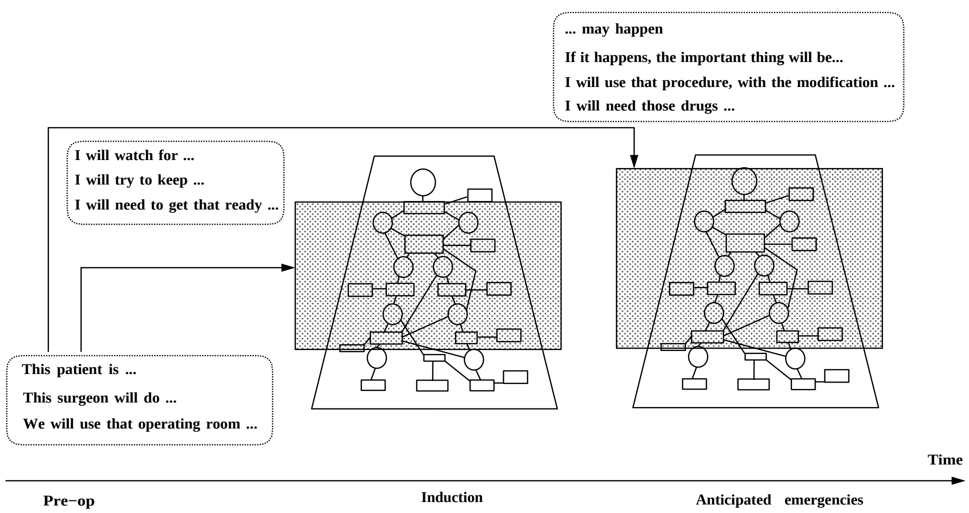
Interacting with Complex Work Environments:
A Field Study and a Planning Model



Yan Xiao

Interacting with Complex Work Environments: A Field Study and a Planning Model

by

Yan Xiao

A thesis submitted in conformity with the requirements
for the Degree of Doctor of Philosophy
Graduate Department of Industrial Engineering
University of Toronto

ABSTRACT

To effectively train and support human activity in interacting with complex dynamic systems, an understanding is required of both the strategies adopted by human operators in actual work settings and the resource demands associated with these strategies. To meet this need, a field study was conducted by taking the task environment of anaesthesiologists as a "laboratory" for investigating how expert human operators achieve successful performance in a complex and event-driven work domain.

A series of field observations over two years revealed that a significant number of the strategies employed by anaesthesiologists were of a *preparatory* and *preventive* nature. Instead of relying on on-site inspirations and complicated mental processing, anaesthesiologists used prior deliberations to prepare both mentally and physically, and to prevent troublesome situations from occurring. A framework was proposed to summarise the results of the field observations. With the guidance of the framework, anaesthesiologists' activities and verbal protocols during eight surgical operations were recorded and were analysed. A number of detailed behavioural patterns associated with anticipatory and preparatory activities were identified. In addition, discussions during four case rounds attended by anaesthesiologists were recorded and analysed. The results of the field study led to the development of a conceptual model of planning and action control in the interaction with a complex work environment. In this model, resources for actions are represented by Rasmussen's "decision ladder," and prior deliberation, or planning, as a means for filling in missing components and changing goal structures in the decision ladder.

The results of the field study and the model developed serve to enlarge the scope of inquiry into the interaction between human and complex systems beyond the event-response unit. They depict the role of anticipation and preparation in the successful interaction with complex work environments. They illustrate an important category of problem solving situations and skills that have not received adequate attention before. They provide answers to some of the key questions for designing effective aids and training programmes. Moreover, they add to our understanding of the adaptive nature of human behaviour manifested as adopting a variety of strategies to overcome obstacles to achieving successful performance.

Acknowledgment

The single authorship of a thesis hides poorly the fact that many other people have made direct or indirect contributions to the thesis. It is a great pleasure for me to thank the many people who have helped me in starting and finishing the thesis.

First, I am indebted to my supervisor and thesis co-advisor, Paul Milgram. He directed me into the fascinating field of cognitive engineering from my previously purely engineering background. I thank him for his trust in my judgment and ability, and for his patience in allowing me with maximum freedom to explore the potential of the field of anaesthesiology as a "laboratory." His support has gone far beyond that of a supervisor, and he has been an English tutor, a personal counselor, and a most supporting friend.

As a co-advisor, John Doyle has been generous in his support beyond what anyone could have expected. Without him, I would not even dream of having an intimate look at anaesthesiologists' work. He without reservation shared his "insider's view" on the job of anaesthesiology, and showed me first hand that anaesthesia is much more than just putting someone to sleep. A big thank-you to John!

I am indebted to Kim Vicente, who has been providing me with much needed encouragement, and is always ready to listen. Conversation with him has been inspiring and stimulating. Through Kim, I had the honor of knowing Jens Rasmussen, to whom I owe a great deal of my understanding of human behaviour. Thanks also go to David Woods, who hosted my visit to his laboratory that encouraged me to use field settings as a window looking into the practitioner's cognition.

I also want to thank my fellow graduate students in the departments of industrial engineering and psychology, in particular, Shumin Zhai, Justin Holland, Randy Sollenberger, Jim Liao, Lia Buarque, Carolyn MacGregor, Sarah Zuberec, Anu Rastogi, David Drascic, Deb Fels, Ferdie Poblete, Martin Krueger, Ken Ruffo, and Peter Wong. They have made the life as a graduate student much more bearable.

Thanks also go to the anaesthesiologists at the Toronto Hospital, General Division. Their support inevitably is vital to the field study.

Special thanks are extended to Arthur Ryman. Through him, I received two terms of IBM Canada Ph.D. Fellowship. Without the support from IBM Canada, finishing this thesis would certainly have been impossible.

I would also like to thank Russ Greiner, Patrick Foley, Mark Chignell, and David Gaba.

Finally, I thank Xiaoping for her understanding and support, and my two wonderful children, Fawna and Eva. Their cheerings lessened the anxiety that I had from time to time, and the maturity they demonstrated in their forgiving attitude towards their Dad has certainly surprised me. Besides, they are such good "subjects" that have been inspiring me constantly. I wish that I had had more time to observe them growing.

Preface

Our world is filled with interesting facts. How should one choose one set of facts over the rest to study? Henri Poincaré (1905/1952) suggested one criterion: choosing facts that are simple. The reason given by him is that the facts we choose should not only be interesting, but have great value, too. The value, as defined by Poincaré, is represented by the chance that a fact repeats itself. Simple facts rely less on the combination of factors, and therefore have a higher chance of recurring. Thus they have more value than do complex facts.

One can see—as Poincaré noted—that biologists and physicists studying cells, molecules, and atoms are lucky: cells and atoms duplicate across species and space. In comparison, those who study the combination of simple facts are not so lucky, or even unfortunate. In particular, "the sociologist is in a more embarrassing position. The elements, which for him are men, are too dissimilar, too variable, too capricious, in a word, too complex themselves" (Poincaré, 1905/1952, p. 17).

Undoubtedly, the variation of human behaviour across both population and time is enormous. But the problem does not seem to be the variation. Gas molecules in a test tube have different movement trajectories, and the location of any one particular molecule can not be known precisely, as prescribed by Heisenberg's Principle of Indeterminacy. However, in thermodynamics, the movement of molecules is represented statistically by aggregate measures such as temperature and pressure, and the variation of movement of individual molecules is removed in the statistical descriptions. The phenomenon—the movement of molecules—could be complex if described at the level of each individual molecule, but simple relationships or facts can be expressed by proper measures, such as in Boyle's Law. So the problem seems to lie in what measures and frames of reference we choose to describe the facts in which we are interested.

This thesis is an empirical inquiry to address this very question of seeking suitable measures and levels of descriptions to capture the observed facts or phenomena in human behaviour, which is surely a domain least promising according to Poincaré's criterion. In fact, the particular category of behaviour to be addressed in the current research relies more on the combination of factors than the usual psychological studies. The subjects are well trained and experienced—anaesthesiologists, who make modern surgery possible, and they are performing in an actual working environment—annulling pain and saving lives, and this working environment is both complex and dynamic—surgical operating rooms. The belief is that this combination of a *proficient practitioner* and a *complex, dynamic environment* has a high chance of repetition in many domains, although it is not usually seen in current laboratory settings. As an analogy to the gas molecule movement, there is an infinite number of possible behavioural trajectories, but the task is to find proper measures and levels of description for the behavioural phenomena, so that the findings obtained repeat themselves not only over time, but over different domains as well.

More than thirty years ago Miller, Galanter, and Pribram named their seminal book *Plans and the Structure of Behaviors* (1960; see also the new printing with a foreword by Donald E. Broadbent: Miller *et al.*, 1960/1986), indicating that human behaviour was more than associations of simple, observable events and activities. They suggested a richer view, asking how physical and mental activities are *organised* in achieving a hierarchy of goals. However, traditional psychological studies have not gone far from the issues presented in that book. As Broadbent pointed out in the foreword of the new printing of Miller *et al.*'s book (Miller *et al.*, 1960/1986, pp. xi–xxix), we are still trying to isolate behaviour into cells, and to use a divide-and-conquer style of research. As a result, the rich structure—the very core of behaviour—is lost in using such a strategy. Such a divide-and-conquer strategy, as concluded in the study of general system theory, can not work well for complex systems such as human behaviour (von Bertalanffy, 1968). The complexity and variations in human behaviour have already suggested that the focus of research should be on the structure and the organisation in behaviour, rather than the characteristics of each functional subsystem or exact sequences of particular mental and physical activities (Hansen *et al.*, 1990).

This thesis is an attempt to understand the structure and organisation in behavioural phenomena observed in an actual working environment. It is empirical in nature, while at the same time it deals with theoretical, sometimes philosophical, issues in trying to explain the observed phenomena in order to present them in a coherent and illuminating fashion.

Contents

List of Figures

List of Tables

Part I

An Old Challenge in A New Context

Introduction 1

An ant, viewed as a behaving system, is quite simple. The apparent complexity of its behavior over time is largely a reflection of the complexity of the environment in which it finds itself.

—*H.A. Simon, 1969, p. 64*

In most industrial, commercial, medical and military settings where daily tasks are investigated there is no hope of the classical "fully factorial balanced repeated measures" design, with its attendent well defined ANOVA. Rather we need a wide variety of "quasi-experimental" investigations, a mixture of experiments, naturalistic observation in field settings, and retrospective analysis of records of individual case histories and actuarial data.

—*N. Moray, 1992, p. 642*

1.1 An old challenge in a new context

Understanding how we think is probably the oldest intellectual challenge that human beings have faced. Until the middle of this century, this challenge had remained a somewhat purely philosophical entertainment. About two hundred years after the start of the Industrial Revolution in the 1760's, the emergence of process plants and large scale automation had created human operators whose physical power and skills were no longer the major determinants of their performance. The success of industrial systems depends more and more on how human operators think, that is, how they solve problems, make decisions, and so on. Consequently, the focus of research on human performance has shifted towards the *cognitive* functions played by the human operator (Bainbridge, 1974; Sheridan & Johannsen, 1976). Designers of machines were forced to be concerned with the intangible, *mental* activities of human operators. The once philosophical question of how humans think has thus acquired a practical, engineering component: How should one design interfaces and training programmes for systems in which the operators' cognitive capabilities are vital to successful performance?

In order to design effective interfaces and training programmes in support of the human operator's cognitive functions, we have to answer the following two basic questions about the complex interactions between the human operator and his or her task environment (Hollnagel & Woods, 1983):

- How does a work domain pose cognitive demands?
- How does a human operator meet those demands?

In other words, we need to know when there is need for cognitive support, and how the need should be fulfilled (Roth *et al.*, 1987). Without answers to these questions, designs will likely be driven by *technologies* rather than by *problems* and *principles*, and solutions put forward will be ineffective, or even worse, the source of new problems (*e.g.*, Bainbridge, 1983; Reason, 1987).

In the search for answers to these two questions, a particular category of systems has attracted the special attention of human factors specialists. These systems are characterised by the following features: (1) containing a dynamic component, which is both internally self-driven and externally supervised and controlled; (2) consisting of highly interactive parts and functions; and (3) having the potential of producing disastrous accidents, mostly because of the performance deficiency of the human element. Examples of such systems include power plants, air traffic control, and large communication networks. Because of the critical role of the human element in these systems, numerous studies in various domains have been carried out to understand how the human element functions in such complex and dynamic environments. Those studies range from early research in process control (*e.g.*, Bainbridge, 1974; Beishon, 1974; Crossman & Cooke, 1974), to critical incident analyses in power plants (*e.g.*, Pew *et al.*, 1981; Rasmussen, 1976; Woods, 1982), to recent studies on expertise in air traffic control (*e.g.*, Redding & Cannon, 1992), and in aviation (*e.g.*, Amalberti & Deblon, 1992).

Despite these and other efforts, a wide range of issues are still to be addressed in areas including information display, training and organisation (Moray & Huey, 1988; Vicente, 1991), all of which depend on the understanding of human behaviour in complex work domains. This thesis addresses this need for understanding the role of the human operator in a dynamic and complex work environment, in particular, how a proficient human operator solves problems to meet task demands.

1.2 Why study human operator's behaviour *in situ*

There are a number of obstacles to a satisfactory understanding of the interaction between the human operator and his or her task environment. The biggest of these is probably human adaptiveness.

The understanding of human behaviour would be much easier if humans were to behave in the same manner in all task environments, including laboratory settings. However, humans are adaptive organisms. Their behaviour changes according to the task environment (Cf. the hypothesis about the ant's behaviour on the beach in Simon, 1969, p. 64). Humans also change their behaviour with accumulation of experience and knowledge. In most domains that we are concerned with, tasks are often highly dynamic and complex (see the four dimensional description of complexity in Woods, 1988, also page 43 of this thesis), and the human operator is well experienced (with about eight years of apprenticeship in the case of nuclear power plant operators, for example). The question is: Will these two factors—complex work environment and experience—affect behaviour in a way that ignoring them would be inappropriate? If the answer is yes, then how much do we know about behaviour in these situations? Will controlled laboratory studies alone suffice?

1.2.1 Humans perform better outside laboratories

For many reasons, human behaviour has been predominantly studied in artificially impoverished environments using novice subjects and context-free tasks. A large body of research considers problem context, domain knowledge, and experience as *obstacles* to the discovery of underlying cognitive mechanisms. Experimental tasks are constructed to "strip the phenomenon of all its accessory conditions" (Gillis & Schneider, 1966, cited after Hammond, 1993, p. 207). This type of research often assumes that the human cognitive system contains a set of basic mechanisms and skills independent of the context of a domain, and should be studied as such. As a consequence, limitations of various kinds become the focus of research, and how we as an adaptive organism succeed in the complicated but context-rich environments attracts little attention (Cf. Toda, 1962). The consistent failures of subjects in carrying out experimental tasks have in fact suggested that "man may be an intellectual cripple"[1] (Slovic, 1976, p. 222).

Nevertheless, "man drives a car, plays complicated games, designs computers, and organizes society" (Toda, 1962, p. 165). So how do we relate discovered limitations to successful performance in daily life? Or is it important to create an artificial environment so that the limitations are manifested in the form of failure?[2]

Studies of Micronesian navigation have taught us an important lesson. There is no clear link between the apparent success of Micronesian navigators and limitations in the human cognitive system (*e.g.*, the limitation that a human being can not orient herself or himself by the magnetic field of the earth). In studying how these navigators succeed (Hutchins, 1983), the discoveries made are not about the manifestation of some limitation, but rather, about the unique representation used, and the skills possessed—the ability to utilise cues from the environment.

Thus it is highly probable that there are some critical and important characteristics in human cognition which are either skewed, or not reflected at all in traditional laboratory studies, let alone the fact that human operators often interact with complex, self-driven systems in a sophisticated, highly technological environment. Consequently, questions have been raised about how much we can utilise the results and theories of classical psychological studies (*e.g.*, Kirlik, in press; Woods, 1993). Even the basic assumptions about features of human cognition relevant to real, non-laboratory tasks have been put to examination (Chi *et al.*, 1988; Hammond, 1993; Rasmussen, 1976).

In summary, it seems clear at this point that there still exists a research need to address the question of how we succeed outside laboratories in natural environments, towards which, it could be argued, that psychological studies should have been oriented in the first place. Answers to this question will certainly provide us with a better basis to understand why we sometimes fail, and to design work places where human limitations are not manifested in the mode of failure.

1.2.2 Field studies can provide guidance needed for laboratory studies

Controlled laboratory studies have some strong appeal over other forms of study. They can create a task environment that can be freely changed. They can also limit or control the factors that contribute to the behaviour observed. However, the very advantage of control over the task environment may lead to the creation of an environment that is too far away from the target situations of interest to be useful,[3] and the separation of factors may just overlook the very key elements and relations that are important to real-life applications (Kirlik, 1993). Researchers' own intuitions alone may be incorrect or inadequate as a guide to the creation of the contrived task environment. Given the complexity of the behaviour exhibited by experienced workers in complex work domains, it is difficult to establish the

---

[1] When I wrote to P. Slovic recently to get his comment on his own quote, he replied: "There are two versions of my article *Towards understanding and improving decisions*. The first (1976) used the term intellectual cripple on p. 222. The second (1982) deleted that term because I decided that it was not appropriate. People's judgments are often erroneous. They need aids and tools to help them with decision making. But this does not mean they are 'cripples.'" The second version is in Slovic (1982). Thanks to Ken Hammond who not only provided the detailed citation for the quote, but also prompted me to obtain Slovic's current comment on his previous position.

[2] One can easily find examples in work places where designers, unintentionally, in one way or other, duplicate what many psychologists have done in the laboratory, *i.e.*, making the limitations apparent. Norman (1988) gives many of such examples. Traditional psychological studies have certainly been successful in providing designers with excuses so that *human errors*—that is, the *user*'s errors—are to be blamed for most, if not all, system failures.

[3] Neisser, when commenting on perception research, has said something very similar (1976, p. 33): "The concept of ecological validity has become familiar to psychologists. It reminds them that the artificial situation created for an experiment may differ from the everyday world in crucial ways. When this is so, the results may be irrelevant to the phenomena that one would really like to explain."

scope of their activities, the kinds of strategies used, and the demands imposed. Guidance from field studies is therefore necessary (Moray, 1992).

After surveying research on supervisory control, Sheridan and Hennessy (1984) outlined three general approaches to performance modelling: (1) studying in actual, operational settings, (2) using large-scale simulations, and (3) developing simplified paradigms suitable for laboratory studies. In particular, Sheridan and Hennessy stress the importance of using operational settings in the study of the human operator's behaviour, because "operational systems are the best vehicle for analyzing behavior to identify such things as ...the information-processing strategies actually used as a function of experience and training, criteria controlling the choice of strategies ..." (p. 29).

Field studies can generate phenomena to reveal the range of cognitive patterns occurring in naturalistic settings (*e.g.*, Scribner, 1984; Redding & Cannon, 1992; Xiao *et al.*, 1992). They can also help to formulate hypotheses and frameworks to guide laboratory research, and to organise results from traditional laboratory studies (Scribner & Cole, 1973; Rasmussen & Jensen, 1973; Baron *et al.*, 1990). Results can also be obtained on human information processing directly from actual settings without having to study each piece of behaviour independently at micro-levels (Rasmussen & Jensen, 1973; Rogoff & Lave, 1984; Klein *et al.*, 1986; Woods, 1993).

1.2.3 New opportunities

The emergence of complex work domains has thus posed challenges as well as opportunities to students of human behaviour. It enables us to observe how human operators overcome limitations in their cognitive systems and how they adapt to highly demanding tasks found in many work places. Several lines of research on human cognitive performance have already converged to the naturalistic approach—studying practitioners' cognitive behaviour in the work environment, such as in everyday cognition (*e.g.*, Rogoff & Lave, 1984; Sinnott, 1989), practical skills (*e.g.*, Lave, 1977; Hutchins, 1980; Scribner, 1984; Singleton, 1978a), and rapid decison making (*e.g.*, Klein, 1989). Cognitive processes are being studied *in situ*—in the setting of the task that calls for them.

This thesis adopts the naturalistic approach. It assumes that very valuable results about human cognitive activities can be obtained outside laboratories and that the question of what and how problems are solved can be answered directly using field data. An actual work setting is used as a "laboratory" to observe experienced workers while they perform their jobs.

1.3 Methodological issues in the naturalistic approach

Field studies differ from laboratory studies in a number of important and fundamental aspects. It is these differences that not only warrant both types of studies, but also call for different ways of conducting and evaluating each. Three areas of methodological issues can be identified in a field study: behavioural data collection, data analysis, and generalisation of results.

1.3.1 Why there are methodological issues

Studying behaviour in uncontrolled environments and during real-life tasks often confronts criticism on methodological issues (*e.g.*, Doherty, 1993). The fact that one has to justify a study of human behaviour in a naturalistic setting, where humans ordinarily behave, as opposed to in a laboratory setting, is certainly extraordinary (see a very critical review of psychological studies in the past forty years in de Groot, 1992). It is even more so that one has to argue that different methods have to be used to deal with different types of phenomena and data in these two kinds of settings. If the quantitative methods developed for controlled laboratory studies are considered as dogmas, then phenomena to be observed and data to be collected will essentially be selected not by their value or significance, but by how well they fit, or are made to fit, those dogmas. Those that do not fit are excluded. de Groot (1992) has listed (among other things) some of them:

- the subject's own view of his or her own mental processes,
- qualitative observations that are resistant to quantitative measurement, and
- phenomenological and narrative descriptions.

Field studies must out of necessity utilise a wide range of data, from self reports to direct, naturalistic observations. Instead of abandoning these "non-conventional" types of data, proper methods of analysing them need to be found. Qualitative and observational studies not only provide us with the understanding at the phenomenological level, they are also necessary steps in parameterisation and quantitative analyses.

1.3.2 Some critical issues

Firstly, there are no well-established methods for collecting behavioural data under field conditions and analysing them. One may run into a chicken-and-egg problem: we have to know what to look for before we can collect useful behavioural data and know how to analyse them, while at the same time without having collected enough data and gained experience in analysing them, we simply do not know what to look for.

The generalisation of obtained results is another critical issue in field studies. As results from studies in naturalistic settings tend to be domain, or even case, specific and loaded with domain language descriptions (see, for example, Scribner, 1984), one has to present the findings in such a language that:

- results from one case can be compared with those from other cases, either within the same domain or across different domains; and
- results obtained from one domain can be applied to another domain.

Two basic strategies have been used to achieve these goals. One approach is to focus on an exemplary domain and to extract the essential properties of human cognitive processes in that domain, and then to verify the findings in related domains (Rasmussen, 1988; Klein, 1989). The other is to study a number of domains and attempt to find similarities (Woods *et al.*, 1990). Both strategies are based on the belief that exploring task demand properties across different work domains with similar characteristics not only gives us a broad exposure to various cognitive situations, but also a better understanding of the effects on humans of the intrinsic properties of complex and dynamic work domains. Examples of domains that have been studied include electronic diagnosis (Rasmussen & Jensen, 1973), firefighting (Klein *et al.*, 1986), fighter flight mission (Amalberti & Deblon, 1992), fast food preparation (Kirlik, in press), and air-traffic control (Redding & Cannon, 1992).

The third issue, described by Woods (1993) as the gap between data and theory (p. 240), is also an issue of how to present findings. Few studies have presented their data in such a manner that allows different interpretations and therefore different theories to account for the data (*e.g.*, Rogoff & Lave, 1984, but see Roth *et al.*, 1987). One cannot generally retrace a study and examine the data collected, and reinterpret the conclusions as one would in general with empirical studies in physical sciences and experimental psychology.

This dissertation promotes the belief that, although there are significant differences across domains, there are potentially common intrinsic cognitive properties present in different complex and dynamic work domains. The task of a field study is not only to identify specific strategies, but also to reveal general characteristics that can be applied widely.

Like pre-Ptolemian astronomers, not only do we lack theories and frameworks, but we are also in need of a basic phenomenological understanding of the kinds of human adaptation that occur in complex work environments.

1.4 An overview

This thesis takes the task environment of anaesthesiologists as a "laboratory." A field study was initiated to attempt to understand what kinds of problems an anaesthesiologist has to solve and how these problems are solved. In other words, it tried to understand how expert human operators achieve successful performance in a complex and event-driven task domain: managing surgical patients and providing anaesthetic service.

The field study first explored the domain of anaesthesiology and various ways of collecting data and analysing them. Direct observations led to a phenomenological account of the task in anaesthesia and a list of general patterns observed in the behaviour. Subsequent study on on-line protocols focused on the *anticipatory* nature of behaviour and substantiated the general findings of the first phase. The thesis then tries to develop a theoretical model to explain the observed data, and to guide both future field studies as well as laboratory studies. It intends to contribute on three fronts: (1) a phenomenological account of anaesthesiologists' behaviour, (2) methodologies of field studies in complex work domains, and (3) a generic conceptual model for capturing the cognitive processes in planning and organising activities in managing complex systems.

The initial phase employed predominantly direct observations as a way to collect behavioural pattern. The confidence and experience built up in this phase led to the second phase, in which more intrusive methods were used, including on-line questionnaires and "thinking-aloud" verbalisations. The initial phase also directed the field study to the analysis of case rounds, where anaesthesiologists (staff and residents) exchange views on previous cases. Although initially the research was largely empirical in nature, considerable effort was spent on building a rather generic model of planning and action control, to resolve the questions encountered in explaining the observed phenomena, and presenting them in a coherent framework.

Compared with other methods of studying behaviour, field studies are probably the most time-consuming. A considerable effort was spent on getting acquainted with a domain, and on making the subject (the person to be observed) feel comfortable with an outsider watching his or her every move. The current field study spanned a period of four years (about the *same* length of time as a residency in anaesthesiology).

See Figure 1.1 for a "road-map" of the dissertation. The main body of it is divided into three parts. Chapter 2 of the current part discusses in detail methodological issues in a field study and reviews previous studies in complex work domains. Part II (Chapters 3 to 8) focuses on empirical issues and presents the field studies. It concludes by raising some critical questions about how to model anticipatory and preparatory behaviour. Chapter 3 introduces the domain of anaesthesiology and includes a work analysis. Chapter 4 describes the direct observation study, which generated a phenomenological account of the task of anaesthesiology and enabled the subsequent protocol study to focus on preparatory behaviour in the task of anaesthesia. Chapter 5 details the protocol study, with an analysis of protocols at different levels of abstraction. Chapter 6 describes the study of peer review protocols. Chapter 7 summarises the field study from the three phases, disucsses several epistemological issues, and highlights the dominant role of anticipatory behaviour in managing complex systems. Chapter 8 examines issues in understanding the observed patterns of behaviour. In particular, questions are raised about how to account for anticipatory and preparatory behaviour and, more fundamentally, the active role that humans play in organising mental and physical activities and in action control.

Part III (Chapters 9 and 10) focuses on modelling issues and the development of a theoretical framework for planning and action control to reflect the planning and preparatory activities found in the field study. Chapter 9 starts with a review of the issue of unit of analysis, and then poses questions of how to account for the preparatory activities. A planning model is proposed using the concept of multi-level action resources, which are developed from Rasmussen's work, in particular the "decision ladder." Chapter 10 uses the planning model to reanalyse the anticipatory and preparatory behaviour observed in the

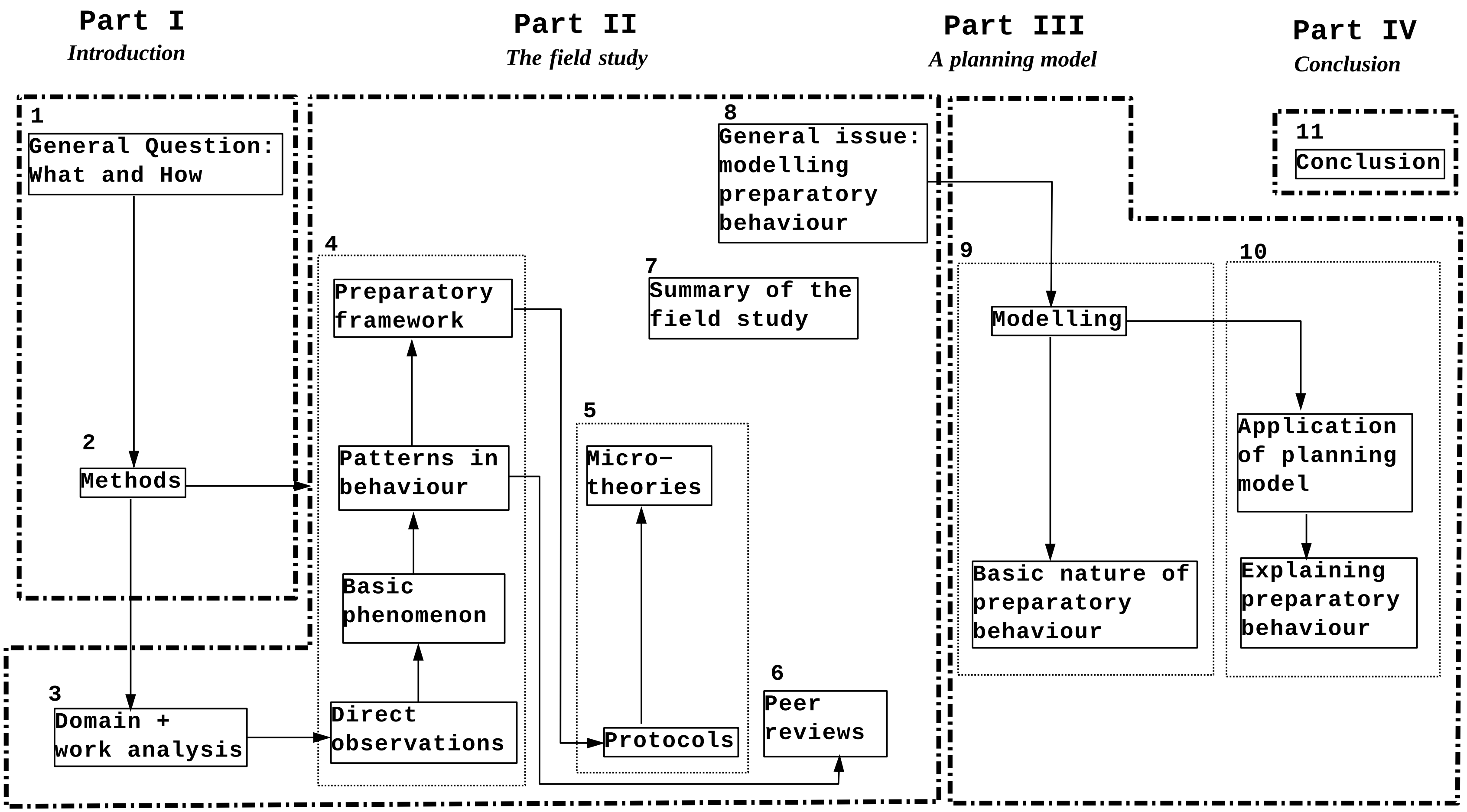


Figure 1.1: Organisation of this dissertation. The Arabic numbers are chapter numbers.

field study.

Chapter 11 concludes the field study and discusses the implications of the field findings and the proposed planning model, and discusses in general the problem solving behaviour of proficient workers in complex domains. Further research questions are posed in light of the findings from the field study.

Field Studies: Theories and Methods

2

> The usual assumption is that the real world is too messy to study directly, so artificial situations are created in which observations are made. On the whole, however, I feel that the analyst's task is no more difficult in the field setting than in the laboratory. The impression that this cannot be so rests primarily on unjustified assumptions regarding the extent to which the behavior subjects in experimental settings is directly revealing of the cognitive processes. Whether the setting is experimental or natural, the investigator must be able to make and support assertions about the representation of the task that the subject brings to the task's solution. The laboratory setting has its advantages, but so has a more naturalistic environment in which assertions about the representation are supported by ethnography rather than by the experimenter's hope that the subject understood what is being asked of him.
>
> —*E. Hutchins,1980, p. 125*

> According to the popular image of science, everything is, in principle, predictable and controllable; and if some event or process is not predictable and controllable in the present state of our knowledge, a little more knowledge and, especially, a little more know-how will enable us to predict and control the wild variables.
>
> This view is wrong, not merely in detail, but in principle.[1]
>
> —*G. Bateson, 1979, pp. 41–42*

This chapter first characterises the nature of field studies. It then reviews previous field studies, in particular those carried out in complex work settings. Methodological issues assocated with field studies are discussed in the context of studying cognitive activities in complex work domains.

2.1 The nature of field studies

By definition, a field study is to be carried out in an environment which is *given* and is not designed by the study; the task of the investigator is not to induce or constrain certain categories of behaviour, but rather, to *understand* how the task environment—as it is known to its inhabitants—fashions the behaviour. In other words, there should be no deliberate control over factors that contribute to behaviour in a field study.[2] Although not implied by the name, a field study also entails that the subject[3] be a *native* inhabitant, who may or may not be selected from among other inhabitants in the studied environment.

[1] Thanks to Kim Vicente for bringing this book of Bateson's and this quote to my attention.

[2] A field study is not to be confused with a field experiment, which is often done in social sciences to manipulate a field setting in order to find out how manipulations change behaviour.

[3] The word "subject" can be misleading. Here it merely indicates the person being studied.

Based on this description of a field study, two general conclusions can be made about the nature of field studies:

(1) that field studies are largely exploratory, as opposed to hypothetico-deductive; and

(2) that field studies consider the actual behavioural *stream* as data, not just some measurement of that stream.

These two points are discussed in detail below.

2.1.1 Field studies are largely exploratory

The search for scientific truth is rarely a random process. Few would wander through the world simply observing, without prior conceptions or search targets. On the other hand, to some degree there are always unanticipated things to be discovered. The characterisation that "we cannot perceive *unless* we anticipate, but we must not see *only* what we anticipate" (Neisser, 1976, p. 43, original emphasis) is certainly not limited to how we see (in a perceptual sense) the world and, indeed, Neisser's (1976) perceptual cycle (Fig. 2 on p. 21 and Fig. 4 on p. 112) can be regarded as portraying the iterative and recursive process of searching for truth in general.

As a particular type of the iterative process, the hypothetico-deductive approach is often perceived as the norm for empirical research, and is followed whenever possible. The idea is that the researcher should first establish a theory, a set of predictions from the theory, and a measurement to be used to collect evidence. Evidence is then selected to verify the predictions. Upon the results of the verification, the theory is either defended or disproved.

However, the hypothetico-deductive approach is not always effective in the pursuit of scientific truth. In a field study, for example, the researcher often starts with some general questions, rather than theories, and these questions are difficult to operationalise in the domain language. The measurement of performance is often the *result* of a field study, and is unclear at the beginning. Compared to an experimental study, a field study is often set to find the unanticipated facts, in particular those that escape brief, casual encounters with a domain. A task appears to be simple but may in fact involve complicated cognitive skills (*e.g.*, the study on the ovenmen in Beishon, 1974; see also the review later in this chapter on page 14), and vice versa (*e.g.*, the study on the Micronesian navigation in Hutchins, 1983). In this sense, a field study is largely an *exploratory* effort, in particular in the starting phase, as opposed to a means of *validating* hypotheses. (Note that this does not mean that *all* field studies are exploratory. Field conditions are sometimes viewed as a necessary setting for confirming results from other fields or laboratory studies. See further treatment of this topic in Scherer & Ekman, 1982 and Sanderson, 1991. Also see the review of the current field study in Chapter 7.)

This fundamental nature of field studies is sometimes overlooked and consequently criteria used for judging experimental studies are often applied to field studies (see the criticism of field studies in, for example, Doherty, 1993). Recognising this aspect of field studies can help one avoid asking improper research questions. The task of processing observed data is often a synthesis procedure—presenting the observed phenomena in a manner that is coherent and relevant to significant research questions. Thus the significance of data analysis does not lie in the statistical power, but in the semantic value of how much of the original problems are solved or illuminated. The generation of phenomenological descriptions in itself is a significant contribution to the answering of research questions. In contrast, an experimental study in the laboratory has to be judged by the power of statistical measures. The significance of a laboratory study is contingent upon the prior theories behind the design of the experiment.

Even though the exploratory nature of field studies may be obvious, it is believed that the discussion presented here is necessary for one to appreciate how various methodological issues are raised and resolved in a field study. The frequently encountered prejudice against field studies is in fact often caused by invoking evaluation criteria which are proper for controlled laboratory studies but improper for field studies (see a discussion by Singleton, 1978b, pp. 14–15).

2.1.2 Behavioural streams are used as data

Another characteristic about field studies is that the investigator is interested not only in the pre-contrived performance measures, if they exist, but also in the stream of behaviour—"a *melody* of movements and postures, adaptively composed to fulfill various requirements" (van Hooff, 1982, p. 363, added emphasis).

The lack of basic understanding about the phenomena under investigation necessitates the search for structures in the observed stream of behaviour, that is, for the recurring themes and subthemes in the "melody" of behaviour. The underlying assumption is that, even though the behavioural stream is contingent on the external environment which is beyond the control of the subject, recurring, relatively invariant patterns exist in the behaviour. Such patterns may not be identifiable if the behaviour stream is examined not directly, but via some measure of the stream, and consequently part of the dynamics of behaviour are lost. Simplistic measures (such as counts, reaction times, and accuracy indices) can, in other words, potentially fail to reveal patterns in the behavioural stream. Studies presented in Newell and Simon (1972) and in Rasmussen and Jensen (1973) are examples of analysing the stream directly, without prior measures.

The basic challenge to the investigator in a field study is therefore whether or not he or she can deduce, from this stream of behaviour, any patterns or organisations that provide answers to the research questions at hand.

Different names have been used to represent the type of study in which the focus is on the behavioural stream instead of on some predefined performance indices. In decision reasearch, for example, it is called the *process method* (Carroll & Johnson, 1990, p.71), to stress the *process* of decision making (as opposed to the *outcome* of a decision). *Protocol analysis* is another popular name for this type of research (*e.g.*, Ericsson & Simon, 1984), as, by definition, the record of the behavioural stream is a set of protocols. *Process tracing* is yet another name (*e.g.*, Woods, 1993). In this thesis, the term "protocol analysis" will be used.

2.2 Field studies: A review

Complex work domains provide a unique opportunity to observe the adaptation of experienced human operators to a multi-task, highly demanding environment. The behaviour observed reflects not only the task requirement as represented by the ultimate goals, but also the surrounding environment, in particular, the characteristics of the tools and interfaces that the designer has provided. The behaviour is also influenced by training programmes, which are often very extensive. Thus, a number of studies in naturalistic settings have been carried out to understand behaviour in actual or simulated complex work domains and to improve design and training. The choice has been simple: simplistic, well-controlled laboratory experiments are judged simply inadequate to reveal the patterns of behaviour in which designers would be interested (Rasmussen & Jensen, 1973; Sheridan & Hennessy, 1984).

The following review of previous studies in naturalistic settings focuses on those that were carried out in complex work domains.

2.2.1 Process control

Process controllers in steel and paper mills were among the first to be studied in detail (Edwards & Lees, 1974), although the uniqueness of process control tasks had attracted psychologists' attention much earlier than that (*e.g.*, Hiscock, 1938, after Lees, 1974). Process control tasks exist in many industrial work domains, and are characterised by supervisory control, whose goal is to ensure autonomous processes are running within prescribed profiles. Often the autonomous processes have long response times with complex, high order dynamics. In the papermill study done by Beishon (1967), for example, it took four minutes for the operators to receive feedback regarding the changes made to the system.

Although some experimental tasks have been constructed in the laboratory to investigate the operator's behaviour in controlling autonomous processes with long feedback delays and high orders of dynamics (such as the "water bath" experiments in Crossman & Cooke, 1962 after Lees, 1974; Moray *et al.*, 1986; Sanderson *et al.*, 1989), a number of studies have also been done in naturalistic settings. Similar to the approach taken by anthropologists studying other cultures, observational methods were usually used to study the human operator in his or her "culture"—the workplace in process control rooms. An intrusive technique has also been used: concurrent verbalisation based on the assumption that one's verbalisation has direct links to one's mental activities and can thereby be used as a basis for inference about cognitive processing (Ericsson & Simon, 1984; Bainbridge, 1979).

The ovenmen

Beishon (1974) studied skills in controlling continuous baking ovens. Aside from long time delays involved in baking, the job of the ovenman was also marked by coordinating multiple tasks (several ovens, each having different kinds of cakes with different requirements) and interruptions from external factors (*e.g.*, deliveries arriving at uncertain times). How to achieve all of these competing tasks was certainly of interest. Beishon showed that the skills involved in operating ovens were in fact complicated. The complication was found largely due to the fact that there were a number of concurrent activities, some of which had a long time span due to the latencies in production. The source of behavioural data used in the study was direct observations with verbal protocol recording. The data were synthesised into an activity chart of the ovenman's behaviour. The focus of the modelling was thus directed at understanding how attention was switched from one activity to another. In analysing the recorded stream of behaviour (*i.e.*, observed physical activities and verbal protocols), difficulties arose in accounting for various "jumps" in the activities: "...the total stream appears disconnected and subject to random jumps from one routine or activity to another" (Beishon, 1974, p. 85).

The exploratory nature of field studies manifested itself in the *discovery* of a key underlying psychological issue: How the ovenman connected activities to fit into a single flow. Beishon summarised his exploratory findings in a proposed model containing a meta-element, called the *executive routine*, to simulate the behavioural pattern that "must follow some kind of cyclic scan procedure ...on a regular basis" (p. 87) to switch attention among various activities.

The melting-shop operators

Bainbridge's (1974) study on melting-shop operators was not a field study in the sense that a *simulated* melting-shop task was used. However, actual operators (as well as university students) were used to do the simulated task. This work appears to be the first to employ a "thinking-aloud" technique to analysis verbal protocols extensively in studying process control behaviour. Verbalisation was used as the major tool to "tap" into the cognitive processes involved mostly in the assessment of the system status, in predictions, and in the choice of actions. Sequences of activities were reconstructed by coding the verbal protocols (including mostly thinking-aloud protocols) and by inferring from the protocols (*i.e.*, "operations which the operator must have carried out," p. 148). Dominant patterns (or "blocks of protocol phrases," p. 148) in the sequence were summarised in flow charts to denote the cognitive control of actions and points of decision and assessment of system status (see a simplified version in Figure 2.1). This model stresses the mental activities involved in the prediction of future inputs to the system and the actions needed to respond to those predicted future system inputs. However, just as Beishon had discovered, the actual sequences of behaviour were flexible.

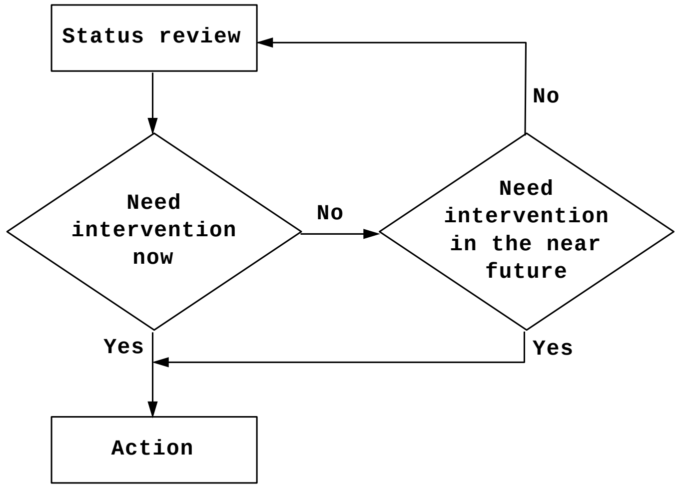


Figure 2.1: Predictive control of a process. Adapted from flow diagrams of activities in a study done by Bainbridge (1974). Bainbridge's study shows that one of the basic sequences of activity of the human operator is this predictive control.

Bainbridge also utilised the advantage of simulation by using inexperienced subjects (university students), and showed that the predictive control strategy was used less frequently by non-operators. Instead, a feedback strategy (*i.e.*, adjust-and-wait) was often used.

The power station operators

Rasmussen (1976) used a similar data collection technique as was used by Bainbridge (1974), but in actual control rooms. Verbal protocols were collected during *startups* of thermal power stations. He found that verbal protocols did not give much information about how data were *processed* by the operator, but rather about the resulting states of that process. Difficulties were met in structuring the protocols, due to the lack of models able to deal with the kinds of behavioural phenomena observed. As a consequence, the role of collected verbal protocols was considered as "a guide or intermittent indication of [the operator's] activity at several levels" (p. 374). Rasmussen established some general phenomenological descriptions about the activities observed. Instead of constructing sequences of activities, he proposed a "*map*" of probable activity sequences which captured the characteristics of flexibility in activity sequences, similar to those observed by both Beishon and Bainbridge in the studies reviewed above.

As a way to characterise the behavioural patterns observed, Rasmussen represented the field study by first formulating a hypothesised normative sequence of information processing activities (*i.e.*, detect, observe, identify, evaluate, choose, and execute) and the resulting knowledge states of these actions. The actual activity sequences were expressed by short-cuts among the nodes in the normative sequence. The result of this conceptualisation is shown in Figure 2.2, which later became known as the "decision-ladder," or the "ladder model" of information processing. According to this framework, the normative sequence of activities and knowledge states involves going up the left leg of the ladder and then down the right leg. Rasmussen's framework was designed to capture the key characteristics of information processing activities of experienced workers: there were shunting leaps among the activities and knowledge states. The observed, actual activities rarely followed the normative sequence (*i.e.*, up and down the two "legs" of the ladder) of activities and knowledge states. Instead, activities or knowledge states were linked together by short-cuts ("rungs" of the ladder) among various nodes in the normative sequence.

The observation that "[t]he operator seems spontaneously to 'know' what is going on and where to focus his attention" (Rasmussen, 1976, p. 373) led to the conclusion that the normative sequence of information processing shown in the decision-ladder was rarely followed. In a very general way, the decision-ladder model captures the notion of directly perceiving typical responses to a situation, and of "knowing where to focus attention": "In any familiar situation, they perceive a small number of alternative plans and they [experts] only need enough information to resolve the choice among those plans" (Rasmussen, 1993, p. 163). It also captures a general property in experts' behaviour that prototypical responses often prevail (Cf. the recognitional model in decisionmaking research reviewed below on page 19).

Thus, the decision ladder is *not* a *model*, in the usual sense, of human information processing (*i.e.*, decision making), but rather a *map* of activities and knowledge states on which actual sequences of activities and activated knowledge states can be drawn. It has the ability to represent human information processing activities under actual task conditions. Not being a model does not reduce the significance of the decision ladder. In complex work settings, the variation in action trajectories is large; the modelling of each trace of cognitive process seems less important than providing a map of possibilities. In other words, an actual action trajectory does not repeat itself very often, but the structure of all possible trajectories is assumed to remain invariant.

Incidents in nuclear power plants

Understanding how industrial accidents occur is of great interest to many people. However, by their nature, accidents happen at unpredictable times. Therefore, studying accidents poses a special problem of how to collect information about what has happened before and during accidents. A number of incident studies (Pew *et al.*, 1981; Woods, 1982; Woods, 1984; Friedman *et al.*, 1985; Sundström, 1991; Payne *et al.*, 1992) used the so-called *process tracing methods* (Woods, 1993). In this method, data from many sources (operational records, retrospective recalls, etc.) are pooled together to recreate activity sequences during accidents. The resulting sequences are analysed in a manner similar to those collected directly in field conditions. Process tracing methods in the study of accidents are used to explore how different factors affect eventual decisions before accidents. Rather than being new approachs to *collecting* behavioural data, process tracing methods use accident records and retrospective reports to *reconstruct* the process of how a decision was actually made, and to explore the role of various factors during the decision making process. The intention is to understand the eventual decision from the perspective of the actual decision maker (Woods, 1993, p. 232). This is in contrast to those incident investigations in which the eventual outcomes are judged without referring to the process of arriving at those outcomes.

Woods (1984) summarised studies of actual and simulated accidents in nuclear power plants, all using this approach of analysing reconstructed activity sequences and then postulating reasons why accidents happened and/or were not corrected. Two kinds of cognitive issues were identified in these studies: fixation on plant state assessment that caused the operator's state identification to be decoupled from actual process state, and "brittle" procedures that could not handle novel situations. The analysis of activity sequences was typically highly semantic: each sequence was examined in its context, and no quantitative analyses were attempted.

Uniformity in the domain in these studies (*i.e.*, accidents in the operation of nuclear power plants) had simplified cross-case synthesisation. Rather than as comprehensive models, the results were presented as a means of identifying and classifying cognitive "bottlenecks" that were considered to have contributed to actual or simulated events. Some critical methodological issues were left unattended, such as the reliability of retrospective verbal protocols. However, the research questions that drove these studies had received adequate answers and the methodological issues mentioned here did not pose particular dangers to the final results.

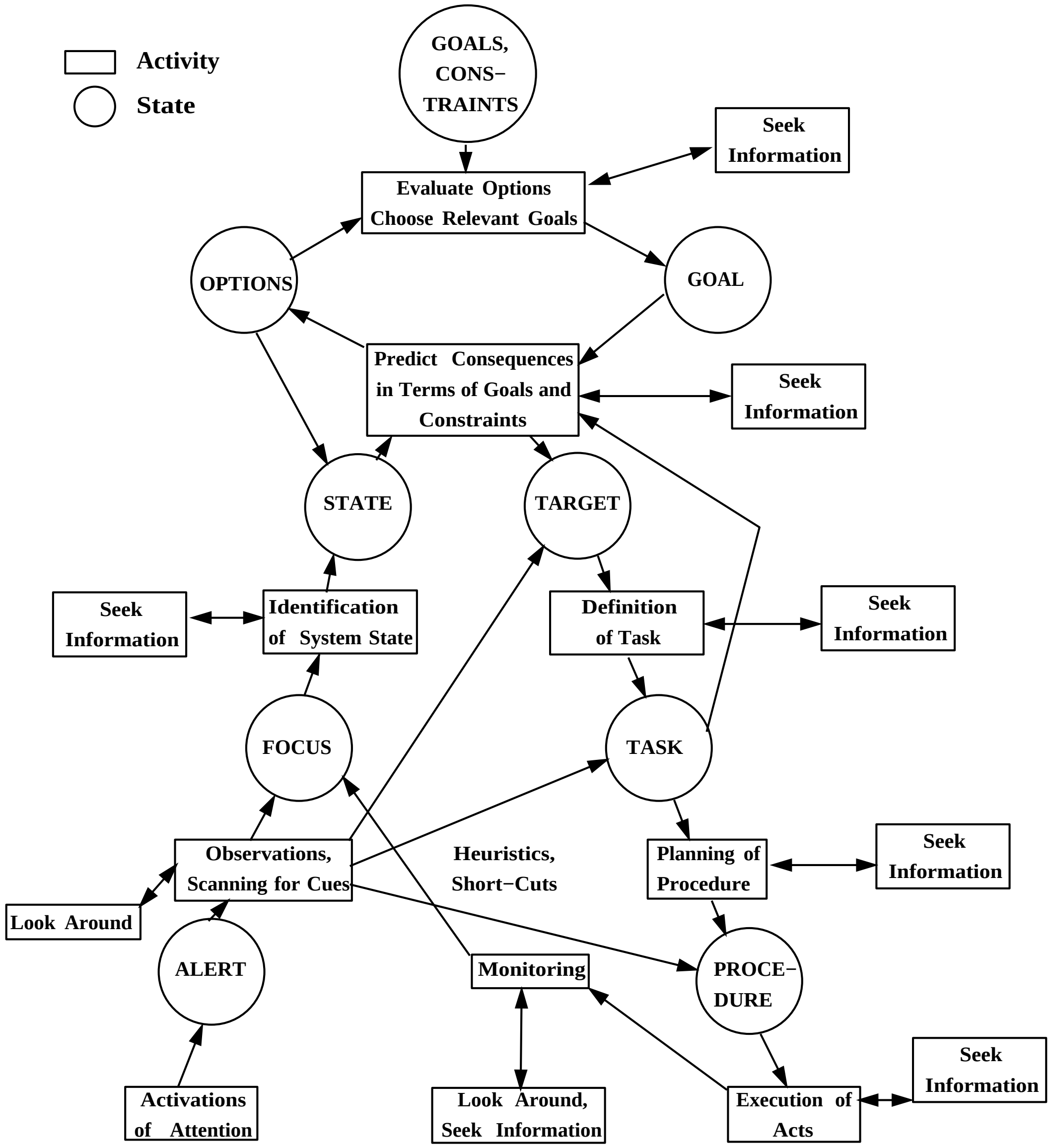


Figure 2.2: Rasmussen's decision ladder, adapted from Rasmussen, 1993, Fig.8.1. The ladder is a network with two types of nodes: information processing activities (in squares) and states of knowledge (in circles). The normative sequence forms the legs of the ladder; all other links (rungs) are possible routes of information processing activities in actual work settings. Note that several variations in the form of the decision ladder have been proposed by Rasmussen.

2.2.2 Other complex systems

Aviation

Pilots in cockpits deal with a rapidly changing environment. Design and training issues, as with other complex work domains, have to be addressed by an understanding of the activities of the operator—the pilot. A few field studies have been conducted for this purpose. For example, Degani and Wiener (1993) investigated how checklists were used in cockpits. Norman (1991) observed how cognitive aids were invented and used in the cockpit by pilots flying aircrafts. In Norman's observation, empty coffee cups were used to cover the flap handle to remind the pilot about flight status.

In a series of studies on pilots, Amalberti and Deblon (1992) videotaped how experienced and novice pilots prepared *and* carried out missions (low altitude, high speed intrusions). Combined with interviews, their observations through video recordings revealed some general strategies used by pilots:

- Utilising local, incidental models or microschemas for short-term predictions, controls, and redundancy checks, to avoid time-consuming processes that would occur if full-scale models are used;
- Anticipating high workload periods and "off-loading" tasks beforehand;
- Simplifying in-flight reasoning processes by conducting elaborate pre-mission preparation and generating trees of possible events; and
- Avoiding situations where there were no ready solutions or no chance of applying known solutions, thus minimising the possibility of having to recompute a new solution. (One way of doing this was to position the system so that prepared solutions could be used.)

The consequence of employing these strategies was significant. For example, seemingly split second life-or-death decisions were in fact dealt with by adopting preprogrammed solutions. The results of this exploratoy field study provided a basis for the subsequent design of decision aids (Amalberti & Deblon, 1992).

Anaesthesiology

The analysis of retrospective reports of incidents requires special precaution, to avoid the kind of hindsight wisdom that has plagued accident investigations (as observed by Rasmussen, 1990). Asking *domain experts* to comment on retrospective reports is one way to reduce the effect of after-the-fact knowledge. The study by Cook *et al.* (1991) used protocols generated at anaesthesia quality assurance conferences to gain an *insider*'s view of the job of anaesthesiologists. Comments from peers on the retrospective reports about the activities during incidents in recent cases were transcribed; the transcripts were used to infer the cognitively loading situations as well as strategies used to cope with those situations. These conferences also provided valuable information that would be impossible to obtain by other means, such as alternative solutions to a problem. Through the comments made during the conference, Cook *et al.* were able to identify a consensus on acceptable performance and the best strategies given a particular case. The knowledge of where the discussion was focused during the conferences gave hints to issues like what information was critical in making certain decisions and judgments.

This approach is unique, in that the evaluation of performance was carried out by practitioners who knew the actual work settings, tasks, goals, and constraints. In particular, Cook *et al.*'s study shows that adequate anticipation of and preparation for future troublesome situations were important indicators of successful performance. This is important, as usually performance is judged by the *eventual outcomes* (demonstrated by a study on closed claims analysis in Caplan *et al.*, 1991). In the conference, it was found that the sensitivity to precursors of complications and stressful factors was often the focus of the discussion in these conferences, instead of the outcomes themselves. A noteworthy point about the analysis of the transcripts in Cook *et al.* (1991) is that the transcripts obtained from the conferences were used essentially for illuminating *cognitive* issues that the authors had formulated. No quantitative analysis on the transcripts was reported.

Judgment and decision making

Driven by similar concerns about the limited utility and relevance of classic psychological studies to the solution of design and training problems, decision making researchers have been concerned over the impact of the prevailing methods of using college students in simplistic laboratory studies. Naturalistic decision making is an emerging field of research, whose goal is to study how real people make real decisions in the real world.[4] The work of Klein and his colleagues (Klein *et al.*, 1986; Klein, 1989; Klein & Calderwood, 1991) best represents this line of research. Retrospective reports were used intensively and often exclusively. Practitioners (fire-fighters, tank platoon commanders, designers, and managers) were asked to report about decision episodes which occurred up to several years ago (Klein, 1989, p. 66). The reports so obtained were coded by independent raters according to whether a decision was made analytically or intuitively, and whether or not concurrent options were evaluated. Results showed the general trend of using intuitive, recognitional strategies when time pressure and experience level were high. A pattern of using prototypical responses emerged by examining detailed evolutions of incidents (*e.g.*, Klein, 1993a, Fig.2 on p. 63). Klein (1989) proposed a model of recognition-primed decision (RPD) to account for this pattern. The RPD model captured several important characteristics of decision making under high time pressure: (1) situation assessment often leads to the recall of a single, prototypical response plan (containing goals, expectancies, actions, and critical cues), as opposed to several concurrent, parallel options; (2) mental simulation is used to verify the prototypical plan and modify it if necessary; (3) only if a plan is rejected by mental simulation is another plan then examined.

The approach that Klein and his colleagues took to study naturalistic decision making represents some key elements in exploratory, empirical inquiries. Starting with a general research question, that is, how decisions are made in naturalistic settings, Klein and his colleagues first collect phenomenological data about various aspects of decision making in one domain (fire ground commanding). With the emergence of several patterns in the process of decision making in naturalistic settings, they proceeded to use these patterns as a way of categorising phenomena collected from several domains (including engineering design and platoon commanding). The findings were presented in a single framework (the RPD model) that could help further empirical studies in collecting data and organising findings. The framework also draws attention to certain aspects of naturalistic decision making (*e.g.*, twenty research questions listed in Klein, 1993b). At the same time, it provides guidance for the design of experimental studies to examine specific aspects in RPD. All these are possible because the findings were presented by removing surface domain features and adding cognitive descriptions (*e.g.*, mental simulations, goals, actions, cues, etc). Studies by Klein and his colleagues demonstrated how a phenomenological study can reveal underlying cognitive mechanisms and how the findings can be presented in a case-*independent* manner.

Electronic diagnosis

Diagnostic tasks constitute an important part of operating most complex systems. Rasmussen and Jensen (1973; also 1974) studied diagnostic behaviour in field settings using

[4] The repeated use of the word "*real*" in the sentence may sound awkward, as so judged by Hammond (1993). He suggested that the word should be "proscribed, eliminated" (p. 217), for the reason that the word "real" implies "emotional tones of a revolt against what many see as a sterile science" (p. 217). Nonetheless it is a fact that the dominant way of research in, for example, decision making is still using sophomores to do simplistic tasks. The word "real" thus not only signifies a revolt, but also the recognition of the naturalistic approach. Naturalistic decision making researchers are certainly not alone nor the first to express the sentiment of revolt. Neisser (1976) has used a similar term—"*ordinary* seeing"—to distinguish his interest in perception from others', presumably in *non*-ordinary seeing.

electronic technicians as subjects. On-line verbal protocols (with thinking aloud) were recorded, and retrospective comments were sought by putting the subjects immediately after each session in the actual working position in front of the instrument. The protocols recorded during diagnosis were read to the subjects for comments and clarification. The protocols were coded in several ways using multiple analysts/coders, and recurring patterns were identified in the streams of protocols. Similar to other field studies, the major task was to synthesise the large amount of behavioural data loaded with domain language and contextual information.[5] Rasmussen and Jensen were able to abstract a few general aspects of searching routines used in diagnosis, and identify two dichotomous search strategies used in diagnosis: topographic versus functional search. This dichotomy spawned other laboratory studies of diagnosic behaviour (*e.g.*, Rouse *et al.*, 1980).

2.2.3 Everyday problem solving

Anthropological studies of skills in other cultures

Finding out what problems people are solving and how they are being solved is a question of interest not only to designers, but to anthropologists and sociologists alike. The following review centres around methodological issues in the study of skilled behaviour.

For centuries, anthropologists have taken up the challenge of understanding how a similar task is performed in other cultures (mostly in the so-called "illiterate" cultures that do not accept Western ways of doing things). The scope of studies covers a wide range of cognitive skills from perception to logical reasoning (*e.g.*, Cole & Scribner, 1974). Studies of skills in different cultures provide some unique opportunities to contrast how a task is accomplished in different cultures, and usually aim at a model of how a task is *actually* performed.

One well-known study was conducted with Micronesian navigators in the South Pacific (Gladwin, 1970). The Micronesians' ability to accomplish the task of navigating on the open seas without maps and compasses certainly posed a mystery to outsiders. Naturally, *participant observation* (Bernard, 1988)—the traditional method of ethnography—was used in the discovery. Micronesian navigators were found to rely on cues that take years of training to interpret. An ingenious scheme is used by them for representation, observation, and calculation. As inferred from these discoveries, the representation used by Micronesian navigators results in a cognitively economical strategy. The computation of current boat position is embedded in the observation and in its interpretation with regard to the destination of the trip (Hutchins, 1983).

These studies represent a fundamental characteristic of ethnographic research which is that, in contrast with the majority of experimental studies, there are no explicit stages of data collection and analysis.[6] It may be misleading to insist on the separation of data collection and analysis, because the goal of ethnographic researchers is often to *understand* how a task is accomplished through examples to demonstrate the general observations.

Studies of skills in everyday tasks

Even within the same culture, problems can be solved in different manners due to experiential differences in solving problems. One such study is on the food-delivery preloaders' task of putting proper numbers of units (*e.g.*, milk cartons) from cases according to orders (Scribner, 1984). The cognitive task being studied involved simple arithmetic calculations of addition or subtraction of small numbers. Direct observations led to two findings: (1) mental work is performed to minimize physical work ("*law of mental efforts*": "*mental work will be expended to save physical work*," p. 21), and (2) the display property of the workplace is utilised to shortcut the arithmetic operations ("*Visual inspection then provides the*

[5] Thanks to Kim Vicente, who told me, based on his personal contact with Jens Rasmussen, that it took Rasmussen and his coworkers *two years* to analyse the protocols collected.

[6] See discussion on the issue of the separation of data from theories in Ericsson and Simon (1984).

quantitative information necessary to fill the order with little or no recoding into the number system," p. 27). This study is significant methodologically in that the direct observation was used for the generation of hypotheses, followed by simulated tasks that were given to people with different backgrounds to support the observational findings. Indeed, experience with the preloading job was found to correlate positively with the deployment of the above-mentioned strategies.

The above study, along with other similar ethnographic studies (*e.g.*, Cole & Scribner, 1974; Lave, 1988), have demonstrated a finding that is echoed in other non-ethnographic studies:

- Cognitive skills are tightly coupled with the subject matters wherein a skill resides;
- The cognitive process is coupled with the content on which the process operates.

The functional value of this pattern of behaviour in terms of cognitive load is illustrated by Hutchins (1983, p. 223):

> The tool box of the Western navigator contains scales and compass roses on charts, dividers, sextants, and chronometres. These are all A/D[7] and D/A[8] converters. In our tradition, the operations of observation, computation, and interpretation are each a different sort of activity and they are executed serially. The Micronesian navigator's tool box is in his mind. There are no A/D or D/A converters because all the computations are analogue. The interpretation of the result (bearing of the reference island, for example) is embedded in the computation (construction of the horizon image) which is itself embedded in the observation (time of day).

Studies on cross-culture comparisons and on everyday problem solving have collectively shown that undue emphasis has been placed on the separation between cognitive process (or structure) and subject matter, and this separation has received unbalanced support from laboratory experiments (Piaget's preformationism is probably the best known in this regard, but see the criticism in, *e.g.*, Carey, 1983). It may be a matter of convenience for a researcher to have such a separation, but unwarranted structures and processes often have to be postulated, as demonstrated in the polar planimeter[9] example proposed by Runeson (1977).

2.2.4 Summary

The studies reviewed above have demonstrated both the uniqueness and the importance of field studies in studying activities in naturalistic settings.

First of all, the behavioural data have been obtained mainly from sources like retrospective reports, concurrent verbalisation, peer reviews, and direct observations. These data are rich in domain language, and their analysis requires a high level of familiarity with the domain being studied. Analysis has rarely been quantitative; patterns are synthesised from streams of behaviour without applying quantitative performance measures. In particular, observational findings often lead directly to conclusions without separate data collection and analysis processes.

Second, the analyses of data have been concentrated at semantic levels. Researchers have used observed data as a basis for building up a description of how a task is accomplished. Patterns of behaviour or idiosyncratic ways of accomplishing a task have been discovered and are represented as *strategies*, the descriptions of which are usually in functional terms—why a strategy was used and in what ways the strategy saved mental and physical efforts and reduced workload. Hutchins' explanation of the Micronesian navigation as quoted above (on page 21) is a typical example. As another example, in explaining why topographic strategies were predominantly used, Rasmussen and Jensen (1973, p. 160) argued that the

[7] A/D: analog to digital.

[8] D/A: digital to analog.

[9] A polar planimeter measures the area of an arbitrary two-dimensional shape directly without doing elementary measurement, such as length.

strategies adopted by technicians were actually rational, in contrast to the earlier criticism that topographic search strategies were inefficient and irrational (Rigney *et al.*, 1968, after Rasmussen, 1993, p. 159)

Such a functional explanation can easily be criticised for its subjectiveness. After all, the reason that a strategy is chosen can be due to entirely something else other than reducing cognitive load. Nonetheless, functional explanations are important in studies on skills, as they offer *plausible* explanations of why a strategy could be chosen in a given situation, which in turn provides guidance and directions for aiding and training for designers and the like. For example, an interface should encourage, not hinder, the use of cognitively less loading strategies.

Third, the strategies observed have often been summarised into a framework using generalised terms. Rasmussen's work on diagnosis and supervisory control and Klein's work on decision making are among the best known in this regard. The generalisation from case-specific data to a domain-independent model of behaviour represents the essence of field studies, the value of which has been proven repeatedly by the impact of such models as Rasmussen's decision-ladder and Klein's RPD model.

2.3 Working with behavioural streams: Issues and difficulties

As discussed and reviewed above, a field study (on behaviour) does not usually start with some pre-defined behavioural measures. Instead, it attempts to *discover* recurring patterns in observed streams of behaviour. The assumption is that there are invariant structures in the behavioural stream that, once identified, can provide answers to the research questions pursued. (In this thesis, the research questions asked are: what kinds of problems are encountered and how they are solved.)

The challenge is therefore whether one can actually detect patterns in behaviour streams that are relevant to the research questions, and how one can do this effectively. This section discusses recognised difficulties and issues in achieving these. During the field study (described in Part II), a few more critical difficulties were encountered, and they will are discussed there (in particular Section 4.2 starting page 53).

2.3.1 Difficulties in processing observed behavioural streams

The basic data in a field study are streams of behaviour unfolding before the observer. The goal is to untangle various threads in this flow of physical and psychological processes, and to understand why the subject behaves in the observed manner. Achieving these goals is easier if the behaviour is highly repetitive and/or changes only in simple ways, and the subject's task is short and is itself simple. Ironically, behaviour in accomplishing repetitive, short, and simple tasks is not very interesting to most researchers.

The difficulties in understanding the observed streams of behaviour can be summarised in four aspects:

- there are large variations in the observed activities;
- the observed activities can be described at a number of levels;
- the multi-tasking and event-driven nature of many work domains makes the behavioural stream difficult to understand; and
- cognitive and physical activities are often intertwined.

There are large variations in behavioural streams

Laboratory experiments are usually conducted in such a manner that the behavioural streams are restrained to a narrow range.[10] In field conditions, on the contrary, the subject's behaviour is not constrained simply to producing repetitive sequences for the convenience of observation. Variation across cases and subjects is an inescapable issue in a field study.

The lack of constraints imposed on the subject leaves a large degree of freedom in producing the sequence of activities. This is reflected in the fact that: (1) the subject usually has a pool of choices of action to accomplish a goal; (2) permutation of action orders is allowed in many cases; and (3) the timing of actions is contingent on the actual situation, which is often unpredictable and nonrepetitive. This freedom not only permits the subject to take advantage of the specific circumstances, but also provides the necessary conditions under which the subject can organise his activities in ways that minimise his mental or physical efforts, while at the same time resolving the remaining degrees of freedom.

Therefore, the investigator will face the challenge of finding recurring patterns in the non-repetitive streams of behaviour, and he or she has to achieve this without artificially constraining the behaviour.

There are multiple levels of descriptions available

A description is a characterisation rather than a recording. Conceptualisation and selection are necessarily involved in the process of describing. Decisions have to be made on what is relevant and what level of detail is necessary. As a consequence, certain aspects of behaviour are deliberately left out, and the continuum of the behavioural stream is segmented and then coded.

For a speech act, the description can be straightforward—the stream of voices is transcribed and then grammatised into sentences.[11] For a non-speech act, however, the segmentation and coding process is not as simple, and the process clearly reflects the observer's *subjective* opinion of what the subject is doing and at what level the stream of activities should be represented. Observers with different levels of familiarity with the subject's activities will provide different descriptions, as illustrated by Scheflen in an example of observing a ball game:

> If you do not know the structural units of the game you will have to record thousands and thousands of fragments .... For example, player number one opens his mouth, scratches arm, lifts bat, looks at player number seven, and so on. But if you do know the units and the system of notation you can codify the game on a single sheet (Scheflen, 1966, p. 270).

One temptation is to use the subject's own verbal description of his or her activities. However, the subject's own description, although informative, is not necessarily a suitable way of segmenting and coding. There are a number of factors that influence the level of details at which the subject verbalises his or her activities. In the studies on action identification (Vallacher & Wegner, 1987), the difficulty of carrying out an act was found to be an important factor, for example. A more difficult act will likely be described at a lower, more detailed level (for example, "*I am trying to reach the alarm-off button*") than

[10] Note that no matter how restrictive a situation is, no two sequences of behaviour are exactly the same. See Bateson's argument (1979, pp. 40–42) about the inherent nature of unpredictability at micro-levels. He maintains that the more precise the control over experimental conditions is, the larger the degree of unpredictability is.

Laboratory studies usually have a pre-set way of measuring behaviour that neglects certain variations. Two most popular measures are reaction time and accuracy indices, both of which do not reveal directly *how* the subject accomplishes the given task. For example, even if two trials produce the same reaction time, the way to finish each trial can be different.

[11] A transcript of a speech act is a partial description of that act. Some researchers have attempted to do this as completely as possible and included pauses, overlaps, durations of syllables, intonations, audible breaths, etc. See Suchman's (1987) treatment of the topic.

an easier one will (for example, "*I am turning off the alarm*"). The latter lacks the physical details, but it describes the action in terms of a higher goal.

In summary, the issue of finding the proper level of description can be resolved only in an iterative manner. Inevitably, one will start with micro levels of description with little or no aggregation. With more experience with the domain under study, higher, more meaningful levels of description can be established.

Multi-tasking and event-driven properties complicate analysis

In naturalistic settings, the subject rarely performs a single task and does so without interruptions for extended periods of time. The observed behavioural stream usually contains transitions from one activity to another. Such transitions can be obvious when caused by external interruptions, but quite often transitions are subtle or even appear random (see the review of Beishon's study on page 14). One activity may be interrupted by another and resumed with few observable indications. Even though the observed behavioural stream is linear and continuous, it may contain several activities which are switched back and forth. Thus difficulties arise when one tries to analyse the stream and to separate it into different threads.

Thinking and doing are intertwined

Quite often the focus of research is on cognitive issues. In a naturalistic setting, however, the division between cognitive and physical activities is not always clear. The term "cognition" is customarily associated with purely theoretical, mental activities and often refers exclusively to the thinking process involved in logical operations and reasoning. Solving mathematical puzzles or playing chess are two typical examples. It is in such a context that a large body of research on human problem solving and decision making has been done (*e.g.*, Polya, 1957).

It would be a strawman's position to suggest that, in naturalistic settings, problems facing people pose similar cognitive demands as do problems of solving mathematical puzzles in laboratories. Nevertheless, by contrasting these two kinds of settings one can readily see the nature of cognitive demands in actual work settings and the implications on how cognitive processes should be studied.

Typically, tasks given to a person in naturalistic settings are not to make decisions or solve problems, *per se*. Rather, cognitive activities are carried out within a larger context (Scribner, 1984), and are means of achieving some global goals (which in domains concerning this thesis are to maintain *safety* and *productivity*). Thinking and doing are embedded in each other, as opposed to a dichotomy, and are not two separated phases of a task, with the former as a purely cognitive phase and the latter a manual implementation of the solution (thus largely ignored). This coupling necessarily implies that cognitive demands in carrying out real life tasks are shaped jointly by factors associated with both thinking and doing. The process of finding a solution is interwoven with the implementation of the solution, which often means incorporating the solution into the stream of physical activities.

Cognitive processes are thus involved not only in the choice of actions, but also in the execution of the choice and the coordination with concurrent activities. Many tasks consisting of well-established routines, and cognitive processes are frequently in the form of action guidance and coordination in a multi-tasking situation. The latter task can be a major challenge in a complex task domain with several concurrent activities (*e.g.*, Kirlik, 1993). Consequently, any model of cognitive processes in naturalistic settings should include the role of mental effort in putting a solution into action as well as in action guidance and coordination (Rasmussen, 1993).

2.3.2 Issues in studying behavioural streams

Four central issues in processing data from field conditions are examined here: (1) studying cognitive activities using field data; (2) syntactic versus semantic analysis; (3) level of analysis; and (4) synthesising case-dependent data into general frameworks of behaviour.

Studying cognition in field conditions

As compared to a motion and time study that addresses the physical aspects of job activities, the current field study focuses on those aspects that largely involve mental activities, in particular problem solving activities. So the first question confronting this field study is: How can one *infer*, from observational and verbal data, the cognitive processes that have taken place in the course of accomplishing a task?

If the subject provides continuous verbalisation of his or her mental activities, there will be abundant *hints* as to what the subject is doing, for what purpose, range of choices, focuses of attention, etc. However, as discovered by previous researchers (*e.g.*, Bainbridge, 1979; Rasmussen, 1976), verbalisation of this kind is rarely effective under working conditions for a number of reasons,[12] which will be discussed later (page 30).

The advantage of field studies in this regard is that there are other sources of information that one can utilise for the inferring of mental activities. It is rare in fact in a work setting that the subject sits still in front of a display with very little observable physical activities, such as physical movement (manual operations, head movement, walk-around, etc.). In a power plant control room, for example, the operator may go to a specific panel to read gauges. Provided that the observer is acquainted with the task domain, a number of valid inferences can be drawn based on direct observations alone. Even though observable activities can not always be relied upon to reveal intentions behind actions, an informed observer can, most of the time, infer the intention behind the observable actions.

The role of syntactical analysis

A behavioural stream can be considered as a sequence of activities indexed along the temporal axis. One natural way of processing this sequence is to search for the inter-relationships among the activities. Such analysis is also called *sequential analysis* (van Hooff, 1982). The first step in a sequential analysis is to code the recorded activities using a fixed vocabulary. The encoded activities thus lose their semantic values after the coding, and statistical tools can be applied to find relationships, such as transition matrices (*i.e.*, frequencies of one activity following another when the two are adjacent temporally) and lagged transition matrices (*i.e.*, frequencies of one activity following another, but the two are separated by other activities). Some tools have been developed to carry out sequential analyses. SHAPA by Sanderson (1990, also 1991) is an example of one such tool.

Sequential analysis is syntactic in the sense that only the coding part of analysis requires the understanding of what is behind each coding label. Due to the semantic-free characteristics, sequential analysis is sometimes highly valued because of the isolation achieved between data processing methods and the theories to be built (Ericsson & Simon, 1984). However, such *syntactic* analysis is possible only after a proper coding scheme is established. For tasks with a closed problem space (*i.e.*, all possible solutions are specified), such as solving mathematical puzzles, an inclusive coding scheme is usually possible. (See cryptarithmetic, logic, and chess examples given by Newell and Simon, 1972.)

As rightly pointed out by van Hooff (1982), important aspects of behaviour may not be covered by sequential analysis, which detects only *serial orderliness*. In other words, analysing order-relationships may not give a complete description of the behaviour. In particular, the *hierarchical* relationships in behaviour are not covered by sequential analysis. Animals, even as simple as the digger wasp, present evidence that behaviour is organised in much richer ways than just a flat, serial sequence. van Hooff's description of the digger wasp, based on Baerends' research (1941, cited after van Hooff), is rather illuminating in this regard, and is quoted here:

[12] Compare with laboratory studies in, for example, Newell and Simon (1972).

> The insect simultaneously takes care of several larvae, each of which has been deposited in a different burrow. The care for each larva is a rather fixed routine that, after the initial stage, operates almost independently of external feedback. The animal attends to its different young in turn. Returning to a particular young after an interval in which she has cared for the others but also has performed other behaviors, she picks up the routine at the point where she left off the last time. So she keeps a record and follows a time schedule, while functioning according to an action plan. Such a process, with dependencies of steps remote in time and interruptions by other behavior, defies explanation by a conventional Markov model[13] (van Hooff, 1982, p. 423).

Obviously there is very little reason to assume that an experienced human operator in a multi-tasking, event-driven environment would present behaviour with less rich structure than a digger wasp does.

As a comparison with sequential analysis, the identification of hierarchical structures in a behaviour stream requires a deep understanding of each segment of the stream in a manner similar to understanding the grammar in a sentence (van Hooff, 1982). For example, the investigator has to rely on a deep understanding to tell where a routine is interrupted and resumed. Each segment (or *episode*) has to be examined in detail; consequently the process of analysing can be very time-consuming (see the footnote on page 20).

Semantic analyses have been carried out under different names. The steps involved usually start with pooling data from all sources, followed by re-constructing events that are assumed to have happened. Episodes are then extracted and analysed individually. Critical incident analyses (Flanagan, 1954) and process tracing methods (*e.g.*, Woods, 1993) are examples of semantic analyses.

As one can see, although there are advantages in syntactic analyses, semantic analyses are essential for discovering rich structures of behaviour, in particular those in naturalistic settings.

Level of analysis

Carroll and Johnson (1990, pp. 77–83) outlined three types of protocol analysis: exploratory, content, and explict modelling. While exploratory analysis deals exclusively in the semantic domain, content analysis involves categorisation and coding. They seem to suggest that explicit modelling is the ultimate goal of protocol analysis: building a computer simulation to account for the protocol data. This series of steps of analyses essentially document what Newell and Simon (1972) have done in their classic studies on human problem solving. "The advantage of the computer simulation is that it forces a level of specificity seldom reached in other ways" (Carroll & Johnson, 1990, p. 81). Along with the gain in specificity, however, comes the price, in the form of limitations on what types of data that one can deal with: studied tasks usually are "(1) fairly short; (2) repetitive; (3) not so highly structured that people simply read materials in order; (4) not so unstructured that responses are idiosyncratic and hard to code; and (5) verbally processed, that is, not based on nonverbal or overlearned skills that are inaccessible to consciousness ..." (Carroll & Johnson, 1990, p. 83).

Clearly, when dealing with protocol data collected under field conditions, one has to be prepared to analyse them at the level compatible with the scope of inquiry and with the type of data at hand. Hollnagel *et al.* (1981) address the issue of level of analysis exactly along these two lines. Instead of taking the process of protocol analysis as a series of operations that specify the mental activities in ever finer detail, Hollnagel *et al.* proposed a scheme

[13] A Markov chain is a sequence of events with the property that the transition probability depends only on one preceding event. This definition is extended by van Hooff to include those events that depend on more than one preceding events. See p. 397 in van Hooff (1982). Cox and Miller (1965) define such chains as m-dependent chains, with Markov chains as a special case. Relationships expressed in the above-mentioned transition matrix or lagged transition matrix are examples of a Markov chain, or more properly, an m-dependent chain. Thanks to J. Templeton for pointing out the reference of m-dependent chain.

of analysis that emphasises almost the opposite: *abstraction* of the raw data to remove context specifics and *identification* of strategies, performance criteria, etc. for pragmatic purposes, that is, training and aiding. Thus the coding of protocol data is only a step in the abstraction process, rather than a means of acquiring specifications. The ultimate goal is no longer a computational model, but the representation of the raw data at different levels of abstraction, each of which is a result of analysis at that level.

The above contrast has made it clear that, given the particular goals and the particular domain selected, the current field study will follow Hollnagel *et al.*'s approach to analysing data.

Synthesising findings

A field study is different from a case study, where the focus is on an individual case. Instead, results must be presented so that the findings can be applied across cases, or even across a class of domains. A wide range of applicability of findings signifies not only usefulness, but also support for the findings.

As discussed in the review of previous field studies in this chapter, the success of a field study critically depends on how the findings are synthesised and presented. Some studies stop at the level of domain language; the results are simply a description of how a group of people accomplish a given task in a given situation (*e.g.*, Scribner, 1984). However, for many practial purposes, results synthesised at this level are not adequate. In particular, for the purpose of improving design and training, the findings have to be presented so that they are not case-dependent or, on occasion, even domain-dependent.

* * *

To summarise, behavioural streams observed in field conditions often require extensive qualitative analysis that relies on the examination of semantic or contextual information. The *semantic* analysis is especially important when the focus is on the structures in the behaviour, which often defy quantitative or syntactic analysis. In carrying out semantic analysis, one has to pay attention to those aspects in behaviour that are *case-independent*, and to presenting findings in ways that enable the results to be widely applicable.

2.4 Techniques of collecting data in field conditions

It is fair to say that data collection in field conditions is the result of a compromise between what one would like to know about the subject's behaviour and what one can obtain given various constraints in practice. This section examines various techniques of data collection in field conditions.

One unique aspect of field studies worth noting is the relationship between raw data—those that one can obtain directly without analysis of any kinds—and conclusions. Not every field study has a clearly separate phase of data collection. (Many ethnographic studies do not. See discussion in Section 2.2.3 on page 20.) In other words, the collection of data and their analysis are sometimes so tightly coupled that only the result of the preliminary analysis of the data is reported (see examples in Lave, 1988). Such an approach is certainly useful at the starting phase of a field study, but it prevents other researchers from re-examining the data used in deriving the findings. The objectives for having a distinctive data collection phase may therefore be identified as (1) providing other researchers the access to not only the final findings, but to behavioural data in some forms and in some levels of abstraction, and (2) allowing close examination and different methods of analysis of the "raw" data obtained.

Data sources in field conditions can be classified in several ways, such as spontaneous versus induced, verbal versus observational, and concurrent versus retrospective. The following discussion on the techniques of collecting behavioural data deals only with the division of data sources into two categories: observational and verbal, as these are the primary sources of the current study.

2.4.1 Direct observations

To simplify wording, in the following by "observation," the intention is *direct* observation: the observer "watches" the subject's on-going activities.

Being in a field and simply watching what the subject is doing is an important step in a field study. Direct observations not only provide one with a record of major events, but also enable one to establish a sketchy phenomenology of general patterns in the behavioural streams. These patterns can give guidance to other formats of inquiry, such as contructing questionnaires and interpreting other, presumably more objective, forms of behavioural data, such as video and audio recordings. Perhaps more importantly, direct observations can orient the researcher to critical aspects of the subject's behaviour.

Observation is an active understanding process

As compared to verbal data, which one can record directly in linguistic form, observational data are themselves a *description* of what is happening. Few are passive data recorders when it comes to observing activities of others (Newtson, 1976). Observation is in fact a process of two steps: comprehending what one sees and recording the results after comprehension. Often, the observer himself or herself is the data collector, the analyser, and the theory builder. Data collection and analysis are often intimately mixed, and both are heavily dependent on the researcher's own reflections on what has been observed. He or she decides what environmental events and activities are revelant to the subject's task, and whether or not the behaviour observed is as natural as it should be. Thus the quality of data is directly determined by the ability of the observer to understand the domain, as well as by the ability to detect cognitive situations.

Video recordings provide one with the opportunity for repeated examination of segments of a behavioural stream, allowing the observer to go back and to check his or her descriptions. It also allows observations in situations where it is impractical to place an observer (as in the study on military flight missions in Amalberti & Deblon, 1992). Video recording itself, as the name suggests, is not a description of what one is doing and for what purpose. Thus the requirement on the ability to observe directly what the subject is doing cannot be alleviated or removed simply because videotaping or other mechanical recording methods provide a means of recording.

Conducting direct observations

Direct observation is convenient and often very informative. Essentially all the domains afford this source of information. However, gaining access and training observers can be difficult, especially in technically complicated domains. Using domain experts as observers can eliminate the need to train the observers in domain knowledge; however, training has to be done to orient the domain "insiders" to let them know what situations are to be observed and the purpose of the study (Flanagan, 1954). In other words, for an observer to be effective, he or she has to be *both* domain-literate and sensitive to cognitive issues.

In ethnographic studies, a special brand of direct observational methods was developed: *participant observation* (Bernard, 1988)—that is, observations are done while the observer is participating in a setting. Quite often this is the only way that one can gain access to a domain and learn the language and the culture in that domain. The advantage of participant observation is essentially to have an observer who is at the same time an "insider" in the observed domain and is playing a role in a work setting. However, it is not always possible for a domain expert to become a researcher of behaviour, nor for a researcher to become a domain participant, especially in many of today's complex work settings, such as cockpits and operating rooms.

2.4.2 Verbal reports

Verbal reports constitute an important part of the behavioural data that one can readily collect. Their role in revealing mental activities is unique, but involves some controversies.

Using verbal reports to study mental activities

The very nature of mental activities prevents an external observer from gaining direct access to them. Mediating or indirect methods have to be found instead and, inevitably, inference has to be used to interpret the behavioural data obtained by indirect means.

In the long history of studying mental activities, the most often used method is verbalisation in various forms. Introspection, used widely before the era of behaviourism, is probably the most "direct" as compared with other mediating methods. Although methodologically speaking there are potentially serious problems with introspection, the nature of questions addressed by introspection (*e.g.*, mental states after reading a paragraph of text) is of a different kind from the current research and not relevant.

Concurrent verbalisation, first used probably by de Groot (1965) in studying chess players and made well known by Newell and Simon (1972), has been favoured by many despite the criticism of using verbal protocols as reliable data about mental activities (*e.g.*, Nisbett & Wilson, 1977). Ericsson and Simon (1984) examined conditions under which concurrent verbalisation is trustworthy, and thus can be used as data for testing behaviour theories just as other directly measurable data. Their basic argument is that, if the verbal report comes from the same source as the targeted mental activities, then the verbal report can be used as reliable data for studying the targeted mental activities. Bainbridge (1979) suggested that, because one can not test the correlation between verbal reports in general (concurrent or retrospective) and mental behaviour, verbal report alone can not be used to test theories of mental behaviours.

Retrospective reports are in general regarded as less reliable than concurrent verbalisation. According to the data model proposed by Ericsson and Simon (1984), if verbal reports derive from the memory resulted from the cognitive process concerned, then the report should be treated as reliable. They also provided specific warnings against certain methods of collecting verbal reports. For example, asking "why" questions, or asking for summarised or generalised processes used, or probing for hypothetical states, "can never tap subjects' memories for their cognitive processes" (p. 29).

Various kinds of supporting methods exist for the purpose of helping subjects recall from memory about the cognitive processes which occurred. Asking the subject to comment or recall while reviewing video and audiotapes is an effective way to help the subject to retrieve past cognitive processes. Using the actual task environment as a setting for retrospective report is also a good support for the subject (Rasmussen & Jensen, 1973). These methods, hereafter referred to as *cued recall methods*, are important tools for collecting behavioural data.

However, a critical problem with verbalisation, either concurrent or retrospective, is that the subject has only a limited access to the cognitive process. This reporting problem is even more acute when the subject is experienced at what he or she is doing. Reports on cognitive processes are fragmented at best, and the cognitive control of behaviour degenerates due to the repeated exposure to the task. With concurrent verbalisation a practical problem is that it taxes the subject's cognitive resources, and it is often those times when no or little verbalisation is reported that we are most interested in (Rasmussen, 1976).

Nevertheless, the utility of verbal reports can not be denied simply because of the methodological concerns. Previous studies (*e.g.*, those reviewed above in Section 2.2) show that verbal reports have been essential in discovering mental activities and strategies. Methodological concerns are probably more acute if the investigator has less confidence in his or her understanding of the activities being studied and/or the purpose of a study is to prove a model of behaviour (Bainbridge, 1979). Below various kinds of verbal reports are discussed in turn.

Sources of verbal reports

Retrospective reports. Retrospective reports have been used widely, especially in industrial incident investigations. A number of reasons exist that make this data source indispensible for studying cognitive activities. For example, in incident analysis, the recall of events is often the only way to obtain information about incidents that led to the final outcome. In addition, retrospective reports provide an easy access to a large number of past incidents, as demonstated in the studies on decision making by Klein and his associates (*e.g.*, Klein *et al.*, 1986).

When used jointly with other source of data, retrospective reports provide additional information regarding some high levels of mental activitities, such as intentions and interpretations of events during a past incident (*e.g.*, Pew *et al.*, 1981). These are not always clear from direct observations or concurrent verbalisations.

Thinking aloud verbal protocols. Without ways of directly measuring and recording conscious mental activities, asking subjects to "think aloud" is a seemingly naive, but in many situations effective way to collect traces of cognitive activities. The concurrent nature of thinking-aloud technique has made it the preferable way of obtaining behavioural data (Ericsson & Simon, 1984).

In naturalistic settings, however, the application of the thinking-aloud technique may not be as efficient as in laboratories, for the following reasons:

- thinking aloud will likely be disrupted by external sources;
- the social environment of the subject can have a negative impact on the verbalisation; and
- the duration of a task can be too long for one to verbalise continuously.

These difficulties drastically reduce the value of thinking-aloud protocols, even though the value of the method has been proven in many laboratory studies.

Concurrent probing. The recall of past events can also be facilitated by carrying out interviews on an on-going basis. Probing questions can be asked whenever the subject's task performance is not deemed to be affected. The subject then reports on the mental activities after a minimal delay. The actual setting (equipment layout, system status, the social environment) should help the recall of cognitive activities. This mode of collecting data has not been reported before, probably due to concerns over possible disturbances introduced by on-going probing questions.

Of course, as compared with retrospective reports, concurrent probing has the disadvantage that neither the observer nor the subject can predict future events accurately. Questions that in retrospect should have been asked may unfortunately not be obvious to the observer during the on-going case.

Team communications. In many work domains there is more than one person involved in the operation, and communication among the team members is an important source of information on cognitive activities. Some researchers (*e.g.*, Suchman, 1987) have used two subjects to carry out a task jointly, for which normal operation requires only one person. This technique essentially creates a team and therefore induces verbal communications that otherwise would not exist.

Team communications consist of rich information about the subject's intention, situation assessment, information seeking activities, and so on. For this reason, this source of behavioural data is very important, and should be collected and analysed when available.

Cued recalls. Difficulties in collecting on-line, concurrent verbal protocols make some researchers rely heavily on retrospective reports. To facilitate recall, providing cues of various forms to the subject is naturally a way to combat some of the problems associated with retrospective methods (Rasmussen & Jensen, 1973; Ericsson & Simon, 1984). The subject is asked to recall the mental activities while on-line recordings (in video, audio, or transcripts) are played back. Comments and clarifications are sought during the replay.

Domain expert commentaries. Domain experts, by using their own reflections, can often provide some unique information on the performance of their peers. Conference transcripts on previous cases, for example, were used by Cook *et al.*(1991) to gain insight into how practitioners view each other's performance. Although the comments collected do not describe what has happened during a task, they provide information on how the past events should be analysed and what factors could have influenced those events.

2.5 The plan of attack

The field study described in this thesis takes a medical domain—anaesthesiology—as a "laboratory" to examine two general questions about problem solving in a naturalistic, complex work environment: What kinds of problems are being solved by practitioners, and in what ways are these problems solved? The approach to be used is straightforward: observing the anaesthesiologist while he or she is working in the operating room. The choice of the anaesthesiology domain was largely due to the similarity between the anaesthesiologist and an industrial process controller: both deal with a dynamic, event-driven, complex system.

Two phases were initially planned in the field study: the *direct observation study* and the *protocol study*. Later it was pointed out that a particular opportunity existed in the hospital that could provide some unique viewpoints into the problem solving activities. This is the morning case rounds, in which the staff and the residents discuss previous cases mainly for training purposes. A third phase, the *peer review protocol study*, was thus included. The intention is that, while direct observations can give a general understanding of the domain, as well as methods to be used in collecting and analysing protocols, protocol analysis can provide finer grain examination of the patterns discovered. The peer review protocol study can then focus on how anaesthesiologists themselves identify problem solving situations and evaluate performance in such situations. The first phase, the direct observation study, generates some general patterns in the problem solving behaviour, while the later two phases enrich those patterns through protocol analysis.