his italics, my bold)

However, he believes that his speculations can be validated through testing and experience. This whole notion of a 'theory' of design is of a piece with the testable nature of his hypothesis of DSD. So, why is he interested in theories and hypothesis with regard to such *fundamental* questions. Are they indeed empirical questions? His own view, just cited, is that they are not. In any event, the kind of testing he has in mind is not even recognisably scientific but involves "the approach taken by the more historically minded philosophers of science" such as Kuhn and Laudan and using documentary evidence.

There is no doubt, finally, that there are formal similarities between the process of design and the process of discovery at some level of description. In general terms, it is my claim that identifying a formal match between the process components of science and design, after a reductive process, leaves confusion behind surrounding a more comprehensive account of science and design. Etymologically, the two words "discover" and "invent" mean the same thing, but it would be as naive to imagine that this semantic identity betokened a profound similarity in the two processes (indeed, anything other than their common feature of contingency) as it would be to interpret formal similarities in models as signifying real identity.

Carroll et al

Carroll & Campbell (1989) and Dasgupta have in common the conflation of science and design/ engineering. However, Carroll & Campbell (who, among others, are tackling the tendency to assimilate HCI to the applied science of psychology – the reason, arguably, being that they share *common domain specificity*) include the following, among their introductory remarks: "human-computer interaction (HCI) springs from a content-*domain* taxonomy"; and later, "an important trend in philosophy of science has been a broadening of focus, from an exclusive consideration of hypotheses and theories, to a recognition of the roles of *domains*" (my italics)). Unfortunately, they fail to continue with this line of reasoning. To have isolated the domain and interrogated its use might have led to an understanding of connection between applied science and pure science, and subsequently an address of exactly what the further difference between science and design might be, viz., descriptive versus prescriptive.

As Carroll & Campbell do not address the issue of what engineering knowledge is, they are forced to rely on the only understanding of knowledge which they possess, that of scientific knowledge. This reliance causes them problems as (i) this knowledge is descriptive and explanatory and not appropriate for design or engineering, and (ii) irrespective of how this knowledge could be used it is not relevantly scoped. This latter problem is the problem addressed by the concept of domain, which is to an extent recognised by Carroll & Campbell but, in effect, translated into an unnatural ontology. That is to say, that their idea that the knowledge is 'embodied' in the artefact is doing the job of constraint best done by the idea of the domain. The difficulty with this solution is that in dealing with (ii) concern with (i) is nudged aside. It is undoubtedly true that there is knowledge in common between psychology and HCI, but the question is how, given the differences of aim, that knowledge is shared. Further, the promising initial emphasis on the idea of the domain to the detriment of theory is lost in the shifting ground of their ontology, and theory is resurrected incarnate in the artefact.

The artefact is at the centre of Carroll & Campbell's view, both as the object of HCI's purpose, but also as its justification, and this belief is sustained by a characterisation of the object's properties and role. These properties are expressed by Carroll & Campbell as 'embodied knowledge', usability, and ontological features, and since it is on these properties that their "novel" view of HCI as a design science is founded, I consider it enough to present them as incoherent in order to undermine their position.

Embodied Knowledge

Carroll & Campbell see knowledge as embodied in the artefact, and because normally inaccessible, still (if not for good) a craft: their long-term position is not clear. They put the options on accessibility, but remain undecided, though their leaning is towards the view that something like Hayek's reading of the intractability of economic phenomena probably applies to psychological phenomena. They seem, in addition, to suggest that whatever the ultimate accessibility of the embodied knowledge it is discovered 'after the fact'. That is to say, they don't claim that it will always be grasped post hoc, but they make no suggestion as to how it would otherwise be acquired. They want to have their cake and eat it: the 'design science's' status (and guarantee) is provided by scientific knowledge, as it were 'concretised'. However, it is not science as we know it because science is explicit, and its knowledge comes along with the successful execution of its practices. Design science is different. Its knowledge is of the same kind, they seem to say: it is explicit but not at the time of the successful execution of its practices – the production of an artefact. Its uncertainties are accounted for at the same time as its mysterious ground is asserted.

This notion of embodiment is really hard to come to grips with. Carroll & Campbell make no attempt to distinguish it from other obvious ways in which it might be employed, and thus, through contrast, making their intent clearer. For example, do natural objects, by analogy, embody physical theories (Are they, like artefacts, "fulfilling many of the functions that are conventionally associated with theories"?), and is it in this sense that it is meant? Do we, as agents, embody psychological theories? Does it mean that the artefact implies the theory, or that it is implied by the theory? But theories explain. Artefacts, if they do anything other than act as media for getting the job done, *demonstrate* something, but it is not at all clear what. It is certainly not the same as 'being identical with', although the title of the paper, "Artifacts *as* psychological theories…." (my italics), suggests that the authors mean the sense of embodiment to be close to identification. They write, "artifacts *support* (?) explanation....In these respects artifacts embody implicit *theories of HCI*," (my italics). They go on, "Although explicit theory is currently scarce in HCI, artifacts are abundant, and are fulfilling many of the functions that are conventionally associated with theories." But these "abundant" artefacts are presumably of diverse value, and because our design knowledge is not explicit they can have none of the value of theories. It is precisely because we do not know how effective they are that we need a proper approach to their specification and design.

Usability

Furthermore, it is interesting to note how Carroll & Campbell shift their position from that of understanding HCI artefacts as embodiments of "implicit theories of HCI" to seeing them as embodiments of psychological theory, and, even, "implicit psychological theories in HCI", or, as noted above, simply "as psychological theories", in the title of the paper. It is easy to see how confused an understanding they have, when one looks closely at a property of the artefact, peculiar to HCI – usability. But Usability is Janus-faced: it expresses not only how the tool fits the user but also how appropriate the tool is for the task. As such it is, at least, a 'contaminated' psychological theory which is 'embodied' in the designed product, but how can we know the extent of the effect of that 'contamination'? The two aspect of usability referred to above can, by contrast, be distinguished in HCIe by means of the separation of the IWS and the domain, *and* consideration of their interaction.

As a result of the uncertain function of key terms, one is left with a confused impression of just what sort of knowledge HCI might generate: psychological, pure and simple, or psychological plus application-effectiveness. Whether it is just cognitive knowledge, in the form of psychological theory, or knowledge of greater scope is further underlined when one sees that (even) what they call the 'weak theory' means that "the working of the artefact can be understood without serious distortion in terms of a central psychological theory or theories, *plus some auxiliary details of 'implementation'*" (my italics). This afterthought shows that we still have that awkward juxtaposition of psychology and HF design. Here again the inadequacy of the term 'usability' is shown up: at times, it expresses that relationship between the user and the artefact; at times, the artefacts effectiveness of application; at times, a combination of the two. The virtue of HCIe is that it permits elucidation of these aspects to better enable the project of design.

Ontology

The assertion that the artefacts "necessarily incorporate psychological assumptions" about their 'usability', which is expressed in the various ways indicated above, is further unpacked as the claim that the artefact has some falsifiable content. It suggests that the authors want to see HCI as identified with declarative knowledge, contradicting other indications. They emphasise that the 'Design Science' they envisage is not a mere study of the design process, but "does design". However, their most general statement of what they think HCI is studying, turns out to be an ontology. They mean it should be concerned with what entities there are, and how they are. It is the construction of categories and their relationships and is, at most, descriptive: "an ecology of tasks and artefacts". The phrase has a ring to it, but I think it is a hollow ring. Presumably, the authors would claim that the importance of seeing HCI in this light is that it might promise an understanding of the impelling logic of the design of particular artefacts in relation to a set of tasks, but how would this logic be used to design further artefacts? Indeed, how could we be sure that our analysis of the connection between the tasks and the artefact was correct? In the absence of a method for design, these putatively understood artefacts could be the result of serendipity.

Ontology's 'Applications'

The authors turn their attention to the scope of this ontology's application, thus described. It should not be too narrow. It should be, therefore, an ecology of 'real' tasks and artefacts, or there will be a mismatch: we shall be studying only tasks and artefacts 'in vitro' and not 'in vivo'. In addition, since the artefacts in HCI are, by definition, to aid people in the use of cognitive tools, they have some relationship with psychology. On the other hand, it cannot have too wide a scope, or it may not have any relationship with psychology. The two extreme positions are identified with two perspectives represented by, respectively, Newell & Card (applied psychology), and Whiteside (a kind of hermeneutics, which they call engineering). But because Carroll & Campbell do not actually argue for their position in a rigorous or systematic way, we can only gain an impressionistic view of the reasons why they adopt the position they do. The force of their argument seems to depend on the extent to which the extreme views fail to demonstrate profitable exercise. They claim that Newell & Card's approach has not borne fruit, and, although they do not say so of Whiteside et al, they imply that the latter approach cannot, in principle. Since it is difficult objectively to assess the success of the various methods, and Carroll & Campbell's approach rests on taking the long view, we must infer that it is the ground on which these other HCI workers base their research which seems to the authors to be at fault, at least in the context of this paper.

Carroll & Campbell's view is that it is self-evident that, in some sense, cognitive behaviour is the common denominator of both psychology and HCI, but that psychology and HCI are not the same: that there is *knowledge* in common. That is to say, for Carroll & Campbell it is not enough to say that the two disciplines bear on the same portion of the world but that their respective systematisations of the knowledge of that portion of the world are of a piece. And why should this be so? Because Carroll & Campbell cannot without this assumption feel confident that HCI has the required stability for repeatability or validation. It is, for example, too easy to dismiss what Whiteside and Wixon do by noting that they approach each situation as individual, as if this fact in itself meant that their conclusions were incoherent. It is true that the latter discourage model-building "or any form of abstraction", but Carroll & Campbell themselves demonstrate no confidence that such will be possible with their view of HCI as a design science. It, of course, does not follow that because Whiteside and Wixon cannot make explicit what underlies their method they do not believe in its efficacy, an efficacy which implies stability and consistency at the level of practice, *and this, in any case, may be all that Carroll & Campbell can claim for their view of HCI.* The truth of the matter is that all three approaches (Newell & Card, Whiteside et al, Carroll & Campbell) are consequences of the failure to appreciate that engineering is not applied science: they have the wrong target in view, which they *call* 'engineering': Newell & Card are scientists, and for them HCI is incoherent without science; Whiteside and Wixon reject the conventional precisions of science as inappropriate predicates for the human user and, since they too see science as the only *explicit* route to knowledge, take refuge in hermeneutics; and, finally, Carroll & Campbell adopt their 'novel' view of *design science* which can, in its qualified form, be nothing more than a genuflection in the direction of the altar of science, since they can specify no way forward. They write,

> "HCI can be appreciated as a distinct sort of science: not a mechanical application of academic psychology, nor an agglomeration of situation-specific interpretations, but a science that designs and evaluates artifacts to help users do their tasks."

But this is a science which has *no* specified practices. Its knowledge is embodied and *not explicit* , with *no guarantee* that it *ever* will be accessible. In all three cases, the protagonists cannot see a way of separating science from engineering without rendering science either *the* container of explicit knowledge or *the* repository of truth and guarantee[83].

Accessibility of Artefact Knowledge

As to the accessibility of the knowledge which may be embodied in the artefact: the weak theory is that "artefacts are a provisional medium of HCI, to be put aside when HCI theories catch up"; the strong theory is that HCI theory will never catch up because the phenomena involved are too complex (this is not, however, an obstacle in applied science, viz. thunderstorms and butterflies' role). From the authors' sympathetic allusion, in the paper, to Hayek's beliefs it is clear that they plump for the 'strong' theory. In any event, it is not absolutely clear what they mean by the 'weak theory': at this stage of their reflections they propose no perspective on how the theory will come about, or should be brought about. The reasoning seems to be based on the conviction that the artefacts are embodiments of theory in some way; but note (above) it is also not obvious what theory. Here we can see another advantage in the HCIe view: there *has* been proposed a perspective on how the discipline should go forward, intimately linked to the design model. Carroll & Campbell may well try to be as rigorous as possible in their way of going about design *in practice*, but they have not set out in an accessible manner what this might mean, and they, therefore, cannot know how successful they have been in the validation of their theoretical perspective.

Dagupta and Carroll

Neither Dasgupta nor Carroll & Campbell square up to the issue of the difference between descriptive and prescriptive knowledge and its effect on the question of what distinguishes science from engineering. Dasgupta by reducing the comparison to involving only formal points of comparison marginalises the place of practices, leaving an image of design which is desiccated. It is inconceivable that the aims of science and design could be radically different while the methodology is identical. Carroll & Campbell attempt to straddle the positive/normative divide, in a similar fashion to that

---

[83]It is not, of course, the case that the term 'engineering' is not used by them: Whiteside et al. use it but mean something quite particular; and even Carroll (1995) entitles his contribution, "Artifacts and Scenarios: an engineering approach".

of Dasgupta, by providing an ontology which serves a descriptive or declarative purpose. It leaves implicit the *process* of problem solving which is (a) particular while scientific knowledge is general, and (b) which is oriented to satisfy requirements with an as-yet-unknown artefact while science addresses a known phenomenon or known phenomena as something to be accounted for. The obviousness of the difference and the tortuousness of the theoretical challenges to this characterisation, as I have already suggested, must rest on Dagupta's and Carroll & Campbell's fear that without some real infusion of scientific knowledge the business of design would degenerate to the status of art or craft with dire implications for its rationality and objectivity.

II. Practical Reason

Introduction

Science may only be the chronological (not the logical) precursor to engineering, but as a consequence of this advantage, it has arrogated to itself all the prestige of concern with 'real' knowledge. Does this view mean that engineering is without a secure foundation? Is there, nevertheless, rigour and stability in the world of values and requirements?

The question of how the worlds of science and engineering are related is not answered by responding yes or no to the question, Can we derive 'ought' from 'is'? Nor does it mean that, since science is thought of as the business of discovering the truth about reality, engineering must be condemned to, either craft-like serendipity or to dependence for its guarantee on science. To an extent, it is both a misunderstanding of what engineering is *and* of what science is that leads us to believe that there is an epistemic gap.

Rules, Practices and Institutions

The present chapter narrows the field of concern from technology (and the domain) to that of engineering: diagnosis and prescription to science's explanation and prediction. And when Searle provided his definition of constitutive and regulative rules, I believe that he introduced an equivocation which was telling in this respect, but he did not question or elaborate it. I addressed it in Chapter 3, but it is worth revisiting and reformulating in the more general context of science and engineering.

Rules

He (Searle, 1969) wrote (and I have already quoted):

> "Regulative rules regulate a pre-existing activity, an activity whose existence is logically independent of the rules. Constitutive rules constitute (*and also regulate*) an activity the existence of which is logically dependent on the rules." (my italics)

I think the temptation to equivocate, in the way Searle does, derives in part from the potential completeness of the chess model[84]. That is to say, it may be close to impossible, to work out the best move, but in principle it is possible. It seems feasible, therefore, to talk simultaneously of the rules which completely make up the game of chess (constitutive) and those which specify the best moves (regulative). Following the rules is just mechanical, like the manipulation of the symbols in a machine. Winning would involve no choice, there would be no *will* to win. Knowing the rules of the game in that complete sense is, however, beyond our capacity therefore with respect to our *de facto* knowledge of the activity to which we apply the rules they have logical independence: those rules are thus of the regulative kind.

Dasgupta and Carroll seem to have a model of science in mind which has these features in common with the chess illustration. They believe in an underlying reality which of course the artefact shares as part of the world. They believe that this world is not subject to chance in any fundamental way (i.e., processes are repeatable and predictable in principle) and neither of them makes the first step towards a view of practical knowledge by partitioning off domains as the object of study. But without this move all they have at their disposal is the constitutive knowledge. This knowledge is indifferent to perspective (is Janus-faced), i.e., it is of the way things are whether designed or natural. Thus Dasgupta can consider that design and science are identical 'methodologically', and Carroll can suggest that the artefact is the embodiment of constitutive knowledge. The problem for them, as for the chess player who abides by the constitutive rules, is that they cannot play to win using these models of design. I pointed out some symptoms of this static view of scientific

---

[84] Discussed in Chapter 3.

knowledge: for Carroll it is seen both in the ambiguity of the concept, usability, which inherits the Janus-faced property of the supporting knowledge, and in the indecision about exactly what sort of knowledge resides in the artefact; for Dasgupta, whose concern is not content of knowledge but the process of its acquisition, there is the contradiction that science and design can have identical methodologies and yet radically different aims.

Practices

So what brings together the constitutive knowledge which is at the heart of pure science, as well as providing a route out into the 'real world'? I noted early in the thesis that certain prescriptions or evaluative statements were meaningless because they were constitutive. My claim was that they came with no means of implementation. However, f=ma (force equals mass times acceleration) is equally constitutive in that it is, on one level, an equation of synonymous expressions – fodder for symbolic computation. But this expression *does* come with a means of implementation since force, mass and acceleration are quantities, and are exemplified and identified experimentally and experientially. It is, in fact, though constitutive and theoretical, intimately linked to *practice*. I can illustrate what *I* mean by 'practice', in this context, using Kuhn's (1970b)) treatment of the equation's alternative expressions acquired by students when they are faced with solving individual problems such as that for pendulums or harmonic oscillators: "Before that acquisition...Newton's Second Law was to them *little or no more than a string of uninterpreted symbols*" (my italics). This "acquisition", which Kuhn talks of, is the acquisition of what I mean by the practices of pure science, and is intimately connected with the knowledge of pure science. As Kuhn points out, it is the 'know-how' which is acquired by the students that gives substance to the knowledge.

The work of 'applied' science and engineering is the search for solutions to particular problems and they each have their own practices analogous to the practices acquired by Kuhn's physics students, but at one remove – in the 'real world'. As with Kuhn's example, what gives substance to applied scientific or engineering knowledge is, respectively, the set of practices associated with tackling particular 'real world' problems of explanation, prediction etc., or those associated with attempting to design an artefact.

Once it is accepted that the perspective on domains within the world plays a part in the definition of the activity then we can begin to discern how knowledge can be common and yet distinguishable by the practices of distinct disciplines or, indeed, by the practices within a discipline. The idea of 'structure' which derives from the HCIe model is what is subtended by constitutive knowledge. The exploitation of this knowledge for practical purposes necessitates drawing on regulative knowledge and is employed in both the solution of particular explanatory/descriptive projects and projects of design. This knowledge is pragmatic knowledge for the same reason as the strategic rules in chess are pragmatic: we do not have a comprehensive semantic grasp of the world. Furthermore, there are distinctive practices and rules of procedure which are associated with these projects and which distinguish them from each other. Dasgupta was forced by his view of science to ignore the obvious difference between a design and a explanatory problem: baldly, in the first, we are faced with an extant phenomenon or set of phenomena awaiting investigation and the outcome of the investigation will be the solution; in the second we have an extant context awaiting something to complete it, the artefact. In the first, we shall describe, analyse, predict etc., and in the second, we shall diagnose, prescribe etc. They may be closely related in fact, but they are never identical.

Institutions

So what general schema might both accommodate the separate treatment of these various properties of science and engineering, and, given the apparently inconsistent moods of positive and normative, what could account for the alacrity with which workers feel they may move between them?

We can employ a device used by Searle (1964), in a paper entitled, "How to derive 'ought' from 'is'" in order to characterise the two separate activities of science and engineering without cutting them adrift from one another. Searle showed how the derivation could take place by treating the constituents of the inference as belonging to the same 'institution'. His institutions were social institutions and were another expression of his view of speech acts. 'Promising', for example, becomes an institution 'constituted' by its practices, and as long as one was practising within the institution one followed its rules. Thus 'institutional facts' were possible: so that an obligation might *follow* from the 'fact' that a promise was made. These facts were to be distinguished from 'brute facts'. For Searle, brute facts "range from 'This stone is next to that stone' to 'Bodies attract with a force inversely proportional to the square of the distance between them and directly proportional to the product of their mass'...."

This idea introduced by Searle allows us to distinguish between the setting-up of the institution and the running of it, or as he puts it: "We need to make a distinction between what is *external* and what is *internal* to an institution" (my italics); and this distinction seems to correspond with the different modes of justifying a rule and justifying an instance which falls under that rule (see Rawls, 1955). So, now the operation of rules (and this is my, not Searle's, use of the rules), with respect to the institution regarded externally, is a regulative operation – internally it is constitutive. The rules and, therefore, the practices which are governed by these rules derive their 'real' value elsewhere (see above: the chess illustration shows how inextricably related are the constitutive and the regulative, and, consequently how comprehensible becomes the equivocation of constitutive rules which also regulate).

Internally, the *fact* (in science), for example, that something is a prediction requires, at least potentially, that one *ought* to be able to specify circumstances in which, or actions through which, the predicted events will take place. One cannot make a prediction without both technical consequences and a certain commitment: like a promise its meaning is constituted of these parts. With respect to the predicted event one can, of course, see the prediction, i.e., the act of prediction as regulative; and now the event, (its representation) as constitutive; likewise, 'diagnosis', 'remedy', 'explanation' etc.; in a manner analogous to the particular promise and its propositional content.

These activities are, thus, analogous to the propositional attitudes, 'belief', 'hope' etc., in that they entail, at least potentially, certain behaviour in order to be used meaningfully. They are tied in with the structure, and that structure both benefits from and sustains them. They are what one can collectively refer to as the practices of a discipline[85]. It is the emphasis on the practices of disciplines which determines the difference between science and technology (including engineering), but these practices, and the rules which constitutively relate them, are themselves regulative over a constitutive representation – structure. The practices, therefore, are not simply supported by knowledge. They are part and parcel of that knowledge; but, conversely, the structure underlying the practices of a discipline permits communication between disciplines which may be radically different as institutions. Consequently, what counts as good engineering practice (fruitful methodology and sound principles) evolves and is justified by its efficacy, just as science is, but it stands on its own and has no need of science as ultimate guarantor. Indeed, science will make it errant if followed as such.

Indeed, if we are sceptical of the move from scientific knowledge to design knowledge (or, indeed, of the authority of engineering knowledge), then we should be equally sceptical of the move from the facts to any generalisation of the facts *within* the discipline (or institution) of science. The alleged process of induction, which justifies the move from particular to general, is on a par with that of the inference from 'is' to 'ought', in the sense that both raise the question of how one passes from the facts to a judgement beyond the facts. Nevertheless, though it is true that nothing substantial follows from 'facts' – they just are, few (if any) facts can be described as facts in this strict sense – 'brute facts'. They usually form part of a project. They are employed for, or offer themselves for, a purpose. Their usefulness is what allows us, in part, to attribute validity to them *and* their employment.

[85] "I use the word 'practice' throughout as a sort of technical term meaning any form of activity specified by a system of rules which defines offices, roles, moves, penalties, defenses, and so on, and which give the activity its structure. As examples, one may think of games and rituals, trials and parliaments." (Rawls, '55).

In other words, at bottom, science can be just as difficult to justify as engineering, in general, and HCI, in particular. What Black (1964) called "Hume's guillotine", that which cuts off the factual from the evaluative, does apply to something like Searle's "brute facts", i.e., that which just is and about which there is no judgement. Without their admission to some system of practices – institutions – they imply nothing. That is to say, as noted above, any such stricture impacts on presupposed mechanisms of science as well as engineering by indicting induction. It should be noted that this parallel between induction and the 'naturalistic' inference happens on two different levels. As with the treatment of 'determinism' in the last chapter, it has a rather peculiar role, and can be viewed as like Kant's 'regulative idea'. It is regulative with respect to the unitary domain of the world characteristic of science. In the next chapter, I shall try to enlarge on this integration of practice and knowledge.

Conclusions

I have examined the opinions and arguments of three representatives of research, either in the area of cognitive ergonomics, in the case of Simon and Carroll; or, in the closely related area of software engineering, all of whom have a view of the how 'traditional' science interacts with engineering (or the more widely defined practical activity of technology). I feel very sympathetic to the views of Simon, and he has many important things to say about design, some of which I have referred to. However, his firm assertion that the realm of 'is' and 'ought' do not coincide, and his analysis of the logic of the relationship of scientific knowledge and engineering knowledge fall short of an adequate account. He continued to treat this systematic knowledge of design as scientific, and although in Simon (1980) he felt there could be no objection to using the term 'cognitive engineering', the idea that engineering could contribute to understanding in the way that science did, cannot be accommodated by the contemporary conceptual framework.

Carroll went further than Simon in constructing such a conceptual framework, but in doing so I believe he strayed further from a coherent perspective, and produced new entities which stretched intellectual credibility far beyond that of Simon's 'possible worlds'. I tried to show that these entities were defined variously and confusingly. Artefacts as 'potential scientific objects' are referred to by Carroll in Monk et al

(1990), and demonstrate how unaccountably rapid is his transition from design activity to scientific knowledge. He appears unable to separate out what I have alluded to as the Janus-faced aspect of design, perceiving artefacts as 'embodied theories', and implying that if they were not then we could have no confidence in the science of design.

The other representative, who takes an extreme and explicit view, Dasgupta, is remarkably close in spirit to Carroll as I have already remarked, though he makes no reference to him. He even writes, "In the domain of artifacts , designs may also be observed to constitute theories....". And it is from his book (Dasgupta, 1991) that the allusion to the Janus-faced property of a design comes. Dasgupta's elaboration of this idea is that systematic designing can be characterised positively or normatively, and that, given his reference model of science, his claim is that "design problem solving is a special instance of (and is indistinguishable from) the process of scientific discovery". He applauds Simon for his view that the design sciences are not *based on* the traditional sciences, and claims that "the Design-as-Scientific-Discovery (DSD) hypothesis seeks to go still further" by asserting "that, methodologically speaking, the 'science' of the natural sciences and that of the artificial sciences are fundamentally alike".

Though he certainly goes further; in an important sense, he does not go far enough. Simon, in spite of not articulating clearly enough how the positive and the normative were linked, did want the independence of the artificial sciences from that natural sciences. He wanted to argue for their own kind of rationale. Dasgupta, however, cannot prevent himself from disclosing his preference for the positive sciences when he writes, as above, that "design problem solving is a *special instance of* scientific discovery" (my italics). Whatever their status, he sees artefacts, since they are like Carroll's embodied theories, as instances of possible theory refutation. I believe this characterisation only occasionally *appears* to fit the facts, and that it may be because of his background as a software engineer, with its formal qualities perhaps only marginally concerned with questions of usability, that he adopts this extreme position. The actual theoretical import of designs is minimal and their real importance lies in the degree to which they satisfy requirements.

Two engineers, Rogers (1983) and Vincenti (1990), who have reflected on what one of them (Rogers) calls the 'philosophy of technology', provide a more realistic appraisal of the part artefacts play. For Rogers, insofar as it can be said that engineers have theory, they must, he says, "because of the catastrophic consequences of engineering failure...try to avoid falsification of their theories"; and he points out that "nuclear power stations are designed using a theory called 'reactor physics'...based on a crude model of the atom", and that "much optical equipment is designed on the basis of the corpuscular theory of light". Rogers' concern is that these theories cannot be viewed as validated by successful design, but it is equally clear that the *designs' value* do not lie either in their potential refutation. Relatedly, Vincenti writes: "What designers did *not* know appears as consequential in its own way as what they *did* know." He further notes, "This observation brings to mind a difference that may have epistemological implications: In science what you don't know about is unlikely to hurt you (except possibly in some unfamiliar experimental situations). In engineering, however, bridges fall and airplanes crash, and what you don't know about can hurt you very much."

Vincenti indicates other feature of engineering knowledge which are distinctive: just as Polanyi's 'operational principle' (Polanyi, 1962) may distinguish the artefact from the natural object, science may know everything about a machine qua physical object but not qua machine. Vincenti writes,

> "It is in no sense knowledge about how an airplane innately is; rather it is knowledge about how an airplane ought to be to enable the pilot to fly the machine with ease, confidence, and precision. Such knowledge had (and has) no interest or importance for scientists; it was discerned and almost entirely generated by engineers and pilots working together for essentially engineering purposes."

and, a little further down the page,

> Part of this knowledge comes from science, but much of it (e.g. the allowable strengths for flush rivets) arises within engineering itself. (Though the sciences deal with how things are, they are not the sole source of such knowledge.) *Clearly, if engineering knowledge is to be understood fully, it must be addressed on its own terms*". (my italics)

Again, with respect to the aims of the different disciplines, Vincenti writes,

> In science…the means acts directly to the end; in engineering it acts through the intermediary of the 'something ', usually a material artefact, that is the immediate object of design (or production or operation)."

How then can it be that two such different views (those of Dasgupta and Vincenti) of the kinship of science and engineering be held by workers in different branches of engineering, who have not arrived at their opinions without considerable reflection and argument?

Apart from practising different kinds of engineering, which might serve as a cognitive account of the difference, I believe that Dasgupta's position in this debate is at least partly due to his reference model of science. It is a sign of the unsettled state of the philosophy of science that it has enabled people like Dasgupta to take advantage of its prevalent instrumentalism to make confusing conflations of the practical aims of engineering with the necessary features of any science, i.e., its practices – something which no systematic activity can do without, however pure or abstract. I shall address this question in the next chapter as a preface to an expression of the foundational framework.

These practices (both scientific and technological) were the subject of Part II of the chapter, and my solution to their integration across the fact/value rift borrowed the concept of the 'institution' from Searle (1964), which is , for Searle (and for me) a schematic basis for some of the distinctions, which I have made before, between the functions of constitutive and regulative rules. But here, in this chapter, the arguments have been developed in order to bridge the putative gap between the positive and the normative. Figuratively, it might be said that, in the form of the 'utterance', i.e., as composed of an illocutionary and a propositional component, the 'institutional' schema accounts for the juxtaposition of facts and values, and for traffic in those facts between the very different 'illocutionary' components of science and engineering. However, I have displaced the application of 'institution' so that 'institutional facts' and 'brute facts' correspond in some way with the illocutionary component and the propositional component of an utterance, and, in that context, brute facts only exist by contrast; and scientific facts are 'institutional'. I have tried to clear up some of the confusion of what engineering is, but as I commented before, misunderstandings of what science is also contributes to confusion of how the two sets of disciplines are related and how different. So, only after turning my attention to an alternative reference model of science, in the next chapter, quite different from that of Dasgupta's, shall I try to enlarge on this modified notion of Searle's 'brute facts'.

Finally, to distinguish, as Long & Dowell (1989) do, sets of *practices* associated with disciplines began the process which will result in a proper regard for the relationship of facts and values. We have been obsessed with an 'observational' and detached perspective on knowledge rather than recognising, as we should, how the manner of our engagement with world determines what we know.

CHAPTER 11

The last two chapters have exposed features of science and technology (including the distinct disciplines of applied science and of engineering) which instantiate concepts deriving and developing from the model of HCIe. In addition, the idea of disciplines as 'institutions' with their proprietary practices has been introduced, devised to accommodate and integrate description and prescription (or facts and values). Together these provide a rationale for the two dimensions within which engineering differs from science: respectively, scope of 'application' of the knowledge, and the widely accepted dichotomy of 'is' and 'ought' formulations. This schematic representation of science and technology shows how we might describe the distinct activities of science and technology with terms which are common.

There seems little doubt that the 'two-stream' view is the correct one, but it is not easy to reconcile this view with our commonly held concept of knowledge is acquired. A re-assessment of the model of scientific knowledge acquisition is required

The outcome is a better balanced model of science and engineering, with applied science straddling the divide, in conjunction with a refined image of knowledge which is common across the range from pure science to engineering. The model is set in the foundational framework which is formulated in a technical vocabulary consisting of terms which have been introduced over the development of the thesis argument.

Philosophy of Science and the Philosophy of Technology

Introduction

I want now to bring together the essential elements of the argument as they have been developed in Chapters 8, 9 and 10. However, the roots of the argument lie in the first part of the thesis, and there are points of contact between the NLD framework and the conceptual framework, to which I shall make more reference later in this chapter and in the concluding chapter of the thesis.

Some Recapitulation

In Chapters 9 and 10 particularly, I developed the argument that consideration of the foundational construction illuminates the thesis claim (i) that science and technology (comprising applied science and engineering) are intimately related but quite distinct, and (ii) that science (comprising pure science and applied science) and engineering are also intimately related but quite distinct. The intimacy is, in one way, a common sense view, since we imagine that we use the same knowledge (at some level of description), but exactly how this intimacy should be articulated is a harder task to determine and to carry out. However, as I have attempted to demonstrate in Chapter 10, some key workers concerned with design knowledge exhibit many unclear ideas about the questions of the relationship of science and engineering, and by implication, the constitution of the two. The origin of the concepts which I have elaborated to encompass these questions is the conception of HCI as an engineering discipline, developed by Dowell and Long and published in the their 1989 paper. In particular, the concepts of 'domain' and 'structure' underpin this framework and, in an elaborated form will thus form part of my foundational framework. It is my opinion that this foundational framework serves not only to ease the problematic arguments referred to above which bear on the epistemological links of science and technology, but also make plainer how the process of the generation or incrementation of engineering or design knowledge might be achieved, and I gave an outline answer using HCIe vocabulary in Chapter 8 to illustrate my elaborated versions of some key terms. In what follows, I shall revisit this question in the context of Dowell's thesis argument (Dowell, 1993)

In Chapter 1, I gave my initial reason for undertaking this thesis as principally one of resolving the discomfort felt about employing certain linguistic resources for the purposes of design. However, in the context of HCIe no practices are put forward using knowledge which is not predicated narrowly on the requirements of projected design, and that is an insufficient basis for NLD as such. By "narrowly predicated" I mean centred on some more or less implicit view of engineering knowledge as distinct from scientific knowledge; or at least, derived independently. Indeed, Dowell (1993) argues cogently that engineering generally is an activity which operates independently – a view endorsed now by most serious students of history of technology – and he writes:

> "the two stream view of science and technology relations posits disciplines of design, parallel with, but autonomous from, science disciplines"

He notes later in the same section, referring to engineering, that "it is unconstrained by scientific theory", which further qualifies what is hinted at in the phrase "parallel with" from the first quotation above. That is to say, of course, the claim is de facto that the division is clear: that engineering does not derive from or depend on science for its progress. It does not exclude some kind of exploitation of scientific knowledge for engineering ends, but this exploitation might be of some contingent nature. However, Dowell's assertions have not been explained except in their negative implications about the relationship between science and engineering. As to a positive view of what it might amount to, he refers to the beliefs of Alexander Koyré as described by Layton (1974), saying that the "influence was more metaphysical". Layton gives the example of

> "the idea of a world governed through Galileo's and Huygens was transmitted to technology by the conversion of the mechanical clock into an instrument of precision. The idea that the universe is governed by precise mathematical laws, it should be noted, was not a scientific result, but one of its presuppositions.... Koyré assumed that the influence was indirect, involving something like a translation of the idea from one medium to another."

So Dowell does not exclude the mutual influence of science and engineering[86]. He describes them as interacting from time to time, but apart from the reference to Koyré's view he does not elaborate. His own expression of the interpretation of cognitive engineering (CE) is as "a design discipline, *possessing its own knowledge* to support its praxis of solving the general design problem of Cognitive Engineering" (my italics). This can be taken in conjunction with Long & Dowell's (1989) criticism of Hammond & Allinson's (1988) project to "use 'encoding specificity' theory to formulate a design heuristic": that it seems as likely as not to lead to a solution to the design problem; or more strongly, "it appears that Cognitive Engineering knowledge is likely to arise, not so much in the theories of Cognitive Science, as in the space of Cognitive Engineering design problems". When Dowell adds, of CE, that "it is unconstrained by scientific theory" (see above), it is clear that something more specific has to be said about this two-stream view in order to account for how parallel streams can interact.

I indicated above that there were two respects in which one might extrapolate the key concepts – 'domain' and 'structure' – of the HCIe conception: as a means of re-unifying science and technology epistemologically; and in order to account for the generation of engineering knowledge . I have tried to make the first one clear, but it is perhaps equally important, though less central to my interests, that I introduce the other. In John Dowell's thesis (Dowell, 1993), the author acknowledges that there is

---

[86]As deSolla Price (1984) puts it, "thermodynamics owed much more to the steam engine than the steam engine ever owed to thermodynamics".

a need to generalise solutions "over sets of instances", in this case, of rATM[87] system: "this expansion in scope must be in many dimensions". It is intuitively clear that we can keep certain features unchanged and modify others, and undoubtedly this outcome will be useful and informative. However, it is not obvious precisely how it would be effected, if we bear in mind that the power of the approach lies in defining the domain in terms of the requirements, and that the 'construction' of the worksystem is consequent on this definition. Dowell has defined a domain as "a world in which goals arise and whose state is transformed by worksystems seeking to realise goals", or, "the domain is not 'the world'...about which we may have common sense or physical theories; rather, it is an abstraction from that world". He goes on to emphasise that it is not there independently of our goals or purposes. It is one of the distinctive properties of the conception of cognitive ergonomics as an engineering discipline that it does not apply to the world "about which we may have common sense or physical theories". In other words, it is to a greater or lesser extent particular.

One of the difficulties which science does not have to face is how to generalise over instances since its 'domain' *is* the world. It is difficult in the case of engineering to see how generalisation can be accounted for, since unlike science it does not start with some general theoretical claims. How can it be known whether a change in the domain under study will not result in some unforeseen results? New requirements may follow which require new specifications supported by new structures. In other words, over what does the generalisation take place? Unless some prior supposition is made about the enlarged or modified domain, and its consequences for further specification, it seems that generalisation cannot be made with any reliability. But this would be putting the theoretical cart before the practical horse. It is just because technology (including engineering) deals with domains of particular aspect that conclusions from the work of either pure or applied science cannot be exploited directly. In the case of pure science, we ignore the very particularity of the design problem; and in the case of applied science, though a domain may be 'objectively'

---

[87]Stands for 'reconstructed Air Traffic Management' and was created as an exemplar of Dowell's (1993) cognitive engineering framework.

similar, the fact that one concerns the way things are (applied science) and the other is qualified by requirements which need to be filled (and, therefore, are not 'objectively' there) leaves open the character of the outcome of any projected link[88].

To sum up, the emphasis on the definition of the domain of work – its content and boundary – precludes the direct application of *general* scientific knowledge because it does not take account of the *particular* design problem. This feature of design problems, together with historical and general observations of the independent development of technology, gives rise to the 'two-stream' view of science and technology. Thus, it is *not* clear that "understanding of the design problems confronted in one instance can be applied prescriptively and with assurance to solving design problems in other instances" (Dowell, 1993) without facing the same dilemma, i.e., either give up the autonomy of engineering knowledge, or face being trapped by the particular. Dowell, in the last quote (which is in the context of a reference to Woods & Roth (1988)) assumes that engineering principles function as the 'deus ex machina', but what principles are is far from clear; and indeed, Wood & Roth express uneasiness in the face of entrapment by the particular:

> "Semantic approaches...are vulnerable to myopia. If each world (i.e., closed world) is seen as completely unique and must be investigated 'tabula rasa', then cognitive engineering can be not more than a set of techniques that are used to investigate every world anew".[89]

It is not my claim that it is unreasonable to assume general properties of design problems (or the operation of pragmatic reasoning with respect to the design problem), it is simply that it is not clear that cognitive engineering autonomy as described by Dowell does not preclude those assumptions. Unless and until the relationship between science and engineering is articulated, the danger of epistemological isolation remains. Another way of putting this is to suggest that if

---

[88] It is important to remember that just exactly what counts as 'real' science is not the point. It is quite likely that psychology might not be regarded as yet privy to natural law but is rather science in its practices – a young science. However, this view does not detract from the argument against using its results directly for technological purposes.

[89] Already referred to in Chapter 9.

there is the possibility of interaction between science and engineering which Dowell acknowledges, *but* this interaction is left implicit, then we cannot know that the possibility of the knowledge generalisation in CE is not a consequence of this unarticulated interaction.

Prima facie, thus, the need for some further justification for domain generalisation implies theoretical suppositions which do not sit easily with the 'two-stream view' of science and technology. They seem, at first, to throw us back on the discredited view that science is the fundamental arbiter and the well-spring of knowledge which is exploited by applied science and engineering. However, perhaps our interpretation of what engineering knowledge can be is influenced by out prejudices of what scientific knowledge is.

Role of Science in the Relationship

It should not be surprising that we may have to re-examine the role science plays in order to understand engineering anew. Views of what engineering and science are appropriate for, and capable of, are interdependent and that is partly what has made discriminating them difficult. Changes in the views of the scope and force of scientific knowledge which have taken place in the latter half of this century's development may, therefore, offer us new perspectives on the place of technology and the significance of its work.

As engineering grows in sophistication it becomes self-supportive – an autonomous and intrinsically valuable discipline. Vincenti (1990), suggests what he understands by knowledge with respect to engineering. He says that though "he goes along with customary usage", he does not "subscribe to the customary perception" that

> "the term *engineering knowledge*, as I understand it to be customarily employed, refers to the knowledge *used* by engineers. *Scientific knowledge*, by contrast, usually means the knowledge *generated* by scientists....Only recently have scholars begun to look seriously at engineers and engineering activities as knowledge producers"

I have argued for the difference between the two activities science and engineering as being one of orientation and purpose, and Vincenti makes the same case. So how can we reconcile this position with the growing tendency to attribute knowledge of the physical world to be in some sense a product of, for instance (and among others), aeronautical engineering? One way would be to introduce an element of symmetry into the equation by interpreting 'scientific knowledge' as knowledge peculiar to science: knowledge of investigative techniques, experiment etc., the counterpart of what is referred to as engineering knowledge (i.e., what Vincenti characterises as "knowledge used by engineers"). In this way, engineering's claim to knowledge as indirectly generated looks less anomalous, if we consider that science too produces knowledge of a comparable kind, indirectly.

Philosophy of Science

Opinions have changed dramatically since the '60s on what kind of discipline, or set of disciplines, science is. As Hacking (1983) puts it,

> "philosophers long made a mummy of science. When they finally unwrapped the cadaver and saw the remnants of an historical process of becoming and discovering, they created for themselves a crisis of rationality"

Hacking, like many others, identifies this crisis with the publication of Kuhn's "The Structure of Scientific Revolutions" (Kuhn, 1970a). The "mummy of science" was the view reflected by Carnap and Popper principally, because they saw science, particularly physics, as exemplifying rationality. Hacking goes on to list the points on which they disagreed: "Carnap's verification was from the bottom up", "Popper's falsification was from the top down"; Carnap believed in induction and Popper did not, and so on. However, "they share an image of science, an image rejected by Kuhn", and, a little unfairly and too concisely, perhaps, Hacking states "if two people genuinely disagreed about great issues, they would not find enough common ground to dispute specifics one by one". The great issues they agreed on, he continues, were that theory and observation are sharply distinguished; that knowledge is cumulative; that scientific reasoning is deductive; that terms in science have to be precise; that science is a unity, which latter property means that sciences share a methodology, and that if you want to be exact and as close to the truth as possible then you need to apply that methodology whatever your sphere of investigation. After the publication of "The Structure of Scientific Revolutions" by Kuhn, it is generally agreed that our view of rationality, as it stood then, was at least inadequate to reinstate respect for science as the royal route to knowledge of the universe. Some thought, and some still do, that it finally confirmed the sovereignty of relativism. As is well known Kuhn was as surprised as anybody that this should be one of the consequences, and he attempted to right the impression the book had given. The shock has passed and many have reinvented their viewpoints to match what they see as a new kind of rationality.

The central problem of the truth of scientific theories had already been questioned by an earlier examination of the reasoning process by which theories were confirmed. J S Mill (Losee, 1972), who is famous for his views on induction, maintained that to verify a hypothesis you have not only to ensure its deductive correctness, but also exclude the possibility, that no other hypothesis can imply the facts to be explained; but he did not provide any such proof in the one case of a theoretical claim which he thought qualified as completely verified. Losee writes that "he was aware of the difficulty of excluding alternative hypotheses". What Kuhn's work did, in addition to this questioning of the logic of verification, was invert the significance of theory and what the theory was about. Popper and Carnap stood for a belief that nature was there objectively and that a systematic investigation would reveal its secrets in a cumulative fashion. This was now no longer certain, and the philosophy of science lost its central emphasis on theory.

Thinkers began to attack the problem of scientific knowledge from more general epistemological positions. Sense data *theories* of perception had been called into question by ordinary language philosophers such as Austin and Wittgenstein and so the scene was set for some regrouping. Perhaps theories at a higher level of analysis – that of the philosophy of science – could also be put into proper perspective.

Hacking's work (particularly, Hacking (1983)) represents one of these movements, though he appeals to the ideas of thinkers who preceded him by many centuries (such as Francis Bacon) he is carried forward on a wave similar to that initiated by Wittgenstein and Austin. His position is that of the 'entity realist'. His argument is woven throughout his book and is well illustrated with examples. Essentially, he takes an unexceptional view of the everyday world and our normally veridical experiences such as seeing and manipulating things, and he enlarges the class of these acts to firstly, looking at distant (normally invisible) or microscopic things and arguing that this kind of seeing is not a special case (in any relevant way), and secondly, manipulating (again not in a special sense) theoretical entities. He gives as an example varying the charge of niobium balls by spraying them with positrons or electrons in order to detect the presence of quarks (which are calculated to have a third of the charge of an electron)[90], i.e., in order to observe whether the charge changes only in units of the third of an electron charge. In this case, Hacking says he is not concerned with the proven existence of quarks (and adds that he might well have been sceptical of electrons' existence in the context of Millikan's experiments in 1908). What persuades him of the reality of such as niobium balls is the gradual adoption of what were at one time unlikely entities as instruments in the search for other still controversial ones.

Hacking sees these behaviours as of a piece with the kind of behaviours which we cannot, treated in general, doubt, and in this his work is in the spirit of Wittgenstein's thinking in "On Certainty" (as quoted by Smithurst (1995))[91]: "The truth of certain empirical propositions belongs to our frame of reference" (Wittgenstein, 1969, , para 83); and as an example Wittgenstein talks of scientific facts as "fused into the foundations of our language game" (para 558). Wittgenstein' view is that language is not simply and only what is said and written but what set of practices it is a part of. Thus, to take another apposite quotation from Wittgenstein (1969):

> "Giving grounds...justifying the evidence, comes to an end; but the end is not certain propositions' striking us immediately as true, i.e., it is not a kind of *seeing* on our part; it is our *acting*, which lies at the bottom of the language-game" (para 204)

So, what Hacking is getting at is that at the core of any activity we have little opportunity to doubt the significance of what we are doing; that as we move away

---

[90] Cartwright quotes Hacking's famous remark (Cartwright, '96), with approval: "If you can spray them, they exist", only to qualify it, as a good empiricist, by saying, "When you can spray them they exist".

[91] Though Hacking's intellectual forbears are more directly the American pragmatists, Peirce and Dewey.

from this centre we do have some leeway, and thus that the position of theory is not so secure. He can, therefore, take the position of being a realist about scientific entities and remain sceptical about theory. This position, for him, obviates the dangers of what he calls, alluding to Dewey, the 'spectator theory of knowledge', which was upset by Kuhn in the '60s. Hacking's eponymous theme was that theory (representing) had played a disproportionate part and that acting (intervening) should be re-asserted. As he forcefully expresses it:

> "Every test of a representation is just another representation. 'Nothing is so much like an idea as an idea,' as Bishop Berkeley had it. To attempt to argue for scientific realism at the level of theory, testing, explanation, predictive success, convergence of theories, and so forth is to be locked into a world of representations".

However, it is worth countering the tendency to swing the other way and against theory altogether, by bearing in mind, as Hacking also pointed out, that the changes of the '60s meant that theory and observation were no longer precisely distinguished, but were interdependent – to quote Koyré:

> "Far from being opposed to each other, experiment and theory are bound together and mutually interdetermined....", and, "an experiment – as Galileo so beautifully has expressed it – being a question put before nature, it is perfectly clear that the activity which results in the asking of this question is a function of the elaboration of the language in which it is formulated; *experimentation is a teleological process of which the goal is determined by theory*."[92] (Koyré, 1968) (my italics)

If entities can have some reliable reality attributed to them then so can theories. The virtue, however, of something like Hacking's position is that it makes practices (with their techniques and instruments) central to any theory of scientific knowledge, and, if this perspective is adopted, perhaps it will be easier to integrate scientific knowledge with that of technology, generally, and engineering in particular.

---

92 An example perhaps of the regulative in the service of the constitutive.

Theory and Domains

My view rests on the assumption that pure science and technology are distinct, because the first deals with the universal expression of knowledge and the second, the particular. This view is what is usually meant by the theoretical as opposed to the practical. It is impossible to deal with the subject thoroughly, but I believe that for my purposes, I need not go deeply into the issue of what kind of general knowledge scientific theoretical knowledge is in order to make my position clearer. However, as a postscript to 'entity realism and before moving on to a consideration of these matters from a technological point of view, I would like to look more closely and critically at the tendency, mentioned above, to swing away from theory altogether by considering a little the views of someone who dismisses theory in the sense of pure science, and in doing so provides a good illustration of what I mean by applied science, though she would just call it science.

Nancy Cartwright (1996), who sympathises with Hacking takes a belligerent attitude to the (often implicit) philosophical views of those who might be described as pure scientists. She describes them as fundamentalists. What she means by fundamentalism is the belief that there are laws which hold universally. Her position is that there is justification for believing that there are laws, but she does not see why they should hold beyond the tightly constrained conditions of the situation within which they are observed to apply. Once we move out of these tightly determined situations the matter becomes unclear:

> "consider a falling object. Not Galileo's from the leaning tower, nor the pound coin I earlier described dropping from the upstairs window, but rather something more vulnerable to non-gravitational influence. Otto Neurath has a nice example".

She goes on to quote Neurath's suggestion that a behavioural psychologist will have more success predicting human behaviour in certain circumstances than a physicist will in predicting how a thousand dollar bill swept by the wind will land in "St Stephen's Square". Her claim is that the accuracy with which universal laws apply is such that we have no justification in believing in them. Rather, she implies, we can only invoke *ceteris paribus* laws of mechanics in such situations thereby undermining the laws' reality, as models of the particular case.

Cartwright then suggests that someone might, however, attempt to model the fluttering paper money using fluid dynamics (a better model), and just might succeed, but this would testify to the action of no universal natural law. It would only hold in that set of circumstances. She says, there are laws where mechanics does not apply, but "fluid dynamics may have loose overlaps and intertwinings with mechanics". Are we (a) to suppose that the connections are then contingent and (b) that any model since it concerns local law or laws is, in my terms, constitutive? As to (b), she writes, "Theories are successful where they are successful and *that's that* (my italics); and (a) leaves us wondering how we make *meaningful* connections between different sets of circumstances in which explanatory models are devised.

The interest of Cartwright's argument is that it appears to support something resembling the notion of applied science which I have sketched, but without the implicit generality of the basis of pure science one is faced with the same problem of generalising from the particular as faces the 'strict' engineer. This view, of course, is not surprising, nor incoherent, to an empiricist, but there is something odd about the scepticism Cartwright evinces towards general laws and their support for encompassing different models in different sets of circumstances, while she sanctions, for example the notion of an overlap between spheres of the 'application' of laws or models. What can this overlap be but one which is meaningful in a general sense?

Cartwright concludes her paper with the sentence, "Reality may well be just a patchwork of laws". However, what follows more precisely from her arguments is that reality may well be a patchwork of *models*, which is how she characterises the instantiation of local laws. What is interesting about this reformulation is that there is much traffic between the models and reality or between models and models – two-way traffic – and the question still needs asking: How does this take place? How do we move between the different modes? If this 'mobility' was a problem for universal laws, in some sense, then it is a problem for local laws too. The alternative seems to be that they lose their explanatory power and become restatements of the facts.

Better to take the view that knowing the laws of nature does not mean that particular problems can be solved easily or even at all. Pylyshyn (1991b), for example, sees these problems as just what cannot be coped with by scientific theory:

> "No physicist can predict from basic principles even the most rudimentary of everyday occurrences, such as what shape you car will take if it runs off the road, or where leaves will land when they fall off a tree",

but Pylyshyn is making a case against there being a 'theory of practice', not questioning the validity of the general laws of nature. On the other hand if Cartwright is correct about these exclusively *ceteris paribus* laws holding universally, and about the possibility of the *strict* but domain-limited application of certain laws, then scientific theory *is* up to the task of solving *real* practical problems, as long as we redescribe the relationship of pure and applied science. A better understanding of just this relationship would surely be preferable to believing as Pylyshyn does when he writes at the end of his paper, "there is no substitute for wisdom in dealing with real life".

As for Pylyshyn, the systematic success of aeronautical engineering as recounted in detail by Vincenti (1990) is no more dependent on wisdom than the progress of pure science. And Cartwright's mistake is to attempt to draw the boundary around areas in which laws might apply strictly, but without giving us an account of why the boundary should be drawn where it is and what is the connection between what is on each side of that boundary. It is no good appealing to her rights as an empiricist to excuse her this responsibility. For, if she can legitimately find a meaningful and intimate connection between general laws of some kind and what takes place in any limited sphere, she cannot exclude the possibility that there is a nexus between that law and others belonging to other spheres, or some more fundamental law; nor, on the other hand, can the possibility be denied that it is legitimate to accept any normative rules even of limited scope.

The problem which empiricism has with causality and law is that it cannot provide what is distinctive about *explanation*: that it has to be more general than the local sphere of concern. One need not, in order to put forward some such account, resort to a metaphysical cement to join cause and effect nor a notion of metaphysical necessity to support the concept of natural law. This view can be distinguished from the 'fundamentalism' which she is opposing, if the basic laws are seen as playing a constitutive role, and this latter sort of fundamentalism is one which claims only that there are *fundamental* convictions, underpinning immediate facts and phenomenological laws, which correspond with necessary features of scientific and technological representation. So, it is better to accept the difference between theory and practice and rethink the status of theory. Pylyshyn, in this connection, writes that "science is only asymptotically concerned with deep and general theories". Perhaps another way of expressing that would be to say, on analogy with ethical statements and injunctions, that natural laws must be universalis*able* to be meaningfully used, and that this quality of scientific statements is intimately linked to the role they play in explanation – understanding how things fit together in mutual dependence. This perspective would offer at least one way of avoiding head to head conflict with the empiricist.

The important result of a comparison of Pylyshyn with Cartwright is that, if we leave aside the different metaphysical perspectives, the two views generate the same opposition between pure science (however characterised) and technology[93] with respect to the particular and the general. It is, however, this opposition, whatever the rights and wrongs of their views, which I want to account for. There are difficulties, it is true, in reaching agreement of a basic scientific kind, and there remain the difficulties peculiar to solving applied science problems, and well exemplified by Cartwright's position[94]. But it makes more sense in these circumstances to agree with Pylyshyn that there is a divide between the theoretical and the practical and attribute some reality to the theoretical, while accepting that the solution to applied science problems involves pragmatic tailoring of models and theories. This latter position is more cogent when it is considered that explanation depends on a perspective which is

---

[93] She makes an illuminating comparison between the wind-swept 1000 dollar bill and the plane: "the thousand dollar bill *comes as it comes*, and we have to hunt for a model for it. Just the reverse is true of the plane. *We build it to fit the models we know work.* Indeed, that is how we manage to get so much into the domain of the laws we know" (my italics) – a clear distinction at least between applied science and engineering, and also a nice illustration of the domain-constrained problem – whether of applied science or engineering.

[94] She gives a very good example of what I call applied science at work.

more general than the particular phenomenon to be explained; and that pure science tends to seek an ever more comprehensive and less complex overall view of nature.

Some Reflections on Technology

Much has been written on the science/technology axis and how it is composed, but, in the main, it has been done by historians and sociologists. This contribution has been valuable because it has demonstrated the independent progress of both science and technology. Most of the analyses, however, though detailed and informative have failed to establish a rationale for the division of the two activities or disciplines, and some who have supported the division have given relative primacy to one or the other: Layton (1976), suggests that

> "we should broaden our conception of 'science' in order to include technological realities which the conventional theory does not take into account",

thereby rendering science more inclusive but still necessary ; and deSolla Price (1984), whose doctrine of 'artificial revelation' (a paradigm shift of a technological nature which impacts, contrary to convention, on science), writes,

> "In a previous paper I have suggested that science and technology move in linked but independent ways, related like a pair of dancers. Now I argue that what keeps them linked is that both dance to the music of instrumentalities. *Normal science begets more normal science. Normal technology begets more normal technology, but an adventitiously new instrumentality can make for a change in the paradigm within science, and an invention leading to a new innovation within technology*" (my italics).

deSolla Price provides, thus, a stab at a mechanism which might relate the two activities, i.e., instrumentality. This is an interesting idea and not too distantly related to that of practice which I have discerned in Hacking's image of science.

Some writers have dismissed altogether the significance of the dichotomy of science and technology. Mayr (1976), another historian of science and technology, accepts that there appears to be a difference in terms of aim between science and technology but concludes that it is illusory, writing,

> "(t)he aim of science, one might argue, is to explain the riddles of nature; the aim of technology, to solve the material problems of human life. The question about characteristic purpose will indeed furnish a criterion that separates the two concepts in a manner that is sharp, simple, and mutually exclusive. Unfortunately, however, it is valid only on the level of semantics. If we analyze actual historical events, we find that the motives behind actions are usually mixed and complex.".

His conviction that history cannot resolve the mix turns into the claim that there is really no distinction to be made – that any difference is what he calls a 'semantic' one:

> "The words 'science' and 'technology' are useful precisely because they serve as vague umbrella terms that roughly and impressionistically suggest general areas of meaning without precisely defining their limits".

Of technologists, however, who have emphasised, and attempted to examine, the epistemology of engineering are Rogers (1983) and Vincenti (1990). Both uphold a two-stream view. Vincenti's work is the best available systematic presentation of thinking on the science/engineering interaction, including detailed technical analysis via examples in aeronautical engineering, and his aim is to bring engineering on to a level playing-field with science. However, his description of the "combined and interpenetrating activities" of science and engineering is as detailed an account as he gives of the interaction. Rogers provides detailed examples of the development of technology and science and his expression of the difference between the aims of science and technology as the "teleological distinction" is important. He also provides material for what might be characterised as distinctively engineering theoretical knowledge equivalent to that of scientific theory, but peculiar to artefact design or technological investigation.

What is lacking in all these writers' work is a framework to account for the interaction of science and technology (once a gap is asserted) and this framework has to be connected at some level with the issue of separate channels of knowledge

Earlier, I suggested the beginning of an answer to the question of how we could reconcile the two-stream version of science and engineering – each having distinct primary roles (of description and explanation; and prescription and design) with the increasingly common sentiment that engineering too brought knowledge, of some kind, with its practice. I suggested that if scientific knowledge was seen as peculiar to science in the same way as engineering knowledge is thought peculiar to engineering we might perceive better how this reconciliation might be possible. Hacking's conclusions bears this perspective out, i.e., that knowledge is of scientific entities, made sense of in the context of and using the methods of science. If science, therefore, is understood as having its eye on the ever-receding horizon of laws and theory which, however, are not unconnected with its immediate practices and 'confirmations', then its knowledge is no longer solely or centrally the theory which issues from those practices, but the confident collection of claims that we know 'what is the case' when we observe stars through telescopes or metal through an electron microscope. Analogously, engineering, whose ever-receding horizon is the fluent and successful design[95] of artefacts issues in a general knowledge of knowing how to make things work as required, and consequently deepens acquaintance with these objects. The metaphor might be of two vessels, with very different attitudes in the world, separately generating wakes which re-form astern, creating a common inheritance.

Philosophy of Technology

There is no distinct philosophy of engineering. It is not thought to suffer from the same conceptual fallibilities as traditional science does. Of course, this is partly a direct result of the fact that its principal aim is not the generation of knowledge about nature. Hence, if it is not thought of as laying claims to knowledge then less room is left for scepticism.[96] There is, however, some interest in what is called the philosophy of technology, but it is not too different from that which is practised by Hacking. The difference lies in the different provenance of the approach: Hacking starts from problems of scientific realism; and Hackman (1995), for example, and others I have tried to indicate are interested in the history of technology, its theory *and* its practices. And, in a paper which covers the familiar territory of supposing and then rejecting the thesis that technology, which is consequent on theory, secures the theory, Smithurst

[95] In both the case of science and engineering, I mean that the scientific or engineering work done will be novel.

[96] Although but for the tacit dependence on science, its reliability would be questioned.

(1995) [97] cites, if anything the converse view: that science has followed on successful technology (in the case of the external combustion engine), and then proposes a more complex mix in the case of the development of chemistry and electricity (cf. deSolla Price, 1984: "Industrial benefits...came directly from the same techniques that had *produced* the theoretical change."). He concludes his examination of the evidence by considering evolutionary theory and surmising that one can possess a lot of obscure data, and may be able to relate predictive successes "say, that of the dual mandibular structures consequent on the evolution of the auditory ossicles.[98]

> "But without the records, without the aid of the anatomist, and without the technology of palaeontology, there is not much chance that this bit of theory, or much else of it, can be re-established as scientific knowledge. *The link of belief with action is broken.* It is not so much that technology evidences theories, as sets the conditions under which they can truly be scientific theories at all." (my italics).

A little later, he goes on "if technology vanishes, so, for the most part, does pure science; for it is technological detail that locks theories onto experience, and without it we have only stories, an imagination of things". But this is perhaps a little one-sided again, showing the dependence of pure science on technology, but, at the same time, relegating engineering to a non-theoretical realm, as the merely practical side of the equation. Where would engineering be without a world-view which comprised the accepted laws of mechanics, the molecular framework of metal, etc? The fact is that it would not survive the abstraction of this knowledge. It is not solely constituted of practices. That is the difference between engineering and craft. But if this knowledge is not theory as understood in the context of pure science, what is it?

It is knowledge of what underlies things and processes. It is the 'background' to which Wittgenstein is referring to when he writes,

> "All testing, all confirmation and disconfirmation of a hypothesis (and one might add – success or failure of a design) takes place already within a system. And this system is not a more or less arbitrary and doubtful point of departure for all our

[97] In a book entitled "Philosophy & Technology" (in which the author talks about "the *makings* and doings of technology"),

[98] Referring to Stephen Gould's "An Earful of Jaw" from "Eight Little Piggies, Reflections in Natural History", Penguin Books, '94.

> arguments (or rationale): no, it belongs to the essence of what we call an argument. The system is not so much the point of the point of departure, as the element in which arguments have their life". (para 105) (Wittgenstein, 1969) (my parentheses)

It is this "system" in the background which is 'structure'. Features of the "system" may be superseded but there are always essential conditions to the rest of the framework of science – the ground on which we can make assessment of past work and speculation for future projects.

I have looked at some of the ideas of science and technology which seem to displace the conventional emphasis on what can be *known* by science and technology, respectively. After summing up my conclusions on these matters I would like to pass on to an exposition of the foundational framework which should not now be unfamiliar, at least in its broad outlines. But before undertaking this summing-up, I would like to make a slight detour.

Whatever differences and connections there are between science and engineering, it is assumed that when we do one or the other there are no doubts that we are doing the right one in the right circumstances. However, the claim in this thesis is that technologists misuse the knowledge resources they have at their disposal, i.e., that they do science when they should be doing engineering. A corollary of the fact that we may not know what discipline we are practising could be that there are fields of interest about which it cannot be stated unequivocally (in current terms) that they should be the subject of scientific or engineering investigation. I shall come back to this in the concluding chapter, but I would like to repeat an illustration which alludes to this equivocation, before the foundational framework exposition.

Understanding or Engineering Cognition?

Early in the thesis, in Chapter 1, I mentioned the statement by Simon (1969) that, "the proper study of mankind is the science of design". I cite the story below as an indication, still only partial and anecdotal, of what might be meant as engineering knowledge. I shall not comment on it further, nor shall I expatiate on the general question until the final chapter, except to say that the foundational framework not only accommodates such a notion but suggests a reason why some areas of study might be more appropriate than others for an engineering approach.

In 1978, Dennett chaired a meeting which he describes in his book, "Darwin's Dangerous Idea" (Dennett, 1995), at which were present Chomsky, Fodor, Shank and Winograd. The occasion was a panel discussion on the prospects and achievements of AI. Dennett says that it became a team match with Chomsky and Fodor attacking AI and Shank and Winograd defending it. Chomsky's view was that if cognition was simply the interaction of a multitude of "jerry-built gizmos" (as Dennett puts it) then it would mean that cognitive science would not be "interesting" (quoted by Dennett, of Chomsky). For him, either cognitive science is like physics or it is preferable to 'practise' it by reading novels such as those of Jane Austen. More discussion followed, "capped by an observation from Chomsky's colleague at MIT Marvin Minsky" who said, "I think only a humanities professor at MIT could be so oblivious to the third 'interesting' possibility: *psychology could turn out to be like engineering*." (my italics) Dennett applauds: "Minsky had put his finger on it". Dennett goes on to say that there is a distaste felt by some in the humanities for engineering, and this distaste is even more intense when it is employed to tackle the problems of comprehending human cognitive behaviour:

> "Better the mind should turn out to be an impenetrable mystery…than that it should turn out to be the sort of entity that might reveal its secrets to an engineering analysis."

It is my contention that the claim for an underlying epistemological unity of science and technology results not only in the rather negative consequence of getting rid of confusion, contradiction and practical difficulties of discipline communication, but also in the possibility that engineering, in particular, has a *positive* role to play (in the manner hinted at by Minsky) in addition to its proprietary role as the set of disciplines concerned with the systematic work of effective design. As hinted at before, this positive role may be more pertinent to some areas of study than to others. It may, indeed, add to our comprehension of the world – both in its physical and mental aspects – but this is not to say that it is usurping the role of science and becoming involved with description and explanation.

Conclusions on Science and Scientific Knowledge

Hacking (1983) and Smithurst (1995) have provided reasons for believing that the theory which distinguishes engineering from science most significantly is less central to the knowledge with which science endows us; and that the practices which constitute much of science and, perhaps, more of technology, have an epistemological role to play. Nevertheless, how this integration takes place may be made clearer only at a finer level of granularity – in the foundational framework.

It has been helpful to review what the philosophy of science offers in the way of models of scientific work and thought in order to pave the way for some detailed relations between scientific knowledge and engineering knowledge, and I have already made some suggestions along these lines. Before broaching the foundational framework I would like to be sure that no essential features can be undermined, such as the reality of scientific theory. Therefore, there is an important conclusion which I believe we should draw from consideration of Hacking's view, and even from Cartwright's more sceptical and finally, I think, unsatisfactory speculations. This conclusion is that the realism with which we endow increasingly numerous new objects – microscopic and remote in space and time – is, collectively, an accretion of theoretical entities which make up the ground on which we stand, and which, on analogy with Newton's 'shoulders of giants', allow us to see further and in greater detail that of which we and our forbears were utterly ignorant. In this sense, reality and theory are intimately linked: what we know of things is that their mass and their weight are connected, and connected by theory. These attributes of the things we know, of real things, are shot through with theory, it has implications for the *reality* of theory[99]. It does not mean (and in this I agree with Cartwright (Hacking, 1983)) that laws are facts.

[99] This view of the matter might seem surprisingly close to Carroll's, but what I was analysing in the last chapter was the status of this observation in conjunction with its bearing on the *distinct practices* of the two activities, science and design.

The examination of such ideas has been undertaken to flesh out the basic concepts which will form part of my foundational framework. Although I have an opinion about the rightness of their views, I do not insist that the thoughts I have expressed count as absolutely cogent arguments . However, I believe them to be good reasons for reflecting on many received views of what sort of a boundary exists between pure science and applied science, as elsewhere for reflecting on what sort of a boundary there is between science and engineering. This thesis, not being directly concerned with the philosophy of science, is not drawing rigorous conclusions of that sort, but does want to make clear the possibility, in today's conceptual setting of science, of holding its implied opinions: that it should not stand apart from certain respectable and coherent positions, nor should it be treated as a work of the imagination.

Foundational Framework

What follows is intended to serve the purpose of providing a conceptual foundation to support what I believe to be the relationship of scientific and technological (including engineering) knowledge. It is based on ideas which originated in philosophy of language, and I believe this basis is not accidental. The primitive ideas of facts and values, representation and action, are already present in language and an understanding of these elements and how they are related is going to elucidate the 'language' of knowledge.

Postulated Concepts of Science & Technology

The exemplar of a framework for the evaluation of NLD systems was proposed as 'plausible'. It could not be asserted to be correct and I could not, in the nature of the case, assume that using descriptive knowledge for prescriptive purposes would be a rigorous process, since it was one of my premisses that science and engineering were radically different activities. Even if the exemplar were plausible, then more had to be done to show why it could be plausible. This was the reason for wanting to establish what I have called the 'foundational framework'. The elementary concepts out of which this framework is made have been exposed and developed in a generalised form over the last few chapters, and I want to unite them, in a more systematic fashion, with those two notions which have been of central importance throughout the thesis arguments: the 'constitutive' and the 'regulative'. Just as the NLD framework has to make sense in the context of the foundational framework, so the foundational framework cannot be an arbitrary construct, but must make general sense: it must fit in with a view of knowledge, truth and reality. The ideas, therefore, are close cousins of philosophical ideas, but the aim is not to solve philosophical problems. The aim is to pave the way to practical 'operational' solutions via conceptual analysis. However, I want to claim that the framework is consistent with a valid philosophical view, in order avoid the charge of building castles in the air.

The axioms of this foundational framework pose certain problems, which are neither practical to answer adequately nor is it, I believe, relevant to the thesis' main point. I shall, therefore, provide a setting for these axiomatic ideas, but it should be regarded as composed of *reasonable* stipulations. They are consistent most explicitly with J L Austin's reasoning. It may seem odd or coincidental that these ideas have sprung from the theory which was considered in the first part of the thesis – used by the arguments for an NLD framework, but it should be borne in mind that these linguistic theories themselves derived from the philosophy of language (i.e., in which, according to Searle (1969), the problems of philosophy can be solved by examining and understanding human language).

In his book, "How to Do Things with Words" (Austin, 1962), the author sums up part of the motivation for undertaking the lectures which make up the book. He says of the categories of speech acts which he has produced: "They are...quite enough to play Old Harry with two fetishes which I admit an inclination to play Old Harry with, viz. (i) the true/false fetish, (ii) the value/fact fetish". I have shown how Searle has resolved, in the context of speech acts, the value/fact relationship by means of 'institutions', and I made mention of the possibility that the speech act of assertion might raise problems with regard to the idea of truth. Both of those fetishes turn up in the foundational framework, but perhaps the one which bears the most elusive character is the true/false fetish, and I shall start with that one in order to accommodate the elementary characterisation of the constitutive and the regulative.

It is accepted generally that there is relationship of dependence between truth and meaning. Some philosophers believe the former is more fundamental than the latter, some the converse. My view here takes neither side, avoiding the issue by asserting that they are mutually dependent. The reason I can do this is that I have systematised the components of meaning and truth, i.e., reified them in the form of the constitutive and the regulative components because they conform to constitutive and regulative rules.

The so-called 'correspondence theory of truth' is appealing because it seems obvious that if a sentence means something and what it means is actually the case then it is true because what it means corresponds with the facts. However, it also seems to me, that it is equally obvious that it can only mean something if we could envisage the circumstances in which it could be true i.e. it can only mean something if what it means *can* actually be the case, or correspond with the facts. In other words, the theory can only be a good theory if it helps us account for the truth, but if we need the truth to account for the meaning then any correspondence theory of truth is circular. Now there are those who adhere to this theory of truth, among them Searle (1995). Its polar opposite option is widely followed – broadly the pragmatic as opposed to the semantic account – in the form of the 'coherence theory of truth', i.e., that sentences are true because they fit into a conceptual scheme. My framework, not being an articulated philosophical argument, can compromise and provide aspects of both, and perhaps Wittgenstein's observation captures that compromise: "A proposition *shows* its sense. A proposition *shows* how things stand *if* it is true. And it *says* they do so stand." (Wittgenstein, 1961) (the author's emphasis).

Rules

Rules concern behaviour: both how things are done, and how they should be done. There is an ambiguity in the expression 'how things, in general, are' which is partly captured by the implicit reference to norms, i.e., that the expression describes how things, if they are normal or as expected, *should* be[100]. When Long & Dowell (1989)

---

[100] "'John is running' is the statement 'I am stating that John is running': and this statement may depend for its truth on the happiness of 'John is running', just as the truth of 'I am apologizing' depends on the happiness of 'I apologize'." (Austin, '62) So "The cat is on the mat" can be 'happy' or not, and this 'happiness' precedes but

write of the scope of science as comprising the explanation of phenomena it is not only that things are recounted simply as they are – that would be the merest description, but rather that the way they are, once it is declared, is the way they are according to a *norm*[101]. We understand processes because their behaviour follows certain general rules. Equally, if promising, requesting, informing etc. are what they appear to be, they must fulfil certain conditions – the constitutive conditions – and, therefore, must *conform* to constitutive rules. In science, then, the principles, being the fundamental elements support the whole conceptual structure, and are its conditions. Analogously, the propositional components, when they are not asserted but embedded in an illocutionary setting, are the fundamental conceptual structure of the utterance, playing a constitutive role and acting as a "norm of representation' for the possibility of the illocutionary utterance.

Regulative rules start where the others leave off, in the world of intentions and tasks. What is governed only by constitutive rules cannot exist in isolation, except analytically, or in the abstract. We must adopt some attitude towards a state or event before we judge whether the representation is what it needs to be. If we consider a class of acts such as promising, in order to understand what the constitutive rules are of the 'bare' promise, one must understand what it is to be governed by regulative rules, directed towards ends: one must not simply understand how the act conforms to the rule but how the rule is justified (Rawls, 1955)

Likewise, what is governed only by regulative rules cannot exist. This amounts to saying that the rules must apply to something, that an action or state is needed to follow the rule. The rules lie on a continuum from the constitutive to the regulative: from the limit of intelligibility near the purely descriptive end to the instinctive drive at a goal nearly blind to its import. In order to grasp the very meaning of something we

---

is presupposed by its truth. In any event, there is a situation such that it *counts* as 'the cat is on the mat' (the realm of the constitutive) and it may *further* hold or not (the realm of the regulative).

[101] Glock (1996b) interpreting Wittgenstein, writes, "Necessary propositions…do not *follow* from the meanings of signs or from linguistic conventions, they partly *constitute* them, being themselves norms of representation." (his italics, my underlining); and later, "Newton's first law of motion is not an empirical statement, but a norm of representation".

understand the articulation of what is understood to meet some standard; and taking an attitude to it, assertion or another performative mode introduces the regulative. This attitude is itself constitutive when viewed within an institution. Thus, though the rules (or corresponding modes of knowledge) are on a continuum, only with reference to such-and-such a function or purpose can it be said to be one thing or the other.

Representations

Broadly, we can characterise what can be represented by two categories: states; and relations to (or attitudes to) states[102]. As observed, the two are on a continuum from the extreme of mere description to that of mere relation (or attitude). As with the rules, these are limiting conditions which are only approached asymptotically, and each requires the other. Science is directed and determined by the need to understand, that is to say how and why things are as they *are*; engineering with what *ought* to be done to fulfil specific requirements of design. Each activity is composed of elements of the other. So, science needs to follow certain practices (i.e., implicitly rules of procedure: cf. Kuhn (1970b)) to fulfill its aim of understanding[103], and its practices include explanation and prediction, the expression of which have constitutive and regulative components, in their turn. Prediction, for instance, might be that Mercury's apparent position will be different from its actual position by so many degrees, viz. that a certain state will be the case. This claim involves a *commitment* to securing the best method for such an observation – the regulative component. Engineering, to fulfill its aim of satisfying requirements, follows the practices of diagnosis and prescription. Prescription, for instance, might be that, for a given rivet, metal sheets of a certain thickness are needed to avoid failure under load, and these loads in their turn are specified by the work to be done. The application of the regulative rule involves 'matching' the metal sheet with the specifications (Vincenti, 1990), and herein lies the attention to the constitutive rules.

---

[102] I am not here concerned with sense data or other incomplete things: "Talk of subjective qualities comes mainly as a derivative idiom.", and, "Entification begins at arm's length; the point of condensation in the primordial conceptual scheme are things glimpsed, not glimpses." (Quine, 1960).

[103] Koyré writes, "experimentation is a teleological process of which the goal is determined by theory." (Koyré, '68).

Domains

Further, science, in its applied form, while still concerned with understanding the way things are, attempts to solve specific practical problems, usually with practical needs in mind: for example, understanding lightning with a view to avoiding its occurrence in a particular setting. Understanding something under specific circumstances – given certain requirements – defines applied science[104]. In other words, when science is done in this way it is determined in its practices by the domain of its application. Its aim remains explanation and prediction of the physical world. Likewise, engineering, concerned with the production of transformations or artefacts, to fulfill certain requirements and reach some level of effectiveness, is determined by the domain of work. The domain mediates between the semantic and the pragmatic no less in science than in engineering. In cognitive engineering, it is the relationship which exists between the Interactive Work System (IWS) and the domain, i.e., the work being done which establishes the semantics. In design, we gather approximations at the margins of the domain, under the heading of heuristics, in order to manage the pragmatic. Likewise, with applied science, explaining particular phenomena draws on pragmatic devices in order to deal with the complexity of, say, turbulence. Applied science is a term, however, which covers a range of activity from simply using physical formulae to calculate, say, the density of a material to what I intend by a domain-constrained applied science problem (applied science proper). Wisdom (1974) makes a similar distinction and gives illustrations (my 'applied science' is his 'technology') : "The application of Newtonian mechanics to resisting media is applied science; if the medium is *highly specific*, so that we take a special interest in it (such as water because we want to fire torpedoes in it), we move into technology." (my italics).

Therefore, (see Figure 7) viewed in terms of the domain concept, there is science, on the one hand, which could be said to be determined by no domain or have the world,

[104] Vincenti (1990), an engineer who recognises the radical difference between science and engineering repeats Rachel Laudan's observation of "an important point of difference...between science and engineering" that "scientists, in their search for understanding, do not aim at rigidly specified goals. Engineers, to carry out their task of designing devices, must work to very concrete objectives...". The discipline of applied science, unlike 'pure' science applied, also aims at rigidly specified goals. This domain-oriented understanding may encourage the confusion between science and explanation, and engineering and design.

in its totality, as its domain; and, on the other, technology which includes applied science and engineering, which is 'domain-driven'. Alternatively, the division could be between science and applied science, on the one hand, and engineering, on the other, since science and applied science are concerned with understanding while engineering is concerned with fulfilling requirements for the production of artefacts, where 'artefact' comprehends any systematic repeatable change to the world, concrete or otherwise[105]. The latter division depending on the primacy, respectively, of following the constitutive path or the regulative one.

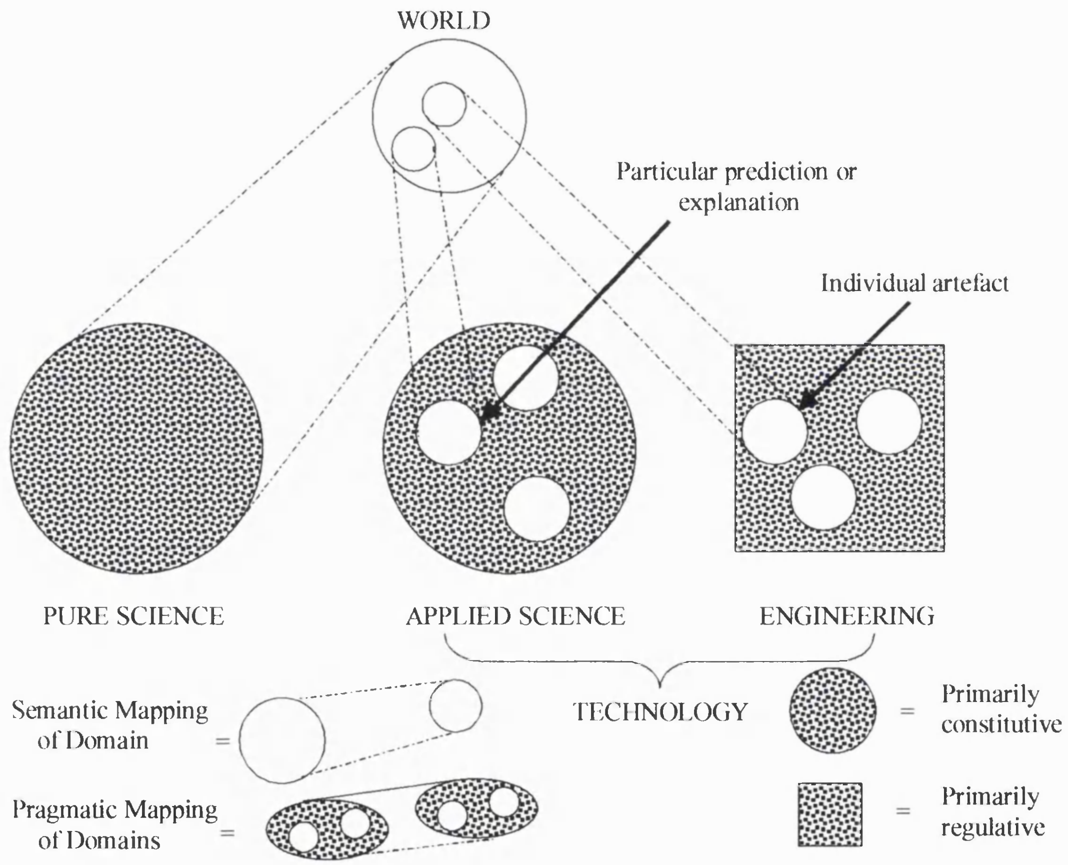


Figure 7: Foundational Framework Schema

To demonstrate the connection between the constitutive/regulative rules and the domain, one could state that a constitutive rule can be applied regulatively, in a domain, *other things being equal* i.e., ignoring the difference the particular domain

[105] Wisdom (1974) writes, "Science is to understand, technology to do. But applied science, though a step on the way to do something, is itself an extension of understanding."

would make. This is the move made by an unthinking adoption of scientific knowledge for design purposes. It might work, but only by accident[106]. So, in contrast with a theoretical explication of, say, 'promising' (a constitutive representation), its application (an instance of promising) could be effected in this *ideal* way, i.e., ceteris paribus: the promisor is aware that the obligation is real, but only if the domain is 'the right one'; something which has to be taken into account.

Fundamental Rules and Representations

This ubiquitousness of the constitutive and the regulative could be represented by a 'Cayley tree' (Figure 8) where the root bifurcates and the bifurcations bifurcate to some practical limit, i.e., defined by the purposes for which the representation is undertaken, with the proviso that one cannot arrive at any elementary constitutive or regulative branch. The root of the tree can be understood as the general object of representation (the world/reality) which possesses the inextricably related properties of'constitutiveness' and 'regulativeness', which separable properties only surface with the appearance of science and engineering.

Structure

The root of the tree might also be seen as structure, and is the ground of science, applied science, and engineering. It is transcendental[107]. Its sense is akin to that of 'reality', and the relationship between structure and structures is analogous to that between reality and what, in particular, is real or considered so. Because it is the paradigm of what *is*, it appears more within the ambit of science than engineering. This interpretation accounts for the assumption that engineering derives from science; that it draws its validity from that of science. On this reading, science is intermediate between reality and practical activity.

This consideration may also account for the tendency of science to slip into metaphysics, and take description to be truly constitutive, thereby plumbing the

---

[106] The Hammond & Allinson attempt to use a scientific theory to express a design rule (Long & Dowell, '89; Dowell, '93).

[107] "An argument is transcendental if it transcends the limits of empirical enquiry, so as to establish the a priori conditions of experience." (Scruton, "Kant", Oxford, '82)

ultimate, sometimes described as answering the 'why' questions as well as the more natural 'how' questions (*this* would be tantamount to the 'fundamentalism' deserving of Cartwright's criticism). As I have suggested, however, the 'thesis' of determinism, for example, is neither a product, nor an assumption (in any explicit sense) of science; and Kant called the closely related concept of causality merely a 'regulative' idea for science. Thus, explanation calls for systematic relations between events and things, and the 'existence' of causality is what allows the practice of explanation and prediction; diagnosis and prescription.

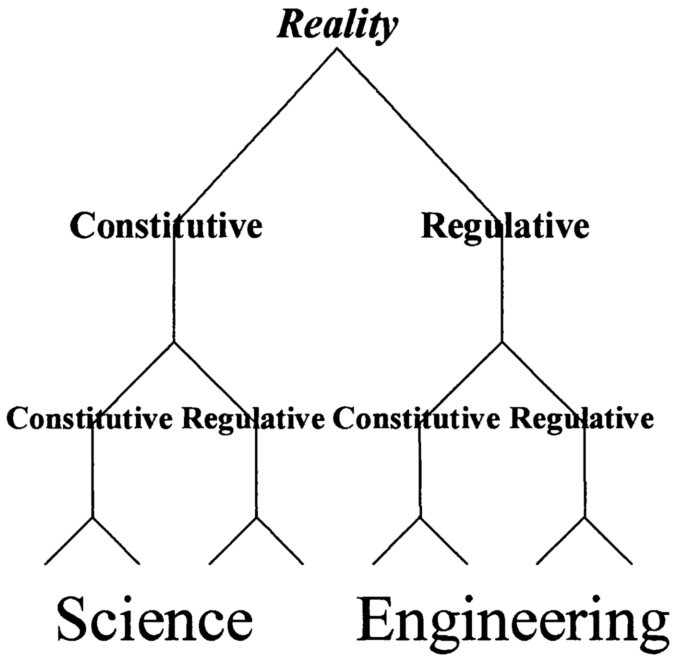


Figure 8: 'Cayley tree'

'Reality', therefore, is not the prerogative of science. What is prescribed and effected is no less of a contribution to grasping 'reality' than what is described and explained. Specifying and reproducing the world would be all the proof necessary for claiming to have understood it[108].

The concept of structure also serves to unite domains. It permits the possibility of generalisation. It is the assumption that something underlies the way in which one can

[108] "The best kinds of evidence for the reality of a postulated or inferred entity is that we can begin to measure it or otherwise understand its causal powers. The best evidence, in turn, that we have this kind of understanding is that we can set out from scratch, to build machines that will work fairly reliably, taking

repeat design procedures in the same domain; as well as work with new but related domains, in a *systematic* fashion. There is a connection here between the laws (the universalisable expressions) as constitutive rules and as structure. Cummins (1984) quotes the 'pendulum law' – the period of the pendulum, **T**, equals **2** times **pi** times the **square root** of the **length** of the pendulum over the gravitational constant, **g** – in order to make the point that, like **f=ma**, the causal relationship is as yet obscure. The expression of the 'pendulum law' is indifferent to the period, T, or the pendulum length being the cause the other being the effect. In this sense the law is a constitutive rule, at bottom. But its interpretation, for scientific work to be done, calculations to be made etc., requires the gradual introduction of regulative rules. The mathematician, however, is often content to view the universe as *essentially* a patterned structure.

Thus, we can attribute meaning to the already quoted remark of Pylyshyn (1991b), that "science is only asymptotically concerned with deep and general theories". *In practice*, in other words, scientific knowledge is never purely theoretical, except as the end-point of its activity, but loosely we consider the calculations of pendulums etc as theoretical workings. Yet they are already "applications of theory", as Kuhn (1970b) has it. There are therefore, he claims, rules to be learned relative to problem-solving, in order to pass from Newton's Second Law (f=ma), which he calls "a law-sketch rather than a law" to other symbolic forms such as those required for free fall, for the pendulum and for coupled harmonic oscillators. He writes that,

> "physicists share few rules, explicit or implicit, by which they make the transition from law-sketch to the specific symbolic forms demanded by individual problems."

Here, even in such intimate contact with the expression of a natural law, are the emergence of regulative rules; and they are regulative with respect to the solution of particular kinds of problems. Kuhn emphasises the difference between the common ownership of such 'knowledge':

---

advantage of this or that causal nexus. *Hence, engineering, not theorizing, is the best proof of scientific realism about entities."* (my italics) (Hacking, 1983)

> "though they shared it, they did not know what it meant and it therefore told them little about nature. What they had yet to learn was not, however, embodied in additional symbolic formulations. Rather it was gained by a process like ostension, the direct exposure to a series of situations each of which, they were told, were (sic) Newtonian."

Structure then is part of the 'background', as Searle (1995) calls it. To continue the parallel with the speech act schema which I have adopted, the structure is the propositional component which is 'subtended' by the illocutionary component. It is not asserted. It is in a sense assumed but this way of putting it does not do it justice. An assumption is often taken to mean some proposition which is tentatively held. In the context of the speech act schema it is not tentatively expressed. Attention can, of course be focused on it and it can be questioned as an asserted proposition, but in its place, as the propositional component, it is not questioned. Likewise, while doing science or engineering, or even going about our daily business, we *take for granted* certain propositional contents. They form part of the ground for the reasonable set of practices which we call science and engineering: part of what I am calling structure, and composed of structures which change in systematic ways.

Conclusions

I began the chapter with a critical look at Dowell's 'two-stream' view of science and technology (particularly engineering): a view with which I agree. My dissatisfaction with his treatment of this view is that, though he acknowledges there is interaction between the two streams he does not express an opinion as to what that interaction can consist in. It seems hard to deny that, in some sense, science and technology have knowledge in common, and that this commonality must play a part in any interaction. However, once the 'two-stream' stance is endorsed, there is no longer any conceptual framework which will allow us to understand how knowledge can be shared. There is also, I am claiming, a further problem with this stance which, I believe, is another consequence of the lack of a proper foundational framework.

As long as engineering was thought of as the application of scientific knowledge the characteristically general knowledge of science could be drawn on at any point. Once the gulf which separates the 'two streams' is introduced, and it is made clear that each design problem must be treated *sui generis*, how can one solution be generalised, without, as Dowell (1993) himself seems to concede, the aid of principles? But what are principles? In the absence of an explanation of them and their role, how can we know that they are not composed of scientific knowledge, and being introduced by the back door? But if we can answer the main question of how knowledge might be shared by science and technology, this subsidiary issue can be resolved.

In Chapter 9, I had suggested a skeletal framework based on components which formed part of the HCIe model. This framework posited a background against which accounts might be made and of which general principles might be predicated, structures. This bare notion would be common to design or explanatory practices. In this chapter, I have tried to clothe this stark concept by promoting a model of science which might allow of a less proprietary attitude to the knowledge of the world; and I have endorsed this model with complementary analyses from a technological perspective. If this is accepted, science and engineering (to take the extremes on the gamut science/technology) are perfectly symmetrical with respect to knowledge of the world. Finally, I completed the body of the chapter with the foundational framework which is a more formal expression of this relationship between practices and knowledge and offers an answer to that main question of how knowledge might be shared by science and technology.

CHAPTER 12

In this, the conclusion to the thesis, I recapitulate some of the difficulties of designing and evaluating natural language processing systems, and how this analysis led to an examination of wider conceptual issues. I then draw together the salient features of the argument, emphasising these wider issues. I do so alluding to the development of ideas about science and technology, pointing out that the reflections of Francis Bacon, at the very start of the modern quest for knowledge, appear to contain what might be considered the essentials of the view argued for in the previous chapters: that we have lost contact with what are arguably his ideas, and we have perhaps fallen prey to a dogmatism similar to the one he faced.

We have passed through a period of approximately four hundred years since Bacon's analysis, and have accomplished things that he could not have imagined. These developments have brought with them questions about how science and technology communicate, and how to formulate public policy with respect to education and research, which his thinking could not have tackled. The ideas which underlie such questions are the ones that have been broached in this thesis. And, in the context of the historical perspective, it might be worth considering the possibility that certain kinds of knowledge may not be achievable by following the 'royal road' of science, but by engineering. The central movement of the thesis is towards a renewed idea of cognitive engineering, the natural consequence of its origins in an operational problem. It is fit that the arguments should issue in a reinforcement of engineering generally (and cognitive engineering in particular) having its own legitimate knowledge, so that the proper and effective relationship between its practices and those of science should be established.

Cognitive Engineering and the Future of Science and Technology

I introduced the thesis claims as a resolution of the operational problem of evaluating a putative Natural Language Dialogue (NLD) system. I construed the problem as an expression of the difficulty of using scientific (descriptive and explanatory) knowledge of language in the pursuit of systematic design (engineering) knowledge of the system as artefact. This problem then formed part of the general question of how scientific knowledge and engineering knowledge are related.

I set out, therefore, to see how descriptive and explanatory knowledge of language might be transformed to design ends. However, in the absence of a good number of worked examples (i.e., 'successfully' implemented systems) a generalised conviction about the validity of such transforms was lacking. The NLD framework was intended to be plausible, and, incidentally, to develop some criteria of knowledge apt for the task of design. The development of that framework thus served those two purposes: (a) to provide some basis from which to operate by suggesting what is required to define the adverb 'successfully'; and (b) in so doing to argue for an important role for the key terms 'constitutive' and 'regulative'. I go on to argue that these two terms could play a more fundamental part in the construction of a foundational framework which would bolster the plausibility of the linguistics/NLD exemplar. Even if the detailed knowledge of linguistics existed and the technological knowledge was also available for exploitation to enable illustrations of transforms, the general argument in favour of an epistemological connection between science and technology might still be needed to overcome conceptual obstacles which I have tried to show are not unimportant or marginal (throughout the thesis but particularly in Chapter 10).

Thus, the first part of the thesis was devoted to addressing what the project to build an NLD system amounted to: that is to say, how it needed linguistic knowledge and yet how this knowledge appeared inadequate to the task. I attempted to describe why the knowledge was not up to the task by illustrating the need for goal-oriented interpretation through the argument of Black & Wilensky (1979). They claim that grammar, if it is going to be helpful in such a context, cannot be simply formal, nor can linguistic meaning rest principally on a straightforward mapping; but that what is meant is closely related to what is *intended*; and that, therefore, purpose plays a central role. It is a natural step from this position to consider the part which planning might play; and, for two reasons, I adopt and adapt Speech Act (SA) theory: (i) that it treats language as action and, thus pre-empts questions of crossing the gap between language as primarily an expression of thought/s and the world; and (ii) that work had already been done on the integration of goals with SA theory in the form of a Plan-based Theory of Speech acts (PBSA) of Cohen & Perrault (1979). I subjected the latter theory to critical examination, concluding that it had to undergo some adaptation in order to widen its application from that of serving as a more or less ad hoc model for NLD design to that of satisfying the needs of a framework for such design (and evaluation). Some work which had been done about the same time (Power, 1979) hinted at the way in which any plan-based view of language should be improved.

I had adopted the terms 'constitutive' and 'regulative' from Searle's (1969) SA theory in order to express what distinguished those resources which were proposed to assist the evaluation of NLD, but failed to do so, from those that genuinely did. These flawed resources satisfied only 'constitutive' rules, characterising by contrast the property they needed to fulfill in order to conform with 'regulative' rules. I extended, therefore, the application of these terms to these two classes of resources treated as classes of knowledge, viz. 'constitutive' knowledge and 'regulative' knowledge. I considered them at this stage as terms of a relative nature, but by the time I came to the second part of the thesis, i.e., the development of a foundational framework, derived from and consistent with the HCIe conception (Dowell & Long, 1989) I began to see them as having a more stable and general use, viz., that of capturing the essential properties of science and engineering.

My initial view of the thesis as a plausible exemplar of the integration of scientific and engineering knowledge *supported* by the a foundational framework is, then, only part of the story. Because the above terms, 'constitutive' and 'regulative', are derived and developed from the NLD framework argument it might seem that the thesis argument as a whole is circular. However, I have also adduced arguments from the philosophy of science and from reflections on the nature of engineering which act as independent corroboration of the foundational framework. The thesis is thus not only an auxiliary framework resting on a foundation; though it is that too, since the possibility of the foundational framework shows that the particular exemplar can be generalised. It is also that the two frameworks each stand as collateral for the other; or to borrow a term from statistics, the thesis is bimodal; so, I claim that the thesis is not two arguments but one.

Development of Thesis Argument in Relation to an Historical Perspective

I prefaced the second half of the thesis with a quotation from Bacon's "Novum Organum" which asserts the convergence (but not the identity) of truth and utility. I chose Bacon for two reasons. The first is because, as Hacking (1983) suggests, he exemplifies a view of scientific knowledge which accords a central position to scientific practices: a view which is contrary to the traditional picture of scientific knowledge as residing in scientific theory, itself the result of the contemplation of nature. Hacking alludes to Bacon's exhortation to "twist the tiger's tail" which hints at more than simple experimentation to confirm theory, but rather an engagement with the world akin to everyday action which provides the experience of reality. In any event, for Hacking, it is consistent with his analysis of scientific knowledge as *intervention* as well as representation. The second reason is that, given the starting point of the thesis as one concerned with design and evaluation, it was apposite that the quotation dealt with the connection between the supposed outcome of science, i.e., truth, and that of design, i.e., utility. Another quotation of Bacon's, mentioned in Rossi (1996), is less pithy but makes clearer the relationship and supports the foundational framework well:

> "The chain of causes cannot by any force be loosed or broken, nor can nature be commanded except by being obeyed. And so those twin objects, human Knowledge and human Power, do really meet in one; and it is from ignorance of causes that operation fails." (Bacon, vol. 4, 1857-74)

For Bacon, as Peltonen puts it in the introduction to "The Cambridge Companion to Bacon" (1996), "the ancient gap between the products of nature and those of human art, had to be bridged". At a new level of analysis, the gap between the 'two streams' should be bridged, and much of Bacon's argument is illuminating in this connection.

Carroll & Campbell (1989) paper, for example, was attempting to bridge the ancient gap between the products of nature and those of human art", but, I argued, only by fudging the dividing line between scientific and engineering disciplines, something which Dasgupta (1991) also failed to wish away . I tried, therefore, to show that (a) the authors in Carroll & Campbell (1989) were not clear about what knowledge was 'embodied', nor how, but also (b) that they sat uncomfortably astride the horns of a dilemma: that the 'design science' is an activity which "does design"; and is in some sense at the same time the process of uncovering an ontology of tasks and artefacts. All these writers, and Simon, whose ideas I examine in the same chapter, are driven by the desire to maximise the interaction between science and engineering, and they appear to do so because without this intimacy with science their practical project of design might either 'degenerate' into an art or, at any rate, would lose the authority and guarantee which science provides. Their arguments' legacy, however, is to confuse the important distinction between the practices of a design discipline and a scientific one. It is this distinction which Long & Dowell (1989), Dowell & Long (1989), and Dowell (1993) have rightly emphasised. In a world in which it is reasonable to assume that like causes produce like effects, i.e., where determinism *as a pragmatic principle* holds, they argue that scientific conclusions cannot be known, beforehand, to be relevant to particular design problems, and that therefore the approach to those design problems has to be expressed in terms of the particular requirements of the problems in question.

Carroll's notion of a cognitive design science in which he too rejects the simple application of experimental psychology is saved from mishap by its determinedly empirical nature: his 'task-artefact' cycle, his scenarios. His practices help him converge on a solution, but they do not vindicate his theoretical stance. As I pointed out elsewhere, Rogers (1983) mentions that 'reactor physics' (a false theory of physics) is what reactor designers employ. So what is vindicated may in fact be wrong. Long and Dowell, in their writings, avoid this consequence. Their view is that the designer in pursuit (ultimately) of engineering knowledge should solve the problem with the minimum of assumptions needed to service the design aim, and moreover, that the outcome of the engineering solution is not a scientific conclusion, nor does it verify a scientific theory. The model, they claim, should be a design model, and refined in the design process; and further, that generalisation of any design knowledge should await the solution of like problems. My task was then to extrapolate those features of Dowell & Long's ontology and address the epistemological issues in their terms. The occasion was the 'operational problem' of the evaluation of NLD.

The aim of addressing this operational problem was not to produce a methodology which might allow the designer to move systematically and automatically from linguistic knowledge to language design. It was to show how the two might be related and to negotiate the difficulties of this relationship, without conflating the descriptive and the prescriptive modes. Having adopted planning as the key ingredient of language *design*, it was also necessary to answer arguments hostile to any planning paradigm serving as a basis for cognitive emulation (typified by Suchman (1987)), before making clearer how planning might be characterised and incorporated in the framework. Further, having concluded that the Plan-based Speech Act theory was adequate only as a measure of consistency (fulfilling the necessary conditions), it was necessary to investigate resources which might be said to answer the need for coherence (fulfilling the sufficient conditions). I took as my example of such a resource, Sperber & Wilson's Relevance theory, and tried to show in what respects it too was inadequate for design purposes. This critique also served as another opportunity to employ the notions of constitutive and regulative knowledge and demonstrate their usefulness in discriminating types of knowledge. I interpreted Sperber & Wilson's theory critically in these terms, and also in those of the philosopher Graham Bird, whose arguments have a more general impact.

The analysis of P-BSA theory and Relevance theory and the kind of knowledge they are each striving for raises the issue of the parallels with the general concepts of semantic and pragmatic; and of their connections with the more particular use of those terms in linguistics. I pursued some resolution of the demarcation disputes which exist between them. My principal aim was not to do so for its own sake, but to provide an introduction to the model of the engineering conception of HCI (Long & Dowell, 1989; Dowell, 1993), showing how it might map those ideas well. The second half of the thesis proposed an interpretation of those other central concepts of the HCIe model, domain and structure, such that the engineering model might be a notion of engineering consistent with a foundational framework which maintains the divide between science and engineering while accounting for the basis for their epistemological interaction. The dangers of an engineering approach of the radical kind proposed by Dowell & Long (1989) and Dowell (1993) are, in my opinion, both apparent and real. The apparent danger is that engineering will lose its secure basis in science; and the real danger is that it will not acknowledge its proper relationship with science. The difference seems to be a subtle one, but it can be expressed roundly as a sharing of a common denominator of knowledge while practising radically distinct activities. At the beginning of modern science and technology, the commonalities were understood by some, and the forces their proponents faced are, in some sense, still influential today. The principal proponent of modern science and technology was Francis Bacon, and the fundamentalist opposition were the Aristotelians.

In some ways, Bacon's real view of science and technology has only been recognised in recent times (Urbach, 1982). He stood not only for the proper grounding of theory, but also for the systematic exploitation of technology; and he saw that these two were connected, i.e., as Farrington (1969) says, "The key to this, in his opinion, lay in the closest possible collaboration between the craftsman and the natural philosopher". What he would get from that was a proper scientific method, and not (automatically) the burgeoning of systematically designed artefacts. However, his view of science brought with it a revision of the meaning of truth and knowledge, such that he could claim their inextricable connection with utility. But until very recently it was widely believed, as Hacking (1983) holds, that scientific knowledge was derived from observation, hypothesis and prediction, where the emphasis was reason and representation. Bacon might have come to see this recent belief as a continuation of the classical Greek tradition: natural philosophy and its proprietorial attitude to knowledge of reality, on which craftsmen, designers and engineers, in present conditions, have to rely for what is known or true. Hacking's (1983) way of expressing the missing component as 'intervention' is what brings Bacon's view of utility into focus. Farrington (1969), emphasises Bacon's attraction to the Presocratic philosophers as the true ancestors of what he saw as the new view of science:

> "They offered, without much sophistication, an explanation of the universe in terms of the familiar operations by which they exercised control over their environment. They drew no distinction in kind between celestial and terrestrial phenomena, but interpreted the more remote, inaccessible, and grandiose phenomena of heaven and earth in terms of the nearer, more intimate, and manageable happenings of their everyday experience".

Here, it is clear that man's scientific knowledge is of a piece with his *experience* of reality, and that although this experience can today be as sophisticated as studying with an electron microscope or manipulating subatomic particles, this emphasis on intervention is what points up the condition of that experience still.

What distinguishes the Presocratic age from the present is the tangled relationship of engineering and pure science. Now our knowledge of the material makeup of things is detailed, and our understanding of the processes involved in material change is complex. We might *imagine* ways of using this knowledge actively to alter how things are to satisfy our needs, but it is widely understood that this assumption is to be treated, at the very least, with caution. I mentioned, in this regard, Pylyshyn's view of the matter expressed in terms of the 'theory-practice gap' (Pylyshyn, 1991b), and elsewhere I drew attention to Cartwright's perspective and how it has influenced her attitude to what scientific knowledge might be (Cartwright, 1996). Bacon's claim was that, unlike the Aristotelians, the Presocratic thinkers realised the intimacy of practices and knowledge. The Aristotelians emphasised the intellectual process of reaching understanding (the origin of word 'theory' was the Greek for 'to observe'). Bacon, by contrast, talked of experiment as playing an equally balanced role with that of reason. Of course, in his time, neither the science nor the technology offered the opportunity of either complex understanding, or complex design or engineering; and so the high-level concerns which exercise Carroll, Long and Dowell would not have arisen for Bacon. He was, however, being prophetic when he rejected both what he saw as the Aristotelian bias and what he saw as the 'empirical', i.e., the *purely* practical bias of the alchemist in favour of his 'modern' view. He expressed this succinctly in the form of a parable:

> "The men of experiment are like the ant; they only collect and use; the reasoners resemble spiders, who make cobwebs out of their own substance. But the bee takes a middle course; it gathers material from the flowers of the garden and the field, but transforms and digests it by power of its own." (quoted by Hacking (1983)

Since neither the practices of science nor those of technology were developed and defined, Bacon's considerations on these matters amounted to a view of knowledge *tout court*. It is not that he anticipated and resolved the present issues as addressed, in part, in this thesis, but that his views unsullied by the present confusions and hostilities offer the vocabulary for supporting such a resolution.

Hacking (1983) makes much of Bacon's manner of expressing his ideas, and finds in it an attitude sympathetic to his own emphasis on 'intervention'; and Hacking himself veers towards the view that proper knowledge is in this sense practical as much as theoretical – not just practical as essential and subservient to theory, but in the sense of practice (engineering) as being on equal terms to theory (pure science). The final paragraph of his book contains the key to his support of my position, quoted in my penultimate chapter: that measurement or other ways of understanding something's causal powers is what amounts to its reality; and our understanding of it is exemplified in our ability to build or reproduce it, i.e., "engineering, not theorizing, is the best proof of scientific realism about entities". One could argue that, given the as yet undeveloped science and engineering, Bacon's view offered the best way forward.

However, perhaps the preponderance of the view that our intellectual heritage was the classical period in Greek history – Platonic and Aristotelian – has maintained the primacy of observation, speculation and theorising as the route to truth, to the detriment of the part played by practices, and particularly to those disciplines in which they are to the fore, such as engineering.

Finally, Bacon's motive for his analysis of what constituted systematic knowledge of nature was not a disinterested one. From his adolescence he was concerned about the fruitlessness of the 'academic' or 'school' knowledge of the times. He believed, not only in knowledge but in progress, and he saw no way this progress could be achieved without a proper scientific method, and what he understood by that was a systematic approach to improvement of what we understood and what we could do practically and industrially. Farrington, in the sub-title of one of his books (Farrington, 1951) refers to Bacon as 'the philosopher of industrial science'. As I suggested above, Bacon did not have to consider science and engineering tightly defined and perhaps at odds with one another. Now we do; I alluded to these difficulties and attempted to suggest connections with the concept of domain which I have adapted from Dowell & Long (1989) and Dowell (1993). In this sense, my approach is a humble continuation of Francis Bacon's project.

Cognitive Science and Cognitive Engineering

The conceptual confusion, which exists between science and engineering and, for which I have tried to provide the means both of extrication and of potentially better and ordered interaction, is at its most confounding when we consider cognitive science and cognitive engineering. It is on this issue that – explicitly or implicitly – many of the conflicts which beset HCI turn. That is to say, they derive from opposing views of what kind of being the human is, or what *true* cognition is. If there is an irony here it is that the opponents of dedicated AI are the descriptivists, and yet they are not prepared to await the outcome of attempts to build robots with 'convincing' cognition.

Conversational Analysis enthusiasts and Quine, an arch behaviourist/pragmatist, are remarkably similar. Both believe in adopting a sceptical position with respect to the manner in which the human comprehends the world. Suchman, who I take as representative of the CA/ethnomethodological position, talks of not being able to track with security/predict the cognition of the world: for her it is bottomless or endless. Likewise, in another area of comprehension or cognition, that of language, Quine (1960) refuses to accept that "Gavagai!" can be interpreted in a determinate way as of a rabbit or various rabbit parts, or a spatio-temporal slice perceived (Goodman, 1965), unless through language (which is not enough if what one is looking for is a solution to the problem of representation). Premack (1986) goes further than my analytic argument by claiming that given this problem for which "psychologists were in their debt", the latter would not need to waste their time

> "by trying to build a learning device that would enable the child to sort through all *logically possible alternatives* before finally settling on the correct one....The psychologists' constructive task was to discover the 'nature' of the constraints." (my italics).[109]

Accordingly, Premack suggests tests with the aid of which, along with the stimulus meaning of a given word, he would be in a position to "specify both the conditions that occasion a word and the interpretation the native puts on the conditions". In this case, the proof, Premack claims, is in the pudding. If CA workers derive their position from a kind of phenomenological description, surely they must accept that arriving at an understanding of a robot's cognition depends simply on observing the way they interact with each other or with us. It is, however, they who appear to be the fundamentalists, whereas it is the engineers who are trying to solve design problems and take a view which is only methodologically sceptical. They, the engineers, are ruling nothing out either epistemologically or ontologically.

There are arguments, therefore, in favour not simply of cognitive science in its practical aspect – building models of cognitive behaviour as a research means – but, of full-blown engineering, since perhaps the aim should be to reproduce cognitive behaviour in order to understand it properly, rather than to 'observe' human cognitive

---

[109] By "psychologist" Premack means at least applied scientists if not cognitive engineers.

behaviour and develop a theory of consciousness. Something like this is what Minsky meant when he said that Chomsky failed to consider a further alternative to the dilemma of psychology as a science aiming at the kind of lawful behaviour which he saw as underlying human linguistic behaviour, i.e., either like physics or without laws and of the kind exemplified by a novelist such as Austen; and that alternative was psychology as engineering (see Chapter 11 and Dennett, 95). There is perhaps a slightly different argument for cognitive engineering. One of the obvious distinctions between physics, say, and psychology, broadly understood, is that one is objectively causal, while the other is subjective and concerned with intention, or goal-oriented. This distinction is expressed as follows by Moore & Carling (1982):

> "We anticipate then that explanation in linguistics will be *teleological* rather than deductive in character. One effect of such a change in the mode of explanation would undoubtedly be that linguistics could no longer compare itself with the more sophisticated of the physical sciences. It might however end the long divergence of linguistics from its subject matter, and allow the field, however slowly to begin to converge with other related fields upon the study of language in operation" (my italics).

Here the emphasis is on the mode of representation of human cognitive behaviour, but we might combine the two approaches and address the field as engineering of cognitive behaviour based on a plan-based model. One might think evolutionary biology a notable example of teleological explanation. It is, however, a causal model, as can be seen from the title "The Blind Watchmaker" – a book by one of its most famous exponents, Richard Dawkins. But it is widely accepted that we have in part taken over our own evolutionary development. Now it can be described as teleological and has become a design issue. Language is more like the latter. Linguistic behaviour, properly represented is goal-driven. It may not be at all clear what these goals are, but that is of its essence. Language is, therefore, an artefact not simply because it is something we *produce* but because it is something we produce *intentionally*; and using the term 'intentionally' broadly means including those intentions which are yet to be discovered. It seems reasonable then to conclude two things: (i) that any further artefact emulating linguistic behaviour should employ a plan-based representation, and (ii) understanding this behaviour might be best served by engineering a replica.

Indeed, as Hacking implies in the quotation cited above, engineering something is the best demonstration that the 'something' in question is understood. When it comes to language, and probably cognitive processes in general, there is an even stronger argument in favour of its being engineered: that it is pointless to attempt to understand linguistic behaviour in a purely scientific way. Perhaps, likewise, the problem of consciousness is not so much a pseudo-problem as some would claim, but a pseudo-*scientific* problem, i.e., better approached as an engineering problem. As quoted in the introduction to the thesis, Simon's view was that "the proper study of mankind is the science of design", and once one has cleared up the equivocation on the concept of science, a clearer more positive view of this remark can be adopted, which is that engineering is a more appropriate approach to capturing the essence of human cognition. Its crucial distinction as an object of study is contained in an eccentric quotation, which I cited in the introductory chapter:

> "The laws that govern these strings of symbols (language, written and spoken), the laws that govern the occasions on which we emit and receive them, the determinants of their content are all *consequences* of our collective artifice." (my parentheses and my italics) (Simon, 1969)

In terms of the foundational framework, we can make sense of this remark through the recognition of the radically 'regulative' substance of cognition. Our understanding of the world is, by contrast, 'constitutive', and science is the re-expression of this kind of understanding. If, however, the principles of cognition are as Simon suggests, and my analysis of language is consistent with that suggestion, then cognition cannot be the subject of what we know as *scientific* knowledge. Instead, only an engineering discipline can explore those regions of reality.

In this connection, it should be noticed that I have used the verb 'understand' in the context of engineering, and yet it is conventional to accept (and I have, in general, kept to the convention) that understanding – the aim of explanation and description – is what science is targeting. This confusion arises because there is (a) knowledge pure and simple, (b) knowledge which is scientific and (c) knowledge which is technological or engineering. Here I am employing the broadest variety, (a), which underpins the others.

I tried to illustrate this common basic knowledge, in the previous chapter, with the simile of the waves of science and engineering converging astern. Understanding or grasping the reality around us can be the immediate *consequence* of either scientific or technological endeavour, but it is not knowledge peculiar to either. The point of this conclusion, and of the historical perspective described above, is that we have perhaps reached a stage in our development when the relative state of science and engineering requires this recognition: one foreshadowed by the reflections of Bacon four hundred years ago.

The claim is not that Bacon had foreseen how science and engineering would develop, but rather that he stands out with his adumbration of a conceptual context within which science and engineering might progress independently yet be epistemologically intimate. Indeed, perhaps his writings are evidence that he underestimated the difficulty of the process of communication between scientific and engineering practices. According to Whitney (1986), Bacon's "method sticks so close to the operative part, the practice of observation and experiment, that generalizations upon which inventions can be based are hardly more than restatements of descriptions of experiments, and *invention mirrors experiment*" (my italics). However, to mirror is not to identify. To complete the section with a more explicit statement of the sentiment contained in the quotation which prefaces the second part of the thesis and is referred to above, the following quotation from the "Novum Organum" should suffice:

> "Human knowledge and human power *meet* in one.... Nature to be commanded must be obeyed; and that which in *contemplation* is as the *cause* is in *operation* as the *rule*" (my italics) (Spedding, Ellis & Heath (eds. & trans.), v4, p47,1857-58).

The two activities of science and engineering are here distinct, but converge, and their products are different but the underlying reality or structure is the same. An important additional feature of this quotation is the recognition that the distinction between 'rule' and 'cause' is determined by the practices of the relevant disciplines.

Bacon's work is sometimes thought of as the prescription of an infallible method, but there is plenty of evidence that Bacon did not see it as such[110]. Whitney (1986) writes, "Bacon recognizes that this goal and method cannot be instituted without a period of muddling through". He writes a little later, referring to Bacon's work, "If Bacon or other investigators make mistakes, they can be easily corrected by those who come after. This confidence about the *possibility* of knowing seems to stand prior to the details of the method..." (my italics). In other words, Bacon is at least partly concerned with the extent and relationship of scientific and technological knowledge – their epistemology.

Epistemology, Science and Engineering

Bacon's view of science as an equal partnership of theory and experiment is reflected in Hacking's (1983) complementary juxtaposition of 'representation' and 'intervention'. They both permit the bringing together of science and engineering, without their conflation. They also break down the barrier between theory and practice. The latter dichotomy is sometimes understood as intra-scientific, and sometimes as coterminous with science, on the one hand, and applied science and engineering, on the other. I think that if we understand the interaction between science, applied science and engineering as I have described it in the last section of the previous chapter (i.e., the foundational framework), where the 'constitutive' elements play the role of 'representation' and the 'regulative' the role of 'intervention', then we can see that, with respect to that framework, though theory and practice are not distinct, neither are they to be confused.

The framework has a further advantage. It allows a better integration of any general epistemological enquiry with that of the philosophy of science. Frequently, problems have arisen about the 'reality' of theory or of theoretical entities, and, in the main, this has been so because the part 'intervention' plays has been ignored or sidelined. Consequently, people have been able to hold perfectly normal views about everyday experience and veracity of perception, while at the same time being deeply sceptical about scientific knowledge. This has led, since scientific knowledge has been

---

110 For example, his reference to the result of the application of his method as a "first vintage".

undermined, to a belief that there may be other avenues to truth, or to a laisser-faire 'anything goes' relativism. If Bacon's and Hacking's view of the matter can be promoted, then the knowledge that flows from science (and technology) can be seen as an extension of everyday experience and the part it plays in our conviction of its reality. The concepts on which the foundational framework are based were derived from philosophical work of a general nature, though employing language as the means of the investigation. If , therefore, these concepts can stand for universal features of knowledge any validity which can be attributed to them in this general capacity can be extended and integrated with that of the foundational framework for pure science, applied science and engineering. This avenue of research could be considered one with value in its own right, but also as further confirmation of the thesis view.

NLD Framework and Other Future Work

In the first part of the thesis, I re-examined the P-BSA theory in the light of the project to relate scientific (descriptive and explanatory) knowledge of language to the general features of design knowledge of NLD. I found that I had to refine the properties of the work done by Cohen & Perrault (1979), and I borrowed some of the ideas which might help that process from Power's perspective on planning (1979), and the analysis done by Ramsay (1990) of epistemic planning. The refinement took the form of extending planning to include attitudes of a personal, social, national, cultural etc. nature. The intention was not to assert that such a technical specification would be required in any design of NLD, but instead to claim that any valuable design specification could be expressed in those general terms. Whatever the notions employed in NLD to date, such as frames, schemata etc., all may be translated into goal-oriented characterisations, and consequently expressed as plans. Of course, semantic networks or conceptual dependency notation might be thought of as simply defining or describing the possible links which could be made by an agent, but their representation without any *motivation* to do one thing or another is useless. The reference earlier in the chapter to Premack's comments on Quine's sceptical observations implies that the designer is not interested in logical possibility as such, but on the de facto constraints; and these constraints include motivation.

Habermas (1979) has put forward a schema which succinctly expresses the properties of language as speech. He has developed this schema from the SA theory of Austin and Searle. It lists domains of reality predicated of speech acts; basic attitudes to those domains and their modes of communication; the criteria for the validation of these attitudes; and the general functions of speech referred to by these attitudes. This listing would be one way of describing the framework for NLD. The domains, attitudes, and criteria would underlie the use of language and be the basis for an analysis to meet the requirements of any given NLD project. Of course, these projects will, for the time being, be of very limited scope. However, the virtue of a framework is that different projects can be related to the large scheme and the principles of the projects' design can be uniform. Much of any design specification based on the concepts supported by this framework would be ad hoc, but gradually these ad hoc features should give way to a more principled and coherent approach. The potential for future work, therefore, on such a scheme is great. The era of speech understanding systems is just beginning and this work on the development of the framework should progress in parallel with these projects, gathering data and producing guidelines and methods.

Finally, in connection with the NLD framework, it might be relevant to emphasise that, consistent with the argument earlier (in this chapter, and hinted at in the previous one) that language is an appropriate target for engineering as it converges on science and has its own consequences for increasing understanding, Habermas distinguishes the study of language from the traditional sciences referring to it as a 'reconstructive science', and contrasting it with the 'empirical/analytic sciences. He points to methodological difficulties inherent in the latter approach to language, and he writes, when he considers whether to take a transcendental approach, that "behind the terminological question, there stands the as-yet insufficiently clarified status of non-nomological (physics would be 'nomological') empirical sciences of the reconstructive type". I would like to suggest that this 'unclarified status' is precisely whether they should be researched as science or engineering[111].

---

[111] In this connection, I am reminded of Simon's enigmatic remark, introduced in the first chapter and quoted again above.

Conclusion: Cognitive Engineering and the Foundational Framework

At the intersection of the first and second part of the thesis lies the chapter on the model which was the product of the 'conception of HCI as an engineering discipline' (HCIe) and the NLD framework: a mapping of the key features of the framework, identifying and elaborating thereby 'structure' and 'domain', and drawing out the connection between the linguistic properties of 'semantic' and 'pragmatic' which characterise the interactions of the 'interactive worksystem' (IWS) and the domain. And, in the second part of the thesis, in order to pursue the aim of better expressing the epistemological relationship of engineering and science, I explicitly broadened the extension of the concepts of 'domain' and 'structure' to adapt them to the purpose of integrating the epistemology of science and technology.

The argument, then, for the foundational framework, with respect to the HCIe model is meant to serve as corroboration. It should show how the essential 'ontological' properties of the model can be epistemologically coherent; and how this coherence, at the same time, justifies the juxtaposition and exploitation of basic science. In this connection, some points of contrast with the formulation of the design problem as defined by Dowell (1993) were made. My exposition of the HCIe model is slightly different, therefore, from that of Dowell & Long (1989) and Dowell (1993). However, it is, I believe, the same in spirit. Thus, I consider that my aim of providing an epistemological treatment of this more ontological model has been fulfilled.

In certain respects, this is the central achievement: to have put forward an epistemological interpretation of HCIe. The NLD framework and the foundational framework are, respectively, 'plausible' and speculative and affirm each other to a degree. They should be cogent in their argument, but they both require considerable further work. They also serve the purpose of providing bulwarks for the important project of developing cognitive engineering as a core discipline of future technology; and, in particular, the further development of the NLD framework should be considered important research work subservient to that discipline.

REFERENCES

Alexander, C. (1964) "Notes on the Synthesis of Form", Harvard University Press,

Allen, J. F. (1979) "A Plan-based Approach to Speech Act Recognition" unpublished PhD thesis, Univ. Toronto

Allen, J. F. & Perrault, C. R. (1980) "Analyzing Intention in Utterances", *Artificial Intelligence*, v15, no. 3, pp143-178

Austin, J. L. (1961) "Ifs and Cans" in "Philosophical Papers" ed. by Urmson & Warnock, Oxford University Press, pp153-180

Austin, J. L. (1962) "How to Do Things with Words", Oxford University Press

Austin, J. L. (1971) "Performative-Constative" (trans. from the French by G J Warnock) in "The Philosophy of Language" ed. by Searle, Oxford University Press, pp13-22

Bacon, F. (1857-74) "The Works of Francis Bacon" in 14 vols., ed. by Spedding, Ellis & Heath, London: Longman

Barwise, J. & Perry, J. (1983) "Situations and Attitudes", Bradford/MIT Press

Bird, G. H. (1994) "Relevance Theory and Speech Acts" in "Foundation of Speech Act Theory" ed. by Tsohatzidis, Routledge, pp292-311

Black, J. B. & Wilensky, R. (1979) "An Evaluation of Story Grammars" *Cognitive Science* v3, pp213-229

Black, M. (1964) "The Gap between 'Is' and 'Should'", *Philosophical Review*, v73, pp165-181

Bresnan, J. & Kaplan, R. M. (1984) "Grammars as Mental Representation of Language" in "Methods and Tactics in Cognitive Science" ed. by Kintsch, Miller & Polson, Lawrence Erlbaum Associates, pp103-135

Bruce, B. C. (1975) "Generation as Social Action" in "Readings in Natural Language Processing" *Proceedings of the Conference on Natural Language Processing*, Cambridge, MA, pp419-422

Bruce, B. C. & Schmidt, C. F. (1974) "Episode Understanding and Belief Guided Parsing", presented at the *Association for Computational Linguistics* Meeting at Amherst, Mass. (July 26-27)

Carberry, S. & Pope, W. A. (1993) "Plan Recognition Strategies for Language Understanding", *International Journal of Man-machine Studies*, v39, pp 529-577

Carroll, J. M.(1993) "Creating a design science of human-computer interaction", *Interacting with Computers*, v5, no1, pp3-12

Carroll, J. M. (1995) "Artefacts and Scenarios: an Engineering Approach", in "Perspectives on HCI – Diverse Approaches", ed. by Monk & Gilbert, Academic Press

Carroll, J. M. (1997) "Human-computer interaction: psychology as a science of design", *International Journal Human-Computer Studies*, v46, pp501-522

Carroll, J. M., & Campbell, R. L. (1989) "Artefacts as psychological theories: the case of human-computer interaction", *Behaviour and Information Technology*, v8, no4, pp247-256

Carroll, J. M., Singley, M. K. & Rosson, M. B.. (1992) "Integrating theory development with design evaluation", *Behaviour & Information Technology*, v11, no5, pp247-255

Cartwright, N. (1996) "Fundamentalism versus the Patchwork of Laws" in "The Philosophy of Science" ed. by Papineau, Oxford University Press, pp314-326

Cartwright, N., Cat, J., Fleck, L. & Uebel, T. E. (1996) "Otto Neurath: Philosophy between Science & Politics", Cambridge University Press

Clark, H. H. (1982) "The Relevance of Common Ground: comments on Sperber & Wilson's Paper" in "Mutual Knowledge" ed. by Smith, Academic Press, pp 124-127

Clark, H. H. (1996) "Using Language", Cambridge University Press

Clark, H. H. & Carlson, T. B. (1982) "Speech Acts and Hearers' Beliefs" in "Mutual Knowledge" ed. by Smith, Academic Press, pp1-59

Clark, H. H. & Malt, B. C. (1984) "Psychological Constraints on Language: A Commentary on Bresnan & Kaplan and Givon" in " Methods and Tactics in Cognitive Science", ed. by Kintsch, Miller & Polson, Lawrence Erlbaum Associates, pp191-214

Clark, H. H. & Marshall, C. R. (1981) "Definite Reference and Mutual Knowledge" in "Elements of Discourse Understanding" ed. by Joshi, Webber & Sag, Cambridge University Press, pp10-63

Cohen, P. R. & Levesque, H. J (1980) "Speech Acts and the Recognition of Shared Plans", *Proceedings of 3rd. Biennial Conference*, Victoria, British Columbia, pp 263-271

Cohen, P. R. & Levesque, H. J. (1990) "Rational Interaction as the Basis for Communication" in "Intentions in Communication" ed. by Cohen, Morgan & Pollack, MIT Press, pp221-255

Cohen, P. R. & Perrault, C. R. (1979) "Elements of a Plan-based Theory of Speech Acts", *Cognitive Science*, v3, pp177-212

Cohen, P. R., Perrault, C. R., & Allen, J. F. (1982) "Beyond question answering" in "Strategies for Natural Processing" ed. by Lehnert & Ringle, Lawrence Erlbaum Associates, pp245-274

Cummins, R. (1984) "The Nature of Psychological Explanation", MIT Press

Dascal, M. (1994) "Speech act theory and Gricean pragmatics" in "Foundations of Speech Act Theory" ed. by Tsohatzidis, Routledge, pp323-334

Dasgupta, S. (1991) "Design Theory and Computer Science", Cambridge University Press

Dennett, D. (1990) "Cognitive Wheels: The Frame Problem of AI" in "The Philosophy of Artificial Intelligence" ed. by Boden, Oxford University Press, pp147-170

Dennett, D. (1995) "Darwin's Dangerous Idea: evolution and the meanings of life", Allen Lane: The Penguin Press

deSolla Price, D. J. (1984) "Notes towards a Philosophy of the Science/Technology Interaction" in "The Nature of Technological Knowledge" ed. by R Laudan, D Reidel Publishing Co., pp105-114

Deutsch, B. (1974) "The Structure of Task-oriented Dialogues", contributed paper in the IEEE Symposium on Speech Recognition, ed. by Erman, pp250-253

Devlin, K. (1997) "Goodbye Descartes", John Wiley & Sons Inc.

Dowell, J. (1993) "Cognitive Engineering and the Rationalisation of the Flight Strip", unpublished PhD thesis, University of London

Dowell, J. (1995) "Interacting with Domains: taking the domain seriously in Cognitive Ergonomics", *European Association of Cognitive Ergonomics*, v2 no. 2, pp17-22

Dowell, J & Long, J. (1989) "Towards a conception for an engineering discipline of human factors", *Ergonomics*, v32 no. 11, pp513-1535

Dowell, J. & Long, J. (1998) Peer Review Target Paper, "Conception of the Cognitive Engineering Design Problem", *Ergonomics*, v41 no. 2, pp126-139

Dowell, J., Smith, W. & Pigeon, N. (1998)"Design of the natural: an engineering process for naturalistic decision-making" ed. by Flin, Salas, Strub & Martin "Decision-making under stress: emerging themes and applications", Ashgate: Aldershot, pp126-136

Dreyfus, H. (1985) "From micro-worlds to representation: AI at an impasse" in "Readings in Knowledge Representation" ed. by Brachman & Levesque, Morgan Kaufmann, pp71-94

Dummett, M. (1973) "Frege: Philosophy of Language", Duckworth

Dummett, M. (1989) "Language and Communication" in "Reflections on Chomsky" ed. by George, Blackwells, pp192-212

Farrington, B. (1951) "Francis Bacon: Philosopher of Industrial Science", Lawrence & Wishart Ltd., London

Fikes, R. E. & Nilsson, N. S. (1971) "STRIPS: A New Approach to the Application of Theorem Proving to Problem Solving", *Artificial Intelligence*, v2, pp189-208

Fodor, J. A. (1983) "The Modularity of Mind", MIT Press, Mass.

Fraser, N. (1991) "Corpus-based Evaluation of the Sundial System", Natural Language Processing Systems Evaluation Workshop, University of California

Fraser, N. M. & Gilbert, G. N. (1991) "Simulating Speech Systems", *Computer Speech and Language*, v5, pp81-99

Frisch, A. M. & Perlis, D. (1981) "A Re-evaluation of Story Grammars", *Cognitive Science*, v5, pp79-86

Galliers, J. R. & Sparck Jones, K. (March, 1993) "Evaluating Natural Language Processing Systems", Technical Report No 291, University of Cambridge Computer Laboratory

Garnham, A. (1983) "What's Wrong with Story Grammars", *Cognition*, v15, pp145-154

Garnham, A. (1985) "Psycholinguistics: Central Topics", Routledge

Gaythwaite, D. M. (in press) "Intellectual Property and Technical Know-how", in "Biotechnology – the Science and the Business" ed. by Moses & Springham, Harwood Academic Publishers: New York & Zurich

Gazdar, G. (1979) "Pragmatics: Implicature, Presupposition and Logical Form", Academic Press

Gibbons, M. (1983) "Is science industrially necessary?" in "Science, Technology and Society Today", ed. by Gibbons & Gummett, Manchester University Press, pp96-116

Gibson, J. J. (1977) "The Theory of Affordances" in "Perceiving, Acting and Knowing: towards an ecological psychology" ed. by Shaw & Bransford, New Jersey: Erlbaum

Givon, T. (1984) "Deductive versus Pragmatic Processing in Natural Language" in "Methods and Tactics in Cognitive Science" ed. by Kintsch, Miller & Polson, pp137-190

Glock, H-J. (1996a) "A Wittgenstein Dictionary", Blackwell, Oxford

Glock, H-J. (1996b) "Necessity and Normativity" in "The Cambridge Companion to Wittgenstein", ed. by Sluga & Stern, Cambridge University Press, pp198-225

Good, D. (1990) "Repair and Cooperation in Conversation" in ""Computers and Conversation" ed. by Luff, Gilbert & Frohlich, Academic Press, pp133-150

Goodman, N. (1965) "Fact, Fiction and Forecast", Indianapolis: Bobbs-Merrill

Grice, H. P. (1957) "Meaning", *Philosophical Review*, v66, pp377-388

Grice, H. P. (1975) "Logic and Conversation" in "Syntax and Semantics" v3 ed. by Cole & Morgan, pp41-58

Grosz, B. (1981) "Focusing and Description in Natural Language Dialogues" in "Elements of Discourse Understanding" ed. by Joshi, Webber & Sag, CUP, New York, pp84-105

Grosz, B. J. & Sidner, C. L. (1990) "Plans for Discourse" in "Intentions in Communication" eds. Cohen, Morgan & Pollack, MIT, pp417-444

Gutting, G. (1984) "Paradigms, Revolutions, and Technology" in "The Nature of Technological Knowledge" ed. by Laudan, D. Reidel Publishing Co., pp47-65

Habermas, J. (1979) "What is Universal Pragmatics?" in "Communication and the Evolution of Society", translated by McCarthy, Beacon Books Boston, pp1-68

Hacking, I. (1983) "Representing and Intervening", Cambridge University Press

Hackmann, W. (1995) "Instrument and Reality" in "Philosophy & Technology" ed. by R Fellows, Cambridge University Press, pp29-51

Hammond, N. & Allinson, L. (1988) "Development and Evaluation of a CAL System for Non-formal Domains: the Hitchhikers Guide to Cognition", Computer Education, v12, pp215-220

Harris, R. (1996) "Sign, Language and Communication", Routledge

Haslett, B. J. (1987) "Communication: strategic action in context", Lawrence Erlbaum Associates

Haugeland, J. (1988) "Semantic Engines: an Introduction to Mind Design" in "Mind Design" ed. by Haugeland, MIT Press, pp1-34

Hayes, P. J. (1985) "Some Problems and Non-Problems in Representation Theory" (Proceedings AISB Summer Conference, University of Sussex, 1974) in "Readings in Knowledge Representation", ed. by Brachman & Levesque, Morgan Kaufmann, pp3-21

Hewitt, C. (1969) "PLANNER: A Language for Proving Theorems in Robots", 1st Int. Joint Conference on AI, Washington

Hobbs, J. R. (1995) "Sketch of an ontology underlying the way we talk about the world", *International Journal of Human-Computer Studies*, v43, pp819-830

Hornsby, J. (1994) "Illocution and its Significance" in "Foundations of Speech Act Theory", ed. by S L Tsohatzidis, Routledge, pp107-207

Johnson-Laird, P. N. (1983) "Mental Models" Cambridge University Press

Kant, I. (1964) "The Critique of Pure Reason", trans. by Norman Kemp Smith, Macmillan Papermac

Koyré, A. (1968) "Metaphysics and Measurement", Chapman & Hall

Kuhn, T. (1970a) "The Structure of Scientific Revolutions", University of Chicago Press

Kuhn, T. (1970b) "Reflections on my Critics" in "Criticisms and the Growth of Knowledge", ed. by Lakatos & Musgrave, pp231-278

Laudan, R. (ed) (1984) "The Nature of Technological Knowledge: Are Models of Scientific Change Relevant?", Dordrecht

Layton, E. (1974) "Technology as Knowledge", *Technology & Culture*, v15, pp31-41

Layton, E. (1976) "American Ideologies of Science and Engineering", *Technology & Culture*, v17, pp688-700

Levinson, S. (1983) "Pragmatics", Cambridge University Press

Long, J. (1996) "Specifying the Relations between Research and the Design of Human-Computer Interactions", *International Journal of Human Computer Studies*, v44, no. 6, pp875-920

Long, J. & Dowell, J. (1989) "Conception of the disciplines of HCI: craft, applied science, and engineering" in Sutcliffe & Macaulay (eds) "Proceedings of the Fifth Conference of the BCS HCI SG", Cambridge University Press

Lyons, J. (1991) "Chomsky", Fontana 'Modern Masters', Harper Collins, London

Mayr, O. (1976) "The Science-Technology Relationship as a Historiographic Problem", *Technology & Culture*, v17, no. 4, pp663-673

McCarthy, J. & Hayes, P. J. (1969) "Some Philosophical Problems from the Standpoint of AI" in Meltzer & Mitchie "Machine Intelligence 4", pp463-502

McCarthy, J. & Monk, A. (1994) "Channels, conversation, cooperation and relevance: all you wanted to know about communication but were afraid to ask", *Collaborative Computing*, v1, pp35-60

Minsky, M. (1975) "A Framework for Representing Knowledge" in "The Psychology of Computer Vision", ed. by Winston, McGraw-Hill, New York

Monk, A., Carroll, J. M., Harrison, M., Long, J. & Young, R (1990) "New approaches to theory in HCI: How should we judge their acceptability?" Panel discussion in *Proceedings of INTERACT'90, Cambridge*, Elsevier North-Holland, pp1055-1058

Moore, T. & Carling, C. (1982) "Understanding Language: towards a Post-Chomskyan Linguistics", Macmillan

Nickerson, R. S. (1976) "On Conversational Interaction with Computers", *Proceedings of ACM/SIGGRAPH Workshop*, Oct 14- 15, Pittsburgh, PA, pp 101-113

Novick, D. (1988) "Control of Mixed-initiative Discourse through Meta-locutionary Acts: A Computational Model", unpublished PhD thesis, University of Oregon

Passmore, J. P. (1985) "Recent Philosophy", Duckworth, London

Peltonen, M. (ed) (1996) "The Cambridge Companion to Bacon", Cambridge University Press

Perlis, D. Purang, K. & Andersen, C. (1998) "Conversational Adequacy: Mistakes are the Essence", *Int. J. Human-Computer Studies*, v48, pp553-575

Perrault, C. R. Allen, J. F. & Cohen, P. R. (1978) "Speech Acts as a Basis for Understanding Dialogue Coherence" in *Proceedings of the 2nd Conference on Theoretical Issues in Natural Language Processing*, Champaign-Urbana, Illinois

Perrault, C. R. & Allen, J. F. (1980) "A Plan-based Analysis of Speech Acts", *American Journal of Computational Linguistics*, v6, pp167-182

Polanyi, M. (1956) "Pure and Applied Science and Their Appropriate Forms of Organization", *Dialectica*, v10, no. 3, pp231-242

Polanyi, M. (1962) "Personal Knowledge", University of Chicago Press,

Polson, P. G., Miller, J. R. & Kintsch, W. (1984) "Methods and Tactics Reconsidered" in " Methods and Tactics in Cognitive Science", ed. by Kintsch Miller & Polson, Lawrence Erlbaum Associates, pp277-296

Popper, R. (1982) "The Open Universe: An Argument for Indeterminism", Hutchinson

Power, R. (1974) "A computer model of conversation" PhD thesis, University of Edinburgh

Power, R. (1979) "The Organisation of Purposeful Dialogue" *Linguistics* 17 pp107-152

Premack, D. (1986) "Gavagai!", A Bradford Book, MIT Press

Pylyshyn, Z. (1991a) "Rules and Representations" in "The Chomskyan Turn" ed. by Kasher, Blackwell, pp231-251

Pylyshyn, Z. (1991b) "Some Remarks on the Theory-Practice Gap" in "Designing Interaction" ed. by J Carroll, Cambridge University Press, pp39-49

Quine, W. V. O. (1960) "Word and Object", MIT Press

Ramsay, A. (1990) "The Logical Structure of English", Pitman

Rawls, J. (1955) "Two Concepts of Rules", *The Philosophical Review*, v64, pp3-32

Rawls, J. (1971) "A Theory of Justice", Oxford University Press

Rogers, G. F. C. (1983) "The Nature of Engineering: a Philosophy of Technology", Macmillan

Rossi, P. (1996) "Bacon's Idea of Science" in "The Cambridge Companion to Bacon", ed. by Marrkku Peltonen, Cambridge University Press, pp25-46

Rumelhart, D. (1975) "Notes on a Schema for Stories" in "Representation and Understanding: Studies in Cognitive Science" , ed. by Bobrow & Collins, Academic Press

Russell, B. (1905) "On Denoting", *Mind*, v14, pp479-493

Sacerdoti E. (1977) "A Structure for Plans and Behavior", New York: Elsevier

Saunders, J. B. (1969) "Words and Phrases: Legally Defined", v4 'O-R', Butterworth

Schegloff, E. A. (1988) "Presequences and Indirection: applying Speech Act Theory to Ordinary Conversations", *Journal of Pragmatics*, v12, pp55-62

Schegloff, E. A. & Sacks, H. (1973) "Opening up Closings", *Semiotica*, v7, pp289-327

Schiffer, S. R. (1972) "Meaning", Oxford University Press

Schwayder, D S (1994) "A Semantics of Utterance, Formalized" in "Foundations of Speech Act Theory", ed. by Tsohatzidis, Routledge, pp80-98

Schwyzer, H. (1969) "Rules & Practices", *Philosophical Review*, pp451-467

Scriven, M. (1967) "The Methodology of Evaluation", in Tyler, Gagne & Scriven "Perspectives of Curriculum Evaluation", Rand McNally, Chicago, pp 39-101

Searle, J. (1965) "What is a Speech Act?" in "Philosophy in America", ed. by M Black, Allen & Unwin Ltd., pp221-239

Searle, J. (1969) "Speech Acts", Cambridge University Press

Searle, J. (1983) "Intentionality: an Essay in the Philosophy of Mind", New York, Cambridge University Press

Searle, J. (1995) "The Construction of Social Reality", Allen Lane: The Penguin Press

Shanon, B. (1993) "The Representational and the Presentational", Harvester Wheatsheaf

Shimanoff, S. B. (1980) "Communication Rules: Theory and Research" Sage Publications Ltd

Simon, H. (1969) "The Sciences of the Artificial", MIT Press

Simon, H. (1980) "Cognitive Science: the Newest Science of the Artificial", *Cognitive Science*, v4, pp33-46

Simon, H. (1996) "Models of my Life", MIT Press

Smithurst, M. (1995) "Do Successes of Technology Evidence the Truth of Theories" in "Philosophy & Technology" ed. by Fellows, Cambridge University Press, pp19-28

Spedding, J., Ellis, R. L., & Heath, D. D. (eds. & trans.) (1857-74) "The Works of Francis Bacon", 14 vols., London, Longman & Co.

Sperber, D. & Wilson, D. (1982) "Mutual Knowledge & Relevance in Theories of Comprehension" in "Mutual Knowledge", ed. by Smith, Academic Press, pp61-85

Sperber, D. & Wilson, D. (1986; 2nd. Edition 1995) "Relevance: Communication & Cognition", Blackwell, Oxford

Sperber, D. & Wilson, D. (1987) "Precis of Relevance: Communication and Cognition" in *Behavioural and Brain Sciences*, v10, pp697- 794

Stalnaker, R. C. (1972) "Pragmatics" in "Semantics of Natural Language" ed. by Davidson, Reidel, Dordrecht-Holland, pp380-397

Stalnaker, R. C. (1987) "Inquiry", MIT Press

Strawson, P. (1950) "On Referring", *Mind*, v59, pp320-344

Strawson, P. (1966) "The Bounds of Sense", Methuen

Strawson, P. (1964) "Intention and Convention in Speech Act", *Philosophical Review*, v73 n.4, pp 429-460

Strawson, P. (1973) "Meaning, Truth, and Communication" in "Linguistics at Large", ed. by N Minnis, Paladin, pp 91-110

Suchman, L. (1987) "Plans and Situated Actions" Cambridge University Press

Urbach, P. (1982) "Francis Bacon as a Precursor to Popper", *British Journal of the Philosophy of Science*, v33, pp113-132

Vincenti, W. G. (1990) "What Engineers Know and How They Know it", John Hopkins,

Walker, R. C. S. (1989) The third review of a Multiple Review of "Relevance: Communication and Cognition", *Mind & Language*, v4 nos. 1 & 2, pp151-159

Waltz, D. L. (1982) "The State of the Art in Natural-Language Understanding" in "Strategies for Natural Language Processing" ed. by Lehnert & Ringle, Lawrence Erlbaum Associates, pp3-32

Whitney, C. (1986) "Francis Bacon and Modernity", Yale University Press,

Winograd, T. (1972) "Understanding Natural Language" New York: Academic Press

Winograd, T. (1975) "Frame Representations and the Declarative/Procedural Controversy" in "Representation and Understanding" ed. by Bobrow & Collins, Academic Press, pp185-210

Winograd, T. (1985) "Moving the Semantic Fulcrum", *Linguistics and Philosophy*, v8, pp91-104

Winograd, T. & Flores, F. (1986) "Understanding Computers and Cognition", Addison-Wesley

Winston, H. W. (1984) "Artificial Intelligence", Addison-Wesley Publishing Co.

Wisdom, J. O. (1974) "The Need for Corroboration" (comments on J Agassi's paper in the same publication) in "Contributions to a Philosophy of Technology", ed. by Rapp, D. Reidel, Dordrecht-Holland, pp64-68

Wittgenstein, L. (1953) "Philosophical Investigations", Basil Blackwell, Oxford

Wittgenstein, L. (1958) "The Blue and Brown Books", Basil Blackwell, Oxford

Wittgenstein, L. (1961) "Tractatus Logico-Philosophicus", Routledge & Kegan Paul

Wittgenstein, L. (1969) "On certainty", ed. by Anscombe and von Wright, Basil Blackwell, Oxford

Wittgenstein, L. (1975) "Philosophical Remarks", ed. by Rhees, Basil Blackwell, Oxford

Woods, W. A. (1970) "Transition Network Grammars for Natural Language Analysis", *Communications of the ACM*, v13, pp591-606

Woods, D. D. & Roth, E. M. (1988) "Cognitive Systems Engineering" in "The Handbook of HCI", ed. by Helander, Elsevier, North Holland, pp3-43