State Misinterpretation in Flight Crew Behaviour: An Incident Based Analysis

Gordon Baxter

School of Psychology

University of Nottingham

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy,

October, 2000.

Abstract

State misinterpretation has been identified as a causal factor in several accidents where humans were operating complex systems in dynamic domains. The concept of state misinterpretation, although undefined, is characterised by its unobservability, and its relative infrequency. These features make gathering data about state misinterpretation difficult. It was therefore decided to use archive incident report data; the NASA Aviation Safety Reporting System database was used. A definition of state misinterpretation was formulated and translated into database queries to retrieve relevant incident reports for a homogeneous set of expert pilots over a fixed time period. These reports were encoded using the Cognitive Reliability and Error Analysis Method suitably adapted to analysis of aviation incidents. Each report was categorised by the type of state misinterpretation, and a taxonomy of these types was developed. Those types which occurred more than 20 times were analysed at three levels of abstraction. First, a concordance of individual actions showed that communication failures, missed observations and distractions were the most common causal factors. Second, the sequences of possible causal actions showed that some sequences are common across different types of error *and* state misinterpretation. Third, the causal trees for each state misinterpretation type were quantitatively compared. The lack of measured similarity between the trees suggests that the types in the taxonomy are distinct. Most of the analysed incidents were preventable by better management of flight crew actions. Two particular sequences of actions dominated the results. The first is where the flight crew missed an observation when they were distracted by a competing task. The second is where a communication failure between the flight crew and air traffic control occurred. Some suggestions are offered about how flight crews can better manage their actions to prevent the occurrence of some types of state misinterpretation, thereby reducing incident numbers.

Acknowledgements

First and foremost I owe a great deal to Frank Ritter who supervised most of the work presented here. Even when Frank returned to the USA he continued to act in a supervisory capacity, passing back comments, advice and encouragement along the way.

I am also indebted to Fernand Gobet who agreed to take over as supervisor when Frank returned to his homeland. The advice and support he offered was much appreciated, and helped to shape the final form of the thesis.

The initial intention had been to investigate behaviour in the operation of chemical process plants. When it became clear that the relevant data source was not going to be available in time, the target domain was changed to aviation. Fortunately Ellen Bass had provided me with a good grounding in aviation whilst we were working together at Nottingham, and helped to point me in the right direction.

Mike Ayers and Alistair Cameron at Thomson Racal Defence graciously provided part time employment for almost the last three years of the work, which meant that I did not have to worry about paying the bills.

Several people graciously responded to questions and requests for papers and offered advice along the way. I owe a debt of gratitude to them all, especially to Michael McGreevy from NASA who was particularly generous in this area, and George Kuk.

Last, but by no means least: I could never have made it this far without the continued encouragement and support of my friends and particularly the members of my family. Thank you to all of them.

Table of Contents

List of Figures

List of Tables

Glossary of Terms and Abbreviations

ASRS

Aviation Safety Reporting System. The voluntary incident reporting system run by NASA on behalf of the Federal Aviation Authority.

ATC

Air Traffic Control.

ATIS

Automatic Terminal Information Service.

CFIT

Controlled Flight Into Terrain.

COCOM

Contextual Control Model. See Hollnagel (1993).

CPC

Common Performance Condition.

The CREAM

The Cognitive Reliability and Error Analysis Method. See Hollnagel (1998a).

CRM

Crew Resource Management.

CSERIAC

Crew Services Ergonomics Information Analysis Center.

DME

Distance Measuring Equipment.

FAA

Federal Aviation Authority. The regularity body for aviation in the United States of America.

HAZOP

Hazard and Operability Study.

ILS

Instrument Landing System.

IMCs

Instrument Meteorological Conditions.

Incident

Used here to describe any event which involves an unsafe or potentially unsafe occurrence or condition.

MHIDAS

Major Hazards Incidents Database Service. A reporting system maintained by the Safety and Reliability Directorate of the UK Atomic Energy Authority.

NTSB

National Transportation Safety Board.

PORV

Pilot Operated Relief Valve.

PRA

Probabilistic Risk Assessment.

RIDDOR

Reporting of Injuries Diseases and Dangerous Occurrences Regulations. A United Kingdom regulation which requires mandatory reporting of incidents.

RNAV

Radio Navigation System.

SID

Standard Instrument Departure route.

STAR

Standard Terminal Arrival Route: A route from an airway to within the airport vicinity.

TCAS

Traffic Alert and Collision Avoidance System.

VOR

VHF Omni-directional Radio Range.

1 Introduction

Errors are an intrinsic part of everyday life. The inseparable nature of knowledge and error was noted by Mach (1905), who pointed out that "Knowledge and error flow from the same mental sources, only success can tell one from the other."(p.84). It is only comparatively recently, however, that researchers have come to widely accept the inherent relationship between correct and erroneous performance (e.g. Hollnagel, 1993b). This introduction provides an overview of human error research, looking at historical aspects, problems, and the taxonomic approach. The part played by human error in complex system is also considered with particular reference to errors that are attributed to state misinterpretation. The chapter closes with an overview of the structure of the rest of the thesis.

1.1 Human Error Research: A Brief History

The understanding of human error as a behavioural phenomenon is an inherent part of psychology. As such, human error research has been conducted since the latter part of the 19th century. Interest in the field has been affected by changes in the dominant school of thinking in psychology, which meant that human error research effectively took a back seat—with a few notable exceptions—between World Wars I and II (Reason, 1990).

The pace of change in machinery developments which came with the Second World War led to the need to find ways of investigating how to adapt tasks to fit the operator. Although this research originated in the military it fairly rapidly spread into non-military domains. The net result was the new field of ergonomics (or human factors) which combined aspects of engineering with those of psychology.

Recent research into human performance in the operation of complex systems is largely founded upon Neisser's (1976) perceptual cycle. Human performance—which includes erroneous performance—is the result of a continuous loop in which the human's model of the system influences the interaction with the system. [1] This interaction changes the status of the system. The human notes the change and updates his model of the system accordingly. Investigations of performance therefore need to consider the features of, and constraints imposed by these three components: the human, the interactions with the system, and the system itself. Hollnagel

[1] The term *system* is used here in its widest sense, and will generally be taken to refer to a human-machine system *and* the context or environment in which that human-machine system is deployed.

(1998a) generalises from this tripartite model of performance by extending the scope of the components into people, technology, and organisation respectively. The latter is perhaps best conceived of as the environment, since it includes external factors such as working conditions and the ambient surroundings as well as organisational factors, such as operating procedures or regulations.

1.2 Research Problems

One of the fundamental notions underpinning human error research is that the same types of behaviour should have the same causes. In order to categorise behaviour, human performance data is required; this data can also be used to validate models of performance. The gathering of data on erroneous performance can be problematic, however, especially when the errors are infrequent, or the consequences of the error may be disastrous. Both of these situations mitigate against the collection of data from operators working with live systems under normal working conditions.

Another problem is that the field of human error research is fraught with terminological ambiguities. Even the term *human error* itself is ambiguous, because it can describe three different elements of performance (Hollnagel, 1998a):

- The cause of an event or action: "The altitude deviation was caused by human error."
- The event or action: "I forgot to change the altimeter setting."
- The consequence of an event or action: "I made the mistake of starting to taxi before being cleared."

Hollnagel (1993) suggested that the more neutral term *erroneous action* should be used to refer to the observable manifestation of an error, although its use has not been widely adopted. Any action can only be considered to be erroneous after the fact (with the exception of deliberate violations). The categorisation of an action as erroneous is therefore a judgement made in hindsight (Woods, Johannesen, Cook, & Sarter, 1994), based on the following criteria (Hollnagel, 1998a):

- The action has to be compared to some measure of expected performance.
- The action has to have resulted in the degradation in performance.
- The actor who performed the action has to have been able to choose to act in a way that would not be considered erroneous.

Erroneous actions are often the result of either performing the right action in the wrong circumstances—a failure in planning, or a mistake (Reason, 1990)—or performing the wrong action in the right circumstances—a failure in action execution, or a slip (Norman, 1981; Reason, 1990). In either case, the action could have been deemed to be correct had the circumstances been slightly different. Rasmussen (1988) summarises this viewpoint by suggesting that erroneous actions are the result of carrying out unsuccessful experiments in unfriendly environments. The margin between success and failure is a fine one, however, so to understand why actions fail, it is necessary to understand why they also work.

1.3 The Taxonomic Approach

Most scientific study is underpinned by a well-defined classification system or taxonomy of the relevant phenomena. The taxonomy provides a frame of reference for the study, and enables other researchers to evaluate the results of that study. In the field of human error research, several taxonomies have been developed, although there is no universal taxonomy which serves all the various purposes of error research (Senders & Moray, 1991). This point is tacitly accepted in Hollnagel's (1998a) Cognitive Reliability Error Analysis Method (CREAM), which provides a generic framework within which a domain specific taxonomy of causes and effects of erroneous actions can be developed.

Rasmussen's (1976) Step Ladder Model has been used as the basis for a number of taxonomies that describe human performance in the operation of complex control systems. These taxonomies include those of Rasmussen (1982); Rouse and Rouse (1983); and that used by Reason's (1990) Generic Error Modelling System (GEMS). Taxonomies have also been developed in industry, such as TAXAC (Brazendale, 1990) which was developed by the UK Atomic Energy Authority's Safety and Reliability Directorate in 1980-81, and the Human Error Taxonomy (Bagnara, Di Martino, Lisanti, et al. 1989) which was funded by the Joint Research Centre of the European Commission.

The critical factor in generating a taxonomy of erroneous behaviour is the decision about the level of abstraction to use for categorisation. Determining an appropriate level of abstraction requires that the purpose of the research be carefully defined. A useful rule of thumb is to model behaviour at a level which allows remedies to be generated to facilitate the avoidance of a repetition of the same type of error in similar circumstances in the future.

1.4 Dealing with Errors in Complex Systems

The increasing use of automation, and computers in particular has led to the development of increasingly complex systems which are used to control external processes such as nuclear power generation, chemical production, or the flight of an aircraft. The increase in complexity, however, seems to have given rise to new ways for the operators to make errors. In some case the errors have given rise to major accidents (e.g. Perrow, 1984; Kletz, 1994; Faith, 1996). Perrow (1984) even goes so far as to suggest that such accidents should be considered as normal. At least part of the problem lies in what Bainbridge (1983) calls the *Ironies of Automation*, which include the fact that the operator is taken out of the control loop for the most part, but still expected to intervene when the system goes wrong.

There is a growing acceptance of the need to manage errors (Frese & Altmann, 1989; Wioland & Amalberti, 1998) for two main reasons. The first is that the complete eradication of error is an unattainable goal. Amalberti (1998) has noted that experts actively ignore some errors which have benign consequences, and also tend to recover from routine errors. In the design of complex systems, the inevitability of errors is usually taken into account by adopting a defence in depth philosophy, whereby several levels of defences are included in order to make it difficult for an erroneous action to give rise to an accident. In such instances the accidents only arise when the gaps in each of the levels of defence coincide.

The second reason is that errors provide a learning opportunity (e.g. Petroski, 1985). This is one of the strongest arguments for keeping the operator in the loop. The only way that the operators can fully understand what is going on is if they are kept informed about what the system is currently doing (and why). A failure to keep the operators informed can lead to what have been euphemistically described as *automation surprises* by Sarter and Woods (1995), whereby the system does something which the operators do not understand in the current context.

There are several major accidents that have occurred during the operation of complex systems, in which state misinterpretation has been acknowledged as a major factor (e.g. the nuclear disaster at Three Mile Island; Perrow, 1984). The state misinterpretation occurs when the operators form a belief that the system being controlled is in a known, normal state, whereas it really is in a state that the operators have only rarely (or never) previously encountered. The factors surrounding the issues of state misinterpretation have received little or no direct attention in human error research, however, beyond the acknowledgement that state misinterpretation does happen.

1.5 Thesis Aims

The general aim of this research is to investigate the role of state misinterpretation in the operation of complex systems. Within this broad remit there are more specific aims:

- To investigate the general role played by the system state in the operation of complex systems, and use this, in combination with known examples of state misinterpretation to formulate a definition of the concept.
- To use the definition of state misinterpretation as a basis for gathering data on instances of state misinterpretation. A taxonomy of types of state misinterpretation shall be developed by categorising the data appropriately.
- To investigate the underlying causes surrounding state misinterpretation. The goal here is to try and establish if there any common causes which can be identified. Once any common features or patterns of behaviour have been identified it should be possible to start to determine how the various types of state misinterpretation can be prevented, possibly by automation, and which types give rise to situations which have to be managed by the operators.

1.6 Thesis Overview

In Chapter 2 the concept of state misinterpretation is introduced, initially by means of examples based on descriptions of real accidents. General issues of taxonomies of human error and classification schemes are introduced. A working definition of state misinterpretation is formulated based on the standardised view of the operator's control task.

Chapter 3 considers issues of data collection are. In particular, how, why, where and when to gather data are all discussed. Three potential methods are considered in light of the specific needs of this research, with the use of archive data being selected as the best compromise. Several possible sources of archive data are identified. A set of criteria was established to guide the selection of the data source; the NASA Aviation Safety Reporting System incident database was chosen.

Chapter 4 combines the definition of state misinterpretation established in Chapter 2, with the findings from Chapter 3 to develop and apply a scheme for extracting incident reports from the ASRS database which relate to state misinterpretation. Three database searches are described. The first takes a general view of the problem to identify how many reports there are for pilots and air traffic controllers. The second refines the results of the first search by focusing in on expert pilots and controllers, defining expertise in a way that the database can understand. The third search refines the results of the second search by concentrating on a shorter time frame. The reports retrieved by the third search were individually examined to remove any false positives.

Chapter 5 provides an overview of the Cognitive Reliability and Error Analysis Method (CREAM). It also describes how the data is extracted out of the database into a more suitable tool for analysing the data. Software tools for analysis in general are considered.

Chapter 6 describes how the CREAM is adapted for use in retrospective analysis of incident reports from the aviation domain to form the CREAM-AIR method. The analysis of the data is also described.

Chapters 7 through 9 present the results of the analysis of the data. Three levels of abstraction are used. Chapter 7 looks at the data at the lowest level of individual actions and events using a static point of view. Chapter 8 looks at the data in terms of sequences of actions and events, with particular reference to the way that the actions and events are linked together. Chapter 9 looks at the data at the highest level of trees. The trees are represented pictorially, and methods for quantitatively comparing trees, and defining the relation between actions and error modes are presented.

Chapter 10 presents the overall conclusions of the research, putting it into context with related work. Some suggestions are made about how to try and mitigate the effects of state misinterpretation. Finally, areas of future work are identified.

2 State Misinterpretation

In this chapter the concept of state misinterpretation is introduced and examined in some depth. The chapter begins with an illustrated overview of state misinterpretation, using examples, and the size of the problem is briefly considered. State misinterpretation is then compared and contrasted with similar existing phenomena, before relating the concept to theories of human error. The chapter concludes with a more detailed analysis of state misinterpretation, leading to the formulation of a working definition, before starting to quantify the problem in the aviation domain using reported incidents from the ASRS database.

2.1 An Illustrated Overview of State Misinterpretation

State misinterpretation has been acknowledged as a significant factor in several major accidents which occurred during the operation of complex systems. The misinterpretation occurs when the operators form the belief that the system is in a particular state which they recognise. In reality the system is in a different, possibly abnormal, state which the operators have either never, or else only rarely previously encountered.

The problem of state misinterpretation is perhaps best illustrated by looking at examples of accidents where state misinterpretation was implicated. Three such examples are considered below, each taken from a different domain.

2.1.1 Example 1: Three Mile Island

One of the most widely known accidents in which state misinterpretation was implicated is that which occurred at the Three Mile Island nuclear power plant in Harrisburg, Pennsylvania, in March, 1979 (e.g. Kletz, 1994). At the time of the accident, Three Mile Island was using pressurised water reactors (PWR) to generate nuclear power—Figure 2.1 shows a simplified schematic of a PWR. Nuclear fission generates heat in the core of the reactor, which is kept cool by the primary water that is pumped around the core. This water is kept under pressure to prevent it from boiling (hence the name pressurised water reactor). The temperature of the secondary water is raised by a heat exchange process (using the heat from the primary water). The secondary water *is* allowed to boil, and the steam it produces drives a turbine to generate power. After passing over the turbine, the secondary water condenses and the resultant condensate is recirculated to be heated by the primary water again. All of the radioactive components of the system are encased in a containment building, to prevent their escape in the event of a leak.

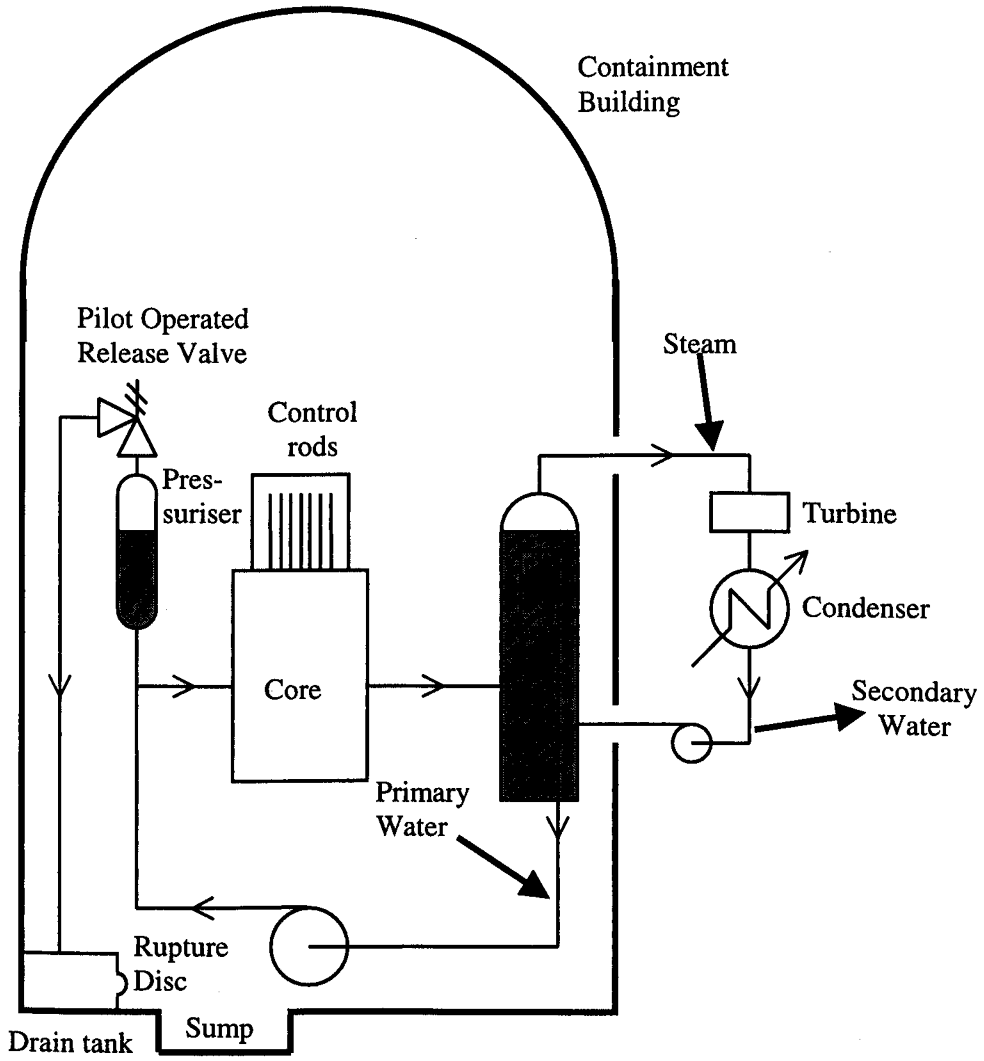

Figure 2.1. Simplified diagram of a pressurised water reactor as used at Three Mile Island. (Adapted from Kletz, 1994).

After passing through the condenser, the secondary water flows through a resin polisher unit (not shown in the figure) which is used to remove any traces of impurities from the water. At the start of the accident one of the paths through the resin polisher became choked. The operators noticed the blockage, and tried to clear it using the instrument air lines (rather than a separate higher pressure source of compressed air). The pressure of the instrument air was lower than that of the water, which led to some water being sucked into the instrument air lines. This water got into several instruments, causing them to fail, which ultimately resulted in the turbine going off-line.

When the turbine stopped, the flow of secondary water was halted, which resulted in the process that removes the heat from the core stopping. The controlled fission process in the core was shut down after a few minutes, which stopped the heating process, although the temperature continued to rise due to basic radioactive decay. Initially the heat being produced amounted to about 6% of the total load, which dropped back to 1.5% after an hour, but the heat being produced by the radioactive decay caused the primary water to boil, and start to evaporate. This resulted in the Pilot Operated Relief Valve (PORV) on the primary circuit of the reactor lifting, and the automatic start up of pumps to replace the water that was being lost through evaporation. The problem was that the PORV stuck open.

The instrument panel appeared to indicate to the operators that the PORV was closed, whereas what it really showed was that the signal to close the PORV had been issued. There were several other indications that the PORV was still open, however:

1. The PORV exit line was 50° hotter than normal (140° instead of 90°), but the operators attributed this problem to the residual heat.
2. The pressure and temperature of the primary water were both lower than would normally have been expected.
3. The level in the building containment sump was higher than normal.
4. The primary water circulation pumps were oscillating.

Counter to these indications, however, was the fact that the level in the pressuriser was high because it was raised by bubbles of steam. The operators attached more salience to this fact, and the PORV indicator, and either ignored, or explained away the other indications.

As part of their training, the operators had been told that having too much water in the reactor was dangerous. So, when they noticed the water level rising they shut down the pumps that were replenishing the primary water supply, in accordance with what they had previously been taught. This was the only action taken by the operators, and it merely exacerbated the problem. If the operators had done nothing, the reactor would have cooled of its own accord.

The net result was that the water level in the primary circuit fell, which led to the top of the core being uncovered. Steam reacted with the alloy cans which are used to enclose the ends of the uranium rods, and this caused hydrogen to be generated. Simultaneously, the steam that was being discharged through the PORV condensed in a drain tank, which then overflowed into the containment building sump, and the contents were automatically pumped outside the building.

The damage in this accident did not arise until two hours after the start. This damage could have been prevented if the operators had realised that the real system state was that the PORV was stuck open, and that the primary water level was low. The operators had developed a mind-set, however, sticking to their original diagnosis even in the face of overwhelming evidence that their perceived system state was different to that of the plant.

It should be noted that it would be an over-simplification to attribute the incident at Three Mile Island to operator error. The operators had not been adequately trained to deal with the situation which arose, and were working under stressful conditions. The crucial instrument indication was misleading, and there were of the order of 100 audible alarms simultaneously sounding to indicate the detection of alarm conditions.

2.1.2 Example 2: TWA Flight 514

On 1st December, 1974, Trans World Airlines flight number 514 crashed into a hill about 25 miles north west of Dulles International Airport, Washington, DC (e.g. Landsberg, 1998). The flight was descending for an approach to the airport in instrument meteorological conditions (IMCs) because the weather conditions were below the set minima required for visual flight rules to operate. All of the occupants of the aircraft were killed, and the aircraft destroyed.

The subsequent investigation by the National Transportation Safety Board (NTSB) concluded that the probable cause was the crew's decision to prematurely descend to the minimum approach altitude of 1800 feet. In other words, the crew believed that they had reached a position where it was safe to descend to 1800 feet, whereas they were really some way short of that point. The decision was apparently based on inadequate and unclear air traffic control (ATC) procedures which meant that the flight crew and the controllers misunderstood their responsibilities in airport terminal areas under IMCs. The flight crew should have spotted that the safe altitude was higher than 1800 feet from the plan view of their approach charts.

The contributing factors to the accident, listed in NTSB report NTSB-AAR-75-16 were:

(1) The failure of the FAA to take timely action to resolve the confusion and misinterpretation of air traffic terminology even though they had known of the problem for several years;

(2) The issuance of an approach clearance when the flight was still 44 miles from the airport, flying along a route which did not have clearly defined minimum altitudes;

(3) Inadequate depiction of altitude restrictions on the profile view for the approach chart for the navigation equipment—VHF Omni-directional Radio Range/Distance Measuring Equipment (VOR/DME)—approach to the required runway at Dulles.

This accident, which is an example of what is euphemistically called Controlled Flight Into Terrain (CFIT), proved to be something of a watershed in the aviation industry in the USA. It was this accident which ultimately led to the development of the Aviation Safety Reporting System (ASRS; Reynard, et al., 1986). The accident is of particular interest here because it provides an initial insight into the complexity of the system state, which is effectively distributed across multiple agents as well as across multiple pieces of equipment.

2.1.3 Example 3: The Farmsum Tanker Incident

In December 1982, three of the four men cleaning the number 6 hold of the *Farmsum* tanker in the Atlantic Ocean were killed when the partition between the number 6 hold and the number 5 hold collapsed (Wagenaar & Groeneweg, 1987). The three men were pinned to the wall as 6000 tons of water flooded into the hold that was being cleaned. The men had all believed that the number 5 hold was empty.

The problem was that the *Farmsum* had been travelling with water in hold number 4 as ballast, whilst hold number 5 was empty. During bad weather a leak developed between holds 4 and 5. As the ballast water flowed into hold 5, the first mate noticed that the level in hold 4 was going down. The first mate misinterpreted the situation, assuming that the problem had just arisen because the sluice valve in hold 4 had been left open. This state of affairs appeared to be confirmed by the fact that the sluice water which was used for cleaning hold 6 could not be removed. The first mate attributed the falling water level in hold 4 to the fact that the sluice pump was removing water from that hold. In order to maintain the ballast level, the first mate responded by refilling hold 4.

Although there was a large amount of water missing from hold 4, the first mate's explanation of events is plausible, given that he did not know how long the (low capacity) sluice pump had been operating. The problem was exacerbated by the fact that he was currently concerned with discovering why it was not possible to pump the water from hold 6. Initially he thought that the pump was broken, but engineers in the operating room knew that it had definitely been working. When the open sluice valve in hold 4 was discovered, this suggested a new explanation for inability to pump water from hold 6, and the reduced water level in hold 4 merely served to confirm this new hypothesis. The first mate accordingly shut off the sluice pump, and started to refill hold 4. Unknowingly he ended up filling holds 4 *and* 5, which eventually led to the failure of the partition between holds 5 and 6, resulting in the loss of life.

Although the situation was exacerbated by the faulty sluice valve indicator, a closer check on the amount of water required to refill hold 4 would have highlighted the problem. Nearly 6600 tons of water were required to refill hold 4, which was close to the hold's maximum capacity. In addition, the sluice pump is a low capacity pump, which would not have been able to remove the missing 3300 tons of water from hold 4 in the time available.

2.1.4 The Scope of State Misinterpretation

The examples provide some initial idea of the variety and scope of state misinterpretation. The concept is not simply limited to computer based control room systems, nor is it peculiar to any one particular domain. Further evidence is provided by Kletz (1991), who reports a number of similar incidents that have occurred in the railway transport industry and in the chemical process industry. The examples also show the very general nature of the state of a system, which is not just limited to the controls and displays with which the operator may directly interact, but also includes aspects of the external environment.

2.1.5 The Size of the Problem

Some insight into the size of the problem of state misinterpretation can be gleaned from previous research which has investigated multiple incidents where state misinterpretation was implicated. In a study of 99 simulated emergency scenarios, using 23 experienced nuclear plant operating crews in 8 different events, Woods (1984) discovered 19 instances of state misinterpretation by the operators. Similarly, Wagenaar and Groeneweg (1987) analysed 100 marine accidents reported by the Dutch Shipping Council during the period 1982–1985 and classified all the errors that were detected. Of the 344 errors detected, 60 involved the generation of an incorrect hypothesis—which often follows from state misinterpretation. Furthermore, 50 of the accidents were also attributed to habitual behaviour. In other words the operators believed the system to be in one state and acted in accordance with that belief, when the system was really in another state.

Whilst it is difficult to make direct comparisons between the two studies because of the different ways in which the data was collected and analysed, the percentage of instances of state misinterpretation in each of the domains is relatively high. In the nuclear power domain, albeit using simulated emergencies, state misinterpretation was discovered in 19.2% (19 out of 99) of the incidents. In the shipping domain the number of errors where state misinterpretation may have occurred is at least 14.5% (50 out of 344 classified errors) and could be as high as 32.0% (110 out of 344 classified errors) if all of the false hypothesis errors can be attributed to state misinterpretation.

2.2 Related Concepts

There are several psychological concepts in the literature which have some similarities to state misinterpretation:

(a) Habit capture (e.g. Reason & Mycielska, 1982; Reason, 1990)

(b) Cognitive lockup or cognitive tunnel vision (Moray, 1987)

(c) Garden path reasoning errors (Johnson, Moen & Thompson, 1988)

(d) Fixation or perseverance (De Keyser & Woods, 1990; Masson, 1991)

(e) Set effects: Functional Fixity, Einstellung, and Aufgabe (e.g. Kletz, 1991)

(f) Mode errors (Sarter & Woods, 1995)

The difference between state misinterpretation and habit capture (e.g. Reason, 1990) is a subtle one. Habit capture arises when the intended behaviour in a particular context is unconsciously replaced by a more habitual behaviour that is normally associated with that context. Reason and Mycielska (1982) report several examples from subjects who were asked to keep a diary of all the errors they made over a selected time period. A typical example is where a person goes into the bedroom in order to change into something more comfortable for the evening, and ends up putting on their pyjamas, as if they were getting ready to go to bed. The habit capture occurs when the person has an intention to perform a particular action or sequence of actions—in the given example changing into something more comfortable—but these actions are *unconsciously* displaced by a habitual sequence of actions that the person normally associates with the context at that time—getting ready for bed.

When state misinterpretation occurs, however, there is no conflict between the intention and the context which can lead to the habit capture. The operators *do* notice the context, because it is the context that usually leads to the intention about what to do next. The operators fail to recognise the context as being novel, however, and so they decide to perform some sequence of actions based on the context that they believe they have recognised. In other words, state misinterpretation leads to a *conscious* choice of which actions to perform.

The concepts shown as items (b) to (e) in the list are all slight variations on a theme, and all relate to diagnosis of an existing problem. Cognitive lockup occurs when the operators get stuck in a single mode of behaviour. Instead of exploring the environment and updating their mental model of the system, the operators try to make the data fit with their existing models. Similarly, garden path reasoning errors arise when early situational cues suggest an incorrect solution, whilst later, often weaker cues suggest the correct answer. Fixation or perseverance, and set effects also describe the situation where the operators stick to the first solution they find even when the evidence may suggest that the chosen solution is incorrect, or has undesired side-effects.

The main similarity between state misinterpretation and the diagnosis related problems (items (b) to (e) in the list) is that the operators believe that their perception of the current situation is correct, when it is not. State misinterpretation can occur at any time during normal system operation too, however, rather than just during diagnosis. When state misinterpretation occurs, the system operators may not be aware that there is a problem, whereas the other concepts happen when trying to solve an existing problem which the operators are aware of.

There is also some similarity between state misinterpretation and mode errors (Sarter & Woods, 1995). On one level it is possible to simply regard state misinterpretation as an instance of mode error, although this is something of an oversimplification. Mode error is generally used to relate to the specific operating modes that a system has been designed to take on. Each of these modes will normally consist of a range of different, but expected, system state configurations. State misinterpretation may coincide with a mode error if the operator identifies an infrequently occurring system state as one that is familiar. The mapping between state misinterpretation and mode error is not straightforward, however, since there will also be cases where a system can enter a state that was not explicitly considered at design time. This type of situation is more likely to occur in systems consisting of hundreds or thousands of individual elements, where it is not possible to consider and account for all possible configurations during system design.

2.3 State Misinterpretation and Taxonomies of Error

The notion of defining different types of errors, and generating taxonomies of errors from these definitions is an important one. Indeed, Senders and Moray (1991) suggest that any empirical science fundamentally depends on the establishment of taxonomies. In the absence of taxonomies it becomes impossible to identify the phenomena that are being investigated. There are several different taxonomies of errors in existence, and the differences between them can largely be ascribed to differences in the purposes of the underlying research. Rasmussen, Pejtersen, and Goodstein (1994), for example, identified six different research perspectives that can be adopted:

1. The common sense perspective, which is used in generating explanations of unusual events in lay terms.

2. The scientist's perspective which is used when trying to understand human behaviour.
3. The reliability analyst's perspective which is adopted when trying to evaluate human performance.
4. The therapist's perspective which is used in trying to improve human performance.
5. The legal perspective which is used in trying to find someone to apportion blame to.
6. The designer's perspective, which is used in trying to improve the way in which the system is configured.

Each of these different perspectives requires its own particular taxonomy. The scientist's perspective—adopted here—normally requires a taxonomy which focuses on the actions and events that make up the human behaviour. In contrast, the legal perspective, for example, will employ a taxonomy which is more concerned with allocating responsibility for what happened.

The difficulty of classifying a particular instance of error as being an instance of state misinterpretation is typical of the general problem of identifying errors, even where a taxonomy exists. It is impossible to define an individual action as erroneous without reference to the context which defines what the appropriate action should have been. The main reason why hindsight is needed to judge a particular action as erroneous is that it is difficult to determine the operator's beliefs and intentions when that action was performed. Furthermore, one also needs to know what the expected outcome of that action was. In general, the expectation will be defined by the operator, but it can also be defined by another agency, such as a supervisor, the operator's employer, or even a regulatory body.

2.4 Existing Classification Schemes

The only existing classification schemes which explicitly mention misinterpretation are those of Rasmussen (1982) and Rouse and Rouse (1983), which is based on Rasmussen's work. Both of these schemes are fairly comprehensive, and are designed to cater for all stages of operator task performance. The schemes assume a model of performance in which operators normally carry out supervisory control in an iterative manner using the current state of the system to determine which course of actions should be performed (e.g. Kelley, 1968). This control task is generally represented in a simplified form as an iterative sequence of steps similar to that shown in Figure 2.2. It should be noted that the figure is very much a static representation of a dynamic situation. In reality, there may be some overlap between the steps, and steps may be omitted or re-ordered.

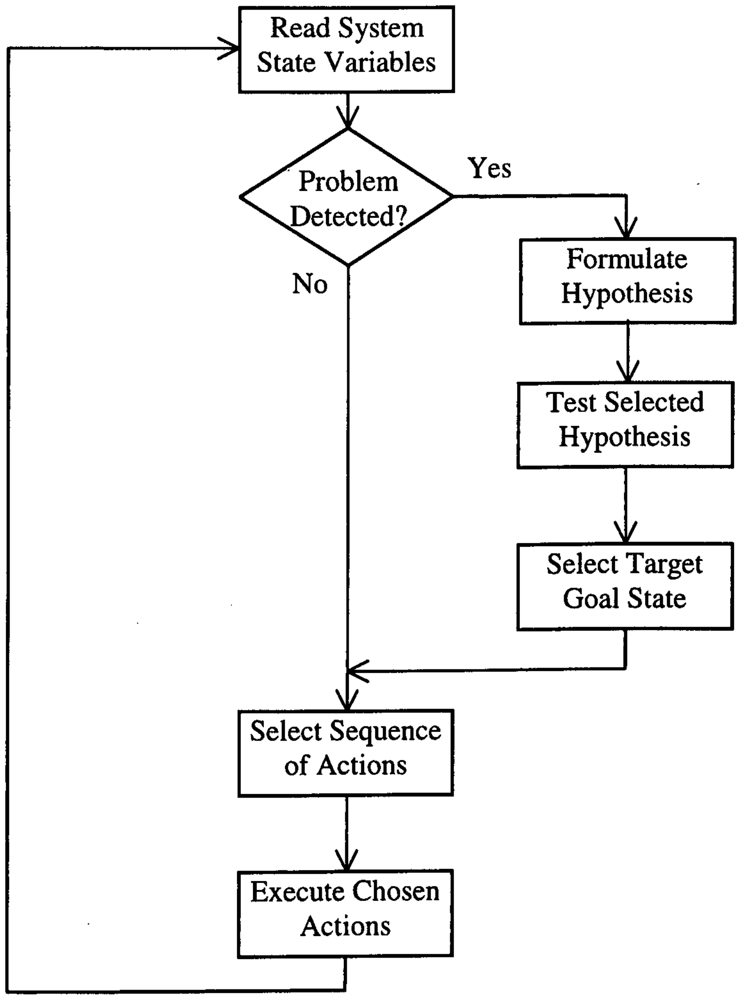

Figure 2.2. Simple conceptual model of operator's control task.

2.4.1 Rasmussen's Classification Scheme

The Rasmussen classification scheme is based on the step ladder model of human performance (e.g. Rasmussen, 1986). In the step ladder model performance is statically represented by a sequence of states of knowledge, connected together by information processes. Behaviour is divided into three levels: skill-based, rule-based, and knowledge-based, as shown in the simplified diagram in Figure 2.3. In reality performance involves a more complex interaction between the three levels than that shown in the diagram.

Skill-based behaviour—shown towards the foot of Figure 2.3—is essentially automatised, requiring little or no conscious control, and generally occurs only in familiar situations where sensorimotor skills can be utilised. Rule-based behaviour—in the middle of the figure—is most common, and consists of the execution of stored rules or procedures which are have preconditions that match the current state of the system and its environment. Knowledge-based

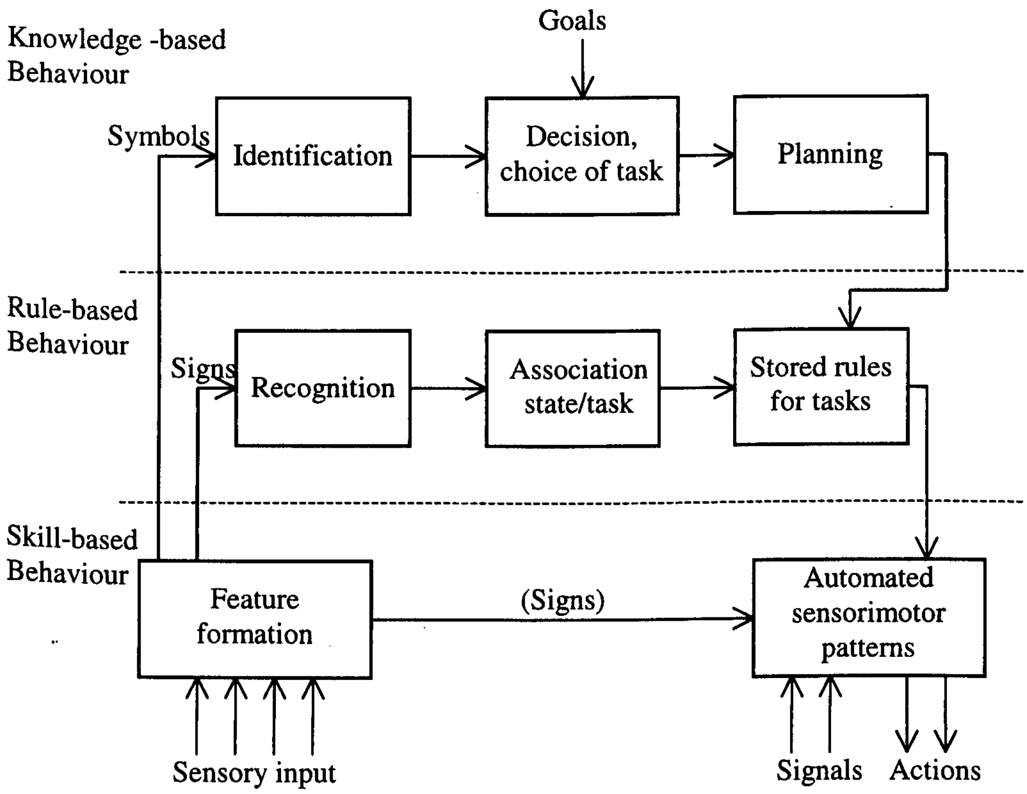

Figure 2.3. A simplified diagram of the interrelationship between the three levels of behaviour.
(Adapted from Rasmussen, 1983).

behaviour—at the top of the figure—is slow and effortful, involving mechanisms such as trial and error as experiments are conducted on an internal model of the system being operated. The boundaries between the three types of behaviour are not clear cut, and performance can be optimised by the use of shortcuts, which are learned through experience.

The model is used to help identify and organise the different groups of categories of human behaviour. Under the general group described as *Mechanisms of Human Malfunction* there is a category which relates to failures of input information processing. This category explicitly includes *misinterpretation*, along with *information not received*—the information was not seen or was not sought—and *assumption*—the information was assumed rather than observed.

Rasmussen's scheme is intended to be a general one, in which the structure is more important than the individual members of the various categories. The scheme does not offer a specific description for *misinterpretation*, although an associated guide is intended to be used to identify which *Mechanism of Human Malfunction* occurred. The guide is essentially a decision tree, in which *information misinterpreted* is one of the three possibilities—the others are *information not received* and *assumption*—if the answer to the question "Does the operator correctly collect the information available for his analysis?" is "No". In essence the guide supports a level of abstraction higher than that of state misinterpretation, because it does not offer any way of discriminating between the three types of *Input Information Processing* failures.

2.4.2 Rouse and Rouse's Classification Scheme

The Rouse and Rouse (1983) model—a development of Rasmussen's earlier model—was used as the basis for generating the general categories of error in their classification scheme. Each of the six processing stages identified in their model is associated with its own general category of errors, which are described as behaviour-oriented. These general categories are each divided into several specific categories, most of which have been proposed elsewhere.

The model is intended to be adapted to different domains and situations as circumstances dictate. The general category of most interest here is *Observation of System State*, which corresponds to the box labelled *Read System State Variables* in Figure 2.2. Within this general category, Rouse and Rouse identified six different specific categories of errors, which they describe as more task-oriented (in contrast to behaviour oriented):

1. Excessive—the operator repeats previous correct readings of appropriate state variables.
2. Misinterpreted—the operator makes an incorrect interpretation of correct readings of the values of the appropriate state variables.
3. Incorrect—the operator makes incorrect readings of the values of the appropriate state variables.
4. Incomplete—the operators fails to observe sufficient values of the appropriate state variables.
5. Inappropriate—the operator observes the values of inappropriate state variables.
6. Lack—the operator fails to observe any of the state variables.

The six categories overlap somewhat. The *Lack* category is simply the extreme case of the *Incomplete* category. In some cases there may be difficulty in distinguishing between the *Excessive* and the *Incomplete* categories because they can both be considered as instances of the operator not knowing the required number that should be read. Matters can also be complicated when trying to distinguish between the *Misinterpreted* category and the *Incorrect* category, since it requires making a fine judgement on whether an incorrect reading was made or a reading was simply interpreted incorrectly.

2.5 Towards a Working Definition of State Misinterpretation

The classification schemes of Rasmussen, and Rouse and Rouse, which explicitly include (state) misinterpretation are of a general nature. They are designed to categorise errors which arise during all stages of performance of a control task. In contrast, this research is only concerned with the earliest stage of that task, which is where state misinterpretation arises. An alternative classification scheme is therefore proposed in this section, which is designed to more closely meet the particular requirements of this research. This classification scheme is then used to formulate a working definition of state misinterpretation.

2.5.1 State Misinterpretation Versus State Interpretation

State misinterpretation is, by definition, a failure in state interpretation. The first step in defining state misinterpretation is therefore to define what is meant by state interpretation. Once such a definition has been established, the characteristic features of state interpretation can be determined. Then, by examining the ways in which these characterising features can vary, the different ways in which state interpretation can fail can be identified and appropriately classified. It should then be possible to start to identify the potential causes of failures in state interpretation—state misinterpretation—which should ultimately make it possible to identify possible ways in which state misinterpretation can be managed.

2.5.2 State Interpretation and the Perceived State of the System

The system being controlled by the operator can be represented by three different types of state which simultaneously coexist (Boy, 1987):

1. The real state.
2. The perceived state.
3. The desired or expected state.

The real state of the system consists of all the possible facts about the system and its environment. By definition it is infinite, and consequently it cannot all be simultaneously accessed by the operator. It would be impossible to keep track of the values of each of the system's thousands of elements individually: the operator would have to continuously take readings of the system variables which would leave little time to determine the overall system state and carry out the relevant control actions. Complex systems are therefore generally controlled directly on the basis of the perceived state of the system, which corresponds to the results of the box labelled *Read System State Variables* in Figure 2.2, using a process described as recognition primed decision making (Klein, 1997). Operators take readings of the variables which define the current system state, and synthesise the values to yield the perceived state of the system. This perceived state, which is essentially the result of state interpretation, will be optimally correct when the operator:

- knows which system variables define the current system state;
- accesses the values of all of these variables; and
- correctly interprets the values of these variables.

If the operator recognises the perceived state as normal, or finds that it matches the expected system state, a set of actions which is consonant with the goal of the current task is executed. If there is a mismatch between the perceived and expected states, the operator analyses the differences to determine the cause of the variance. When the problem has been diagnosed the operator acts to bring the system to a normal or safe state, as appropriate.

2.5.3 Identifying Different Types of Perceived State

Each of the three factors that determine the operator's perceived state can be considered as a binary dimension, in that it is either true or false for any particular situation. By logically combining the truth values for the three factors, eight different types of perceived state can be identified, as shown in Table 2.1. The labelling of the columns underneath the heading *State Variables* relates to the factors that affect the correctness of the perceived state, identified above. So, *Known* means that the operator knows which system variables define the current system state when *True* appears in that column. Similarly, when the operator accesses the values of all of these variables, and correctly interprets those values, *True* will appear in the *Accessed* and *Interpreted* columns respectively.

The table illustrates the major difficulty in defining and identifying state misinterpretation: none

Table 2.1. The Dimensions of the Different Types of Perceived State

State Variables			Perceived state
Known	Accessed	Interpreted	
True	True	True	Optimally correct
True	True	False	Incorrect
True	False	True	May be coincidentally correct
True	False	False	May be non-optimally correct
False	True	True	Incorrect
False	True	False	Incorrect
False	False	True	Incorrect
False	False	False	Incorrect

of the different types of state will definitely lead to problems. Aircraft pilots, for example, can successfully fly their aircraft even when their perceived state is not optimally correct, as suggested by rows three and four in the table. If the pilot does not access all the required variables, the perceived state may be coincidentally correct in that all the variables that have not been accessed are within process limits. If the pilot accesses too many variables, then the perceived state may still be correct, but the pilot will simply have wasted effort accessing and interpreting the additional data. It is these combinations of factors that can give rise to a perceived state which does not match the true state of the system. In other words, state misinterpretation can potentially occur when the pilot knows all the state defining variables, but fails to access all of these variables, and may, or may not, correctly interpret the values of the state defining variables. This working definition of state misinterpretation is critical to the rest of this research and will be used to select an appropriate set of incidents for further investigation and analysis.

A lack of knowledge about the state's defining variables is assumed to preclude the ability to generate the correct perceived state. This assumption is based on the fact that operators need to know which variables define the system state in order to be able to distinguish between normal and abnormal states. So, for example, pilots and air traffic controllers have to demonstrate advanced levels of knowledge and competence before they are allowed to fly an aircraft or control real traffic. Pilots and air traffic controllers should therefore know the state defining variables, so, for these two groups of operators, the expected number of incidents where the state defining variables are unknown (shown in the bottom four rows of Table 2.1) should be close to zero.

2.6 Summary

In this chapter the concept of state misinterpretation was introduced by means of examples taken from different domains, which provided some illustration of the general nature of the problem. The general nature of the concept was compared and contrasted with existing psychological concepts, before considering the position of state misinterpretation in relation to existing taxonomies of human error. A taxonomy of state interpretation was then drawn up, on the basis that state misinterpretation is simply a failure in state interpretation. A working definition of state misinterpretation was then formulated, which takes some account of the expertise of the operators, and developments in the field of naturalistic decision making (e.g. Klein, Orasanu, Calderwood, & Zsambok, 1993; Zsambok & Klein, 1997), in particular the concept of recognition primed decision making. In the next chapter, consideration is given to the issues surrounding the collection of data about instances of state misinterpretation.

3 Collecting Human Error Data

The issue of data collection is central to any study of human performance. It is an issue which causes some controversy, however, especially when dealing with expert performance, and there is no universally accepted best way of gathering such data. This chapter delineates some of the issues involved in deciding how data is to be collected in light of the constraints of human error research. In particular, the specific requirements of investigating state misinterpretation by highly skilled operators in complex domains are considered.

The first part of the chapter directly addresses the standard range of questions that are applicable to the collection of human performance data:

- Why gather data?
- What sort of data to gather?
- Where (and when) to gather data?
- How much data to gather?
- How to gather data?

The questions provide a basic framework for the discussion of the various issues involved. These issues interact and overlap, so it is not possible to answer each of the questions in isolation from the issues raised by the other questions. In answering the questions, three possible methods of data collection are identified. The general strengths and weaknesses of each of the methods are assessed.

Based upon the specific needs of this research a set of selection criteria is established and used to choose the most appropriate method for collecting data. The chosen method is then discussed in more detail in order to address any outstanding pitfalls of using that method, and the ways in which those pitfalls may be avoided.

The final part of the chapter considers potential sources of data. The suitability and availability of each of the different data sources are assessed when selecting the single source to be used here.

3.1 Why Gather Data?

This research is concerned with problems of state misinterpretation which arise as part of system performance. The approach proposed here is to identify the conditions that give rise to the deviations in behaviour in the first place, with the aim of generating a set of recommendations that can be used to inform the design or operation of systems. These recommendations will be based on the particular properties—events, features and behaviour—which facilitate state misinterpretation. In order to generate such a set of recommendations, it is necessary to first gather data about cases of state misinterpretation so that the underlying properties can be identified.

It is important not to try and read too much into the data for an individual case. Accidents, for example, should be regarded as unique in that the precise circumstances surrounding each individual accident are invariably different (Kletz, 1994). In order to avoid developing procedures that will only prevent a recurrence of exactly the same accident under exactly the same conditions, the details of the accident have to be analysed using different levels of abstraction. Using higher levels of analysis facilitates the identification of the properties that are common across several events. This type of approach was advocated by Toft and Reynolds (1994), who identified four types of isomorphism by using several levels of abstraction during the analysis of event data:

- Event isomorphism: separate events with different manifestations which lead to identical problems.
- Cross-organisational isomorphisms: all of the organisations involved are in the same industry.
- Common mode isomorphism: organisations of an apparently disparate nature, and in separate industries have similar qualities, which lead to the same problems.
- Self-isomorphism: internal and external contingencies within sub-units in a single organisation lead to the same problems.

The identification of the common properties of the selected events will form the basis for establishing the recommendations for improving the design or operation of the system. This approach is intended to complement other techniques for preventing accidents during the design or operation of systems, such as Hazard and Operability studies (HAZOPs; e.g. Kletz, 1999), or Probabilistic Risk Assessments (PRAs; e.g. Stewart & Melchers, 1997), which are widely used in the Chemical Processing and Nuclear Power industries.

3.2 What Sort of Data to Gather?

For the purposes of this research, data relating to situations in which state misinterpretation occurred needs to be collected for analysis. This data needs to be sufficiently detailed to allow for:

- the detection and identification of the occurrence of state misinterpretation
- the determination of the properties or features associated with each instance of state misinterpretation

State misinterpretation is categorised here as a type of mistake, in both the general sense of the word, and the narrower sense used by Reason (1990) who restricts the use of the term *mistake* to describe a failure in the overall planning process. The fundamental principle underpinning this research is that people can learn by reflecting on previous mistakes (in the general sense). The importance of this general tenet was recognised by Mach (1905) who pointed out that:

> A clearly recognised error, by way of corrective, can benefit knowledge just as a positive piece of knowledge can. (p. 84).

Indeed, this principle underpins much of the progress of engineering as a discipline (Petroski, 1985). When engineered artefacts (bridges, roads, planes, computer systems and so on) fail, a post mortem is normally carried out to determine precisely what went wrong, and why. The results of this analysis are then published so that others can be made aware of the reasons for the failure and guard against a recurrence in the future.

If people are to learn from previous mistakes, the data associated with those mistakes need to be made accessible and available for further analysis. So, for the purposes of this research, the details of the sorts of erroneous actions that are associated with the occurrence of state misinterpretation during the system performance need to be available for analysis. The feasibility of adopting this approach of analysing data from previous accidents to see what can be learned was demonstrated by Rasmussen (1980). A set of published event reports from the nuclear power industry was analysed to identify the different error modes involved. When the error modes were categorised, they highlighted the mismatch between the operators and the systems that they were expected to use. Rasmussen concluded that there was a need for effective feedback on operational performance to inform the system design process, so that the resultant systems would fit the operators better.

An instance of state misinterpretation is not normally a directly observable event. What generally happens is that the operator forms a belief that the system is in one particular state, which is different from the true system state. The operator then executes one or more actions which are consonant with the perceived state of the system. It therefore follows that state misinterpretation does not directly cause accidents, although it may be an indirect cause.

In many cases the operating staff detect the effects of state misinterpretation whilst there is still time to correct the state of affairs before the situation develops into an accident. The term *incident* will be used here to describe any such events which involve an unsafe or potentially unsafe occurrence or condition (Chappell, 1994). This usage of the term is not universally accepted. van der Schaaf (1991a), for example, uses the term *near miss* to describe the same sort of events, and uses the term *incident* to encompass *both* accidents and near misses, in common with many in the chemical process industries. Similarly, in the nuclear power industry, a standard rating scale is used which runs from 0 to 7, with the term *incident* being reserved for the lowest 4 levels (0-3), rather than a single level.

Terminological ambiguities and inconsistencies abound in human error research. Senders and Moray (1991) noted this problem when they tried to get a panel of experts in the field to agree on a definition of terms such as *human error*. One reason for the problems may be historical, in that each of the application domains (aviation, chemical process, nuclear power and so on) independently developed its own terminology, using a vocabulary that was best suited to their industry. The differences only emerged when researchers began to compare and contrast performance across domains.

In general, incidents occur one or more orders of magnitude more frequently than accidents. The numerical relationship between accidents and incidents is often depicted using a pyramid (Bird, 1969, cited in Heinrich, Petersen, & Roos, 1980), as shown in Figure 3.1, although the relationship between the different levels in the pyramid should be viewed as a generalisation,

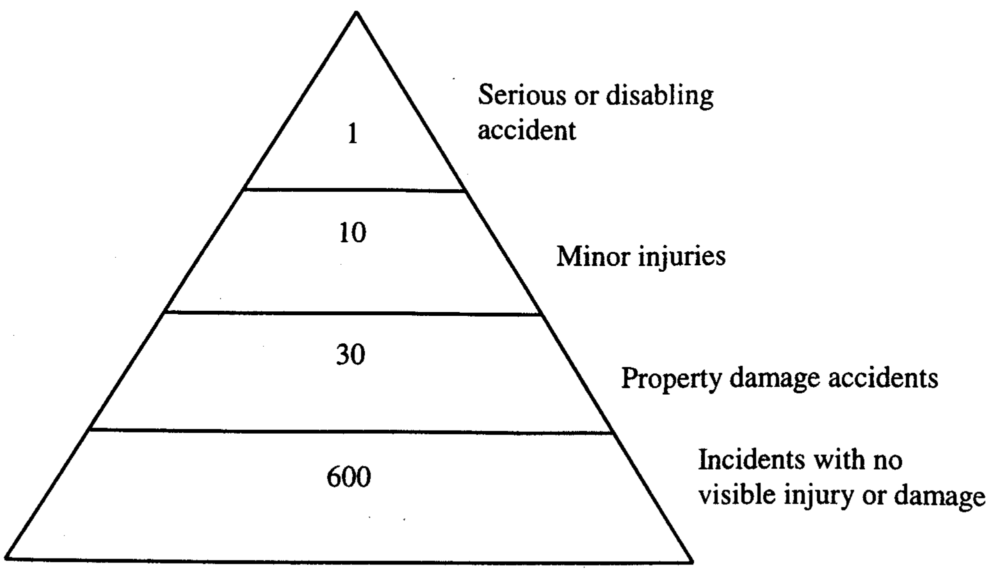

Figure 3.1. Pyramid relationship accidents and incidents.
(Adapted from Heinrich, Petersen & Roos, 1980.)

rather than a fixed mathematical relation (Heinrich et al., 1980; Groeneweg, 1994). The pyramid representation has been employed in many industries, although the terms used to describe the different levels in the pyramid vary somewhat. The most important point to note is that there is tacit acceptance that the accidents at the peak of the pyramid arise from the same foundations as the incidents at the base of the pyramid.

In a given time period the number of incidents that occur is much greater than the number of accidents that occur. So, given that the foundations of incidents and accidents are the same—an incident can be considered as an accident that was detected and prevented before it happened—collecting data about incidents in any given period will normally yield a much larger data set for analysis than if accident data is collected.

3.3 Where (and When) to Gather Data

In this section the general features which characterise those situations that will give rise to state misinterpretation are discussed. The identification of these features provides the basis for determining where to look for data relating to incidents involving state misinterpretation. At the most abstract level, the two main identifiable components are the system which is being operated, and the person who is operating that system. The complexity of the systems, and the expertise of the operators involved are discussed, before giving some consideration to the frequency of occurrence of state misinterpretation.

3.3.1 System complexity

State misinterpretation occurs relatively infrequently, but in those reported accidents where it has been implicated, the consequences were often disastrous, involving major loss of life, money, or production. In most of these cases the accident occurred during the operation of what Perrow (1984) calls a *complex system*, a concept he defines using the dimensions of interactions and coupling.

The interactions dimension is used to describe how the various components of the system are related. At one extreme lie linear interactions which are relatively well behaved, in that they occur in fixed, familiar sequences, and even when they are unplanned they are normally visible to the operator. At the other extreme lie complex interactions, which are more difficult for the operators to predict and understand, and their effects are often invisible. Most systems usually involve a mixture of linear and complex interactions. Normally there will be a large number of linear interactions, whilst the number of complex interactions is relatively small, even in the most complicated of systems. Measuring a system's complexity in terms of interactions is therefore a matter of degree, rather than a simple dichotomy.

The coupling dimension is used to describe the relationship between successive steps in the production or control process. At one extreme lies loose coupling, where the system includes an inherent amount of slack in the form of buffers which can be used during delays in the process. The steps in the process can also be reordered, thereby making the process more flexible. At the other extreme, tightly coupled systems depend on a fixed, often time critical ordering of the steps in the production process. In general, most continuous process systems are tightly coupled.

Although Perrow's (1984) definition provides a useful way of delineating system complexity, it is overly simplistic. As Perrow himself acknowledges, complex interactions can be made more linear by increasing the understanding of the underlying process, either through further research, or by increasing the training of the operators. Perrow also suggests that coupling can be reduced by introducing intermediate steps in a process such as holding inventories of intermediate products in storage tanks. The danger with this approach is that it may simply change the nature of the risk of an accident, rather than removing it. On chemical processing plants, for example, intermediate inventories are generally frowned upon, on the grounds that what is not stored cannot leak, catch fire, or explode (e.g. Kletz, 1991; 1994).

This research is mainly concerned with those domains where the operator uses, or interacts with, a complex system to monitor and control an external process, such as nuclear power generation, heavy chemicals production, or aviation. These *complex interactive systems* (Weir, 1991) effectively form a (large) subset of Perrow's complex systems. In complex interactive systems the operator interacts with the hardware or the software of the system. The complexity of the system is a function of the complexity of the domain (similar to Perrow's interactions), the complexity of the control requirements (similar to Perrow's coupling), *and* the complexity of the required operator interaction. Although the scope of this research is limited to state misinterpretation in complex interactive systems, the problem of state misinterpretation also occurs in other complex systems, such as rail, and road transport (e.g. Kletz, 1991; van der Schaaf, Lucas & Hale, 1991).

3.3.2 Operator skill

A high level of skill is required to operate complex interactive systems. The staff that operate such systems typically undergo extensive training to help familiarise them with the basic operation of the system. On top of this training they will normally have of the order of 10^2 or 10^3 hours experience of working with the system, at least some of which may have been gathered under the auspices of a more experienced operator or supervisor. It is not just the number of hours of experience that is important, however, it is also the variety of the experience over time. The link between the two is not necessarily a causal one, although there will normally be a good correlation between the length of time an operator spends working with a system, and the number of different system states that the operator will encounter over that period.

Over time, the operator's comprehension of the way that the system works will change, as the inadequacy and limitations of the operator's initial understanding of the system's workings—encapsulated in some form of mental model—become apparent. This need for a change of representation underpins Cognitive Flexibility Theory (e.g. Feltovich, Spiro, & Coulson, 1993) which describes how complex concepts can only be understood if they are correctly represented. The initial simplifications which are often employed when teaching people about such concepts can form barriers to forming a correct understanding of the whole of the concept.

3.3.3 Frequency of occurrence

Although accidents are reported relatively infrequently, it is generally possible to quantify accident rates a priori on a temporal basis. The figures vary considerably within and across domains, and are often defined with respect to the number of opportunities for the occurrence of a particular accident. Typically rates range between once per hundred, and once per thousand such opportunities. When such rates are converted to a temporal basis, these superficial rates reduce somewhat because the opportunity for the occurrence of particular accidents arises so rarely. Habberley, Shaddick and Taylor (1986), for example, in their study of maritime collisions suggested that the frequency of occurrence of accidents in which a ship is lost through a collision is about once every six ship lifetimes. This finding adds further weight to the preference for using incident data rather than accident data.

3.3.4 Key features summarised

On the basis of the discussion presented above, the following characteristics can be noted about the sorts of situations in which state misinterpretation may manifest itself:

- State misinterpretation arises during the operation of complex interactive systems.
- Complex interactive systems operators typically interact with the system hardware or software (operator or user interface) to control some external process.
- The operating staff are highly skilled, often with of the order of 1000 hours exposure to the system.

- The frequency of occurrence of state misinterpretation will be relatively low (although it will be higher for incidents than for accidents).

3.4 *How Much Data to Gather?*

There are three important factors which need to be considered when deciding on the amount of data required in order to draw valid conclusions from the analysis of that data: the level of detail; the scope of the incidents; and the number of incidents. The first two will fundamentally determine the effectiveness of the results of this research; the third is related to the practicality of analysing the data.

The quality of the results of the data analysis will reflect both the quality and the quantity of instances of different types of state misinterpretation. The quality of the incident data is important in that the level of detail and the scope of the incidents will determine how much information can be gleaned from each incident. The level of detail ideally needs to be such that it allows for the detection of variation across incidents, since it is this variation which forms the basis for the identification of different types of state misinterpretation. The data should therefore relate to a wide variety of circumstances and events, whilst still being representative of a homogeneous population of systems.

The number of incidents that are analysed is also of critical importance. The identification of different types of state misinterpretation necessarily depends on multiple occurrences of common properties across incidents to ensure that any detected isomorphisms are not simply a coincidence. Furthermore, having a larger data set should make it easier to identify and apply several different multiple levels of abstraction during the process of trying to identify isomorphic cases of state misinterpretation.

As part of the data analysis, a taxonomy of the different types of state misinterpretation will be developed. The quality and utility of the taxonomy will reflect the quantity of data that is analysed. Based on available evidence from existing studies carried out in the aviation domain, the minimum number of instances of incidents that are normally used in data analysis is of the order of 10^2. The final amount of data that will be collected will depend on the amount of time and effort required to causally analyse each of the incidents. The incidents have to be individually analysed and documented since there are no readily available tools readily that automatically perform causal analyses.

3.5 How to Gather Data?

There are three basic methods for gathering data on human performance when working in complex domains where the system can change dynamically without operator intervention: laboratory based experiments; field based observation; and archive data. The strengths and weaknesses of these methods are assessed below; the method best suited to the requirements of this research is then selected.

3.5.1 Laboratory based experiments

The first method is the traditional behavioural science method of using laboratory based experiments. The main advantage of this method is that it allows for the independent variables associated with a particular phenomenon to be experimentally controlled. By varying one (or more) independent variables, the effect on the dependent variable can be observed. There are several disadvantages to using laboratory based experiments, however.

The main drawback of this method is that there is a lack of face validity between laboratory based experiments and the real world situation. This lack of validity makes it inappropriate, at best, and impossible, at worst, to try to generalise from the results obtained in the laboratory to the situation in the real world. The use of laboratory based experiments largely ignores the current consensus of opinion among researchers which emphasises the important role that context plays in the shaping of human performance (e.g. Hollnagel, 1993b; Hutchins, 1995; Nardi, 1996).

Given the criteria of the situations needed to generate instances of state misinterpretation, it is a long and difficult task to develop and conduct laboratory experiments that would meet all the requirements. The first problem is the availability and selection of appropriate subjects. The subjects would need lengthy experience of operating the system being used in the experiment. Whilst it might be possible to use operators from the relevant domain, their experience is liable to be biased towards the particular system that they normally work with. The practicality of getting access to operators for the length of time needed to conduct the experiments also mitigates against using this approach.

The second problem is the choice of experimental task(s). Complex systems often require operators to perform multiple tasks, sometimes simultaneously. The difficulty is to gain access to a complex system (or an appropriate simulation) which could be readily deployed under laboratory conditions. The system would also need to be familiar to the subjects, in order to fulfil the criterion regarding expertise. Whilst some laboratories have their own high fidelity simulations, such as the Halden Error Assessment Programme (HEAP; Hollnagel, Drøivoldsmo,

& Kirwan, 1996) which has a nuclear power plant simulator, they are the exception rather than the rule.

The fact that erroneous behaviour is an inherent part of human performance also needs to be heeded. There is no single part of cognition which is responsible for the production of errors, which re-emphasises Mach's (1905) observation that:

> Knowledge and error flow from the same mental resources; only success can tell the one from the other. (p.84)

In other words, knowledge or data relating to correct performance is also needed in order to make an informed judgement regarding each potential instance of state misinterpretation.

Although it is possible to design laboratory experiments based on scenarios designed to lead to state misinterpretation, the practicalities of doing so are long, complicated, and expensive. There are potential problems in running subjects on such experiments too: if they can identify the purpose of the experiment they may behave more cautiously than normal, in order to deliberately guard against performing erroneous actions.

3.5.2 Field based observation

The second method is to carry out longitudinal observation of experienced operators. The main advantage of this method is that it guarantees the face validity of the data. In general, observation is a valid technique, particularly if the aim is to investigate human reliability per se. In light of Mach's observation about the relationship between knowledge and error, there is much that can be learned about operator performance under abnormal system operating conditions from observing performance under normal operating conditions. This research focuses on one specific aspect of operator behaviour, however, rather than on operator behaviour per se. The observational method therefore has a number of drawbacks in this particular situation.

The first drawback is the relatively low frequency of occurrence of state misinterpretation. Although there may be a deterministic element to the occurrence of state misinterpretation, it is very difficult to predict precisely when a sequence of events or actions giving rise to an observable error will occur. There is consequently no guarantee that state misinterpretation will occur even during extended periods of observation.

The second drawback is the high costs associated with extended observation and the subsequent analysis of the data. Even if it could be guaranteed that state misinterpretation would occur once in every 24 hour period, for example, then to gather enough data for 100 instances would require 2400 hours of recordings. Since analysis of recorded data can take an order of magnitude longer than the recording, the time taken to analyse such a large amount of data quickly becomes prohibitive.

The third drawback is the inherent adaptability of human behaviour. One of the characteristics of experienced operators is their ability to detect and recover from potential erroneous actions. In other words, recovery is performed before any unwanted consequences arise. Unless verbal reports are also recorded, which can be used to try and identify these recoveries, it will be difficult to detect where a recovery has taken place.

The fourth drawback is that the mere presence of an outside observer may affect the operator's performance. In particular, the operators may be more careful than normal if they know that their performance is to be captured on video tape for subsequent analysis. This may help to reduce the frequency of occurrence of state misinterpretation, thereby requiring even longer periods of observation in order to gather enough data.

3.5.3 Archive data

The third possible method is to use existing available data. This method has been successfully employed elsewhere in the field of human error research by making use of existing accident and error reports (e.g. Rasmussen, 1980; Pew, Miller & Feeher, 1981; Woods, 1984; Wagenaar & Groeneweg, 1987). As with the other methods of data collection there are associated pros and cons (Chappell, 1994).

There are several advantages to using archive data. The first is that the data comes directly from the participants in the incident. The second is that there are large numbers of observations available (in contrast to accident data). The third is that the data has high ecological validity because it relates to real incidents which occur under normal operating conditions. In addition, the cost of collecting the data will generally be low, since the data already exists.

The main disadvantage of using archive data is that it has usually not been gathered for the specific purpose of the investigation at hand. The net effect of this is that the data has to be re-ordered and possibly re-represented before it can be appropriately analysed. The second disadvantage is that the detail in the reports may not have been validated. The third is that the data may be subject to reporter biases, in terms of who reports, and the information that gets reported which may be biased by selective retrieval and rational reconstruction (Ericsson & Simon, 1993).

3.5.4 Selecting the most appropriate data collection method

The aim of each of the methods described above is to generate data that can be used to develop theories and models of human performance. In recent years there has been a shift towards an increased use of data gathered from the real world (e.g. Hutchins, 1995), especially when studying expert performance in dynamic environments. The ideal method, however, would be one which combines the contextual richness that occurs in real world situations with some of the experimental control that laboratory conditions can provide. The work on the HEAP (Hollnagel, Drøivoldsmo, & Kirwan, 1996) comes close to this ideal. The HEAP involved observing teams of operators running a high fidelity nuclear power plant simulator. Even with the HEAP work, however, there are problems. The operators—experimental subjects—either know, or can guess the purpose of the work. As a result they have some expectation that an abnormal situation will arise during a simulator session. The sessions were also normally limited to an hour, rather than the duration of a normal shift. On a more general level, the high financial cost of building (or buying) a simulator has to taken into consideration too.

The ultimate choice of method depends on the weights associated with the pros and cons of the individual methods. Ecological validity is regarded as the most important criterion here because this research is concerned with human performance when operating complex systems in dynamic environments. For this reason, the use of laboratory experiments can be eliminated. The hit and miss nature of using field observation for relatively infrequent events, such as state misinterpretation, mitigates against its use, except perhaps in longitudinal studies.

Archive data, on the other hand, naturally has high ecological validity. Furthermore, a judicious choice of data source should make it possible to gain access to the hundreds of instances of state misinterpretation required for analysis. For these reasons, the use of archive data was selected as the most appropriate way of gathering data here.

3.5.5 Critiquing the use of archive incident data

Reason (1991) criticises the notion of recording incident data, suggesting that it can only be used to identify tokens of types of instances, rather than the types themselves. The basis of his argument is that the types reside at a point much further back in the causal chain of events, and can often be traced back to decisions and actions made at the organisational level. The effect of these corporate events lies dormant in the socio-technical system until the event is activated by some triggering conditions. Reason likens this to the existence of resident pathogens within the human body, and refers to these decisions or actions as latent failures, in contrast to active failures, which are decisions or actions that have an immediate effect on the integrity of the system.

Reason's criticisms of the use of existing incident reporting systems were directly addressed by van der Schaaf (1991b). Whilst agreeing that reactive methods, such as reporting systems, are not a panacea, van der Schaaf also pointed out that the use of more proactive methods is insufficient by itself. Van der Schaaf's (1992) Near Miss Management System (NMMS) demonstrates how it is possible to backtrack beyond incident tokens, using incident reporting systems. The key is to explicitly attempt to identify the types of causes involved in individual incidents by trying to identify common patterns of features across incidents. Such an approach is consonant with the identification of isomorphic events (Toft & Reynolds, 1994), discussed earlier in Section 3.1.

It is not always possible to prevent all errors, something that Reason (1990) readily acknowledges. In order to deal with these situations some form of error management is required (Frese & Altman, 1989). Managing errors at the sharp end of the system, however, critically depends on the ability to discriminate between different types of incidents, which in turn relies on being able to discriminate between different manifestations of errors. The process of discrimination is founded on the ability to appropriately categorise incidents on the basis of their defining features. These features, in turn, can only be identified by contrasting and comparing several examples of incidents. Incident databases provide one way of gathering together information about several incidents in one place, thereby creating a basis for identifying and defining different types of incidents, and categorising the incidents accordingly.

3.6 Establishing Selection Criteria for Archive Data

Having decided upon the use of archive data as the most appropriate method of data gathering, the next step is to select a source of data. There are a number of issues surrounding the collection and use of archive data which need to be directly addressed, however, before searching for potential data sources. The nature and interaction of these concerns directly influences the availability and suitability of individual data sets, and can be used to generate several criteria to direct the search for incident data sources, and to guide the final choice of data source.

Archive data is inevitably only as good as the system that is used to collect and record it. The existence of archive incident data relies on the presence of an incident reporting system. That system has to be available to the people who need to be able to report incidents, and they need to utilise the system, in order to record the data. The introduction of an incident reporting system generally requires a change in the culture of the environment where it is deployed. It is therefore important that the data source selected for this research has a history which takes appropriate account of the need for acceptance and usage of the system to develop.

A coherent body of incident data can only be established if there is a framework available for recording that data. In this way, it becomes easier to analyse multiple incidents for common features, since the individual incidents should be reported in a consistent manner. The framework should also include an explicit method for categorising incidents, so that the incidents can be appropriately classified when they are reported. The use of explicit categorisation also helps to avoid potential ambiguity about the terms used to classify individual incidents, during subsequent analysis of the data. It should be noted, however, that the classification scheme employed when the data is initially recorded is often designed with a particular purpose in mind. Subsequent analyses of the data will need to take appropriate account of the original classification scheme, which, in some cases may mean that the data has to reclassified using a different scheme. There is consequently a need to ensure that the original data classification scheme is available alongside the data.

The number of incident reports stored in a reporting database is a direct reflection of the usage of the (complex) system. Usage, however, can be affected by acceptance of the system, or may simply reflect problems of finding time and resources to report incidents. Ives (1991), for example, describes a case on a nuclear power plant where time pressures made it difficult to report incidents during the early phases of commissioning, which led to the recurrence of similar incidents. The management addressed the problem for the start-up of the plant by employing a number of experts for the operators to call on when they encountered a problem. The operators were thus able to continue with their normal business, whilst the experts analysed the situation at hand.

The selected data source should therefore have an incident reporting rate which indicates that the system is being actively used by multiple sources. If an industry wide system has a reporting rate of less than 100 reports per month, for example, this suggests that the system is not being very actively used.

The quality and depth of information available for any particular incident can vary widely. The net effect of this variation is that the value of the reported data correspondingly varies in terms of its usefulness as a tool for deployment in research or management. Adopting a sceptical attitude towards the utility of incident data, however, merely generates a self-fulfilling prophecy: if accident data is perceived as not being particularly useful, then there is little enthusiasm for collecting and recording it, so any data that is collected turns out to be ineffective (Hale, Karczewski, Koornneef, & Otto, 1991). The corollary of this argument is that any potential source of data should preferably have been widely used, and its use publicised and accepted by the operators within the appropriate domain.

The full benefits of incident data can only be realised if the data is analysed and utilised, rather than simply filed away. Careful consideration needs to be given to how the data is to be used, however, since it is very easy to unwittingly generate a culture which will not report any incidents. Ives (1991) reports how the management of a nuclear power plant decided to measure plant performance by the reduction in the number of incident reports. This policy directly led to the number of reported incidents falling by 50% in the first month. The conflicts that developed between various parts of the organisation eventually resulted in the company having to develop a completely new incident reporting system. Furthermore, the data on the unreported incidents was lost forever. It is therefore important that the usage of the reported incident data be appropriately considered during the selection process.

The need for making public the nature and details of accidents involving structural failure has been around for at least 140 years (Petroski, 1985). The specific need to maintain databases of accident and error reports has also been noted by Rasmussen (1980), and Brazendale (1990). There is additional impetus from the regulatory authorities too. So, for example, in the United Kingdom there are regulations on industry that require mandatory reporting of incidents, such as the Reporting of Injuries Diseases and Dangerous Occurrences Regulations (RIDDOR) and similar regulations exist for the railway and offshore industries. The Civil Aviation (Investigation of Air Accidents and Incidents) Regulations 1996 also explicitly identify the purpose of the investigation and accidents under those regulations to be the prevention of accidents and incidents, and explicitly states that the purpose is *not* to apportion blame or liability. These examples suggest that any industries which are centrally regulated should be a fruitful place to look for incident data sources.

In spite of the recognised need to make incident data available, and the various regulations, the number of incident databases that are open to public scrutiny remains comparatively small. Up until recently there was evidence to suggest a trend in some industries towards a reduction in the number of published reports (e.g. Petroski, 1985; Kletz 1993). This trend looks to have been reversed, however, with the release of two databases for the chemical process industry (one published by the Institution of Chemical Engineers in the UK, and the other published by the American Institute of Chemical Engineers), an increasing number of aviation incident reporting systems being established, and the proposed establishment of a national incident reporting system for the railways in the UK, following the accident at Ladbroke Grove in 1999 (HSE, 2000).

3.7 Possible Sources Of Archive Incident Data[2]

There is no doubt that there is a rich supply of incident data available. There is a major problem in accessing much of this data for research purposes, however. Some considerable time was spent in trying to identify sources of data which were directly accessible. Several companies in various domains were contacted, in an attempt to obtain a source of incident data. The general attitude, however, was a marked reticence to allow anyone to analyse incident data, partly for fear of possible litigation, and an unspoken fear about possible adverse effects on competitive edge. Where incident databases do exist and are accessible they tend to be industry wide, and are managed by a central organisation.

Although public enquiry reports are a potential source of data, they can be eliminated because they are too disparate and too detailed for the purposes required here. Such reports are usually only produced when a major accident occurs, and it is deemed to be in the public interest to hold an open inquiry into the circumstances surrounding that accident. The enquiry often lasts months, with the resulting reports extending to hundreds of pages, containing several sections of expert testimony about the various individual facets of the accident. These reports are generally difficult to use as a way of establishing what happened because the way in which the information is often presented makes it difficult to generate a coherent view of the true nature of the accident (Johnson, McCarthy and Wright, 1995).

Whilst there are a number of perceived drawbacks to making incident data publicly available, most organisations do collect incident data in some form. The motivations behind making incident reports generally available has helped to provide an impetus to produce databases that are available in the public domain. The remainder of this section offers an insight into the number of these databases that are available, along with other potential sources of incident data. The databases tend to be divided into two basic categories. The first is a bibliographic database which provides pointers to relevant publications. The second is a database which contains specific records of individual incidents. Where the latter type is maintained by an individual company, the data is usually not publicly accessible. There are some databases which are

---

[2] The information in this section is partly based on responses received from the Crew Services Ergonomics Information Analysis Center (CSERIAC) at Wright Paterson Air Force Base in Ohio, the Health and Safety Executive in the UK, and various individual companies and professional bodies.

maintained by a central authority, such as a regulatory or professional body, however, and in general, these are more likely to be available to the public.

3.7.1 Nuclear power

The Major Hazardous Incidents Database Service (MHIDAS) is maintained by the Safety and Reliability Directorate of the UK Atomic Energy Authority. It is distributed by Silverplatter as a CD-ROM on a subscriptions basis. MHIDAS is essentially a bibliographic database, containing references to articles on a range of incidents. It has, however, been used by researchers to analyse incidents and categorise failures.

The International Nuclear Information System (INIS) is co-ordinated by the International Atomic Energy Agency (IAEA) in Austria. INIS is an extensive bibliographic database which is concerned with peaceful applications of nuclear science and technology. Details about INIS can be found at the IAEA web site (*www.iaea.or.at*), and it is available as a CD-ROM from Silverplatter. The IAEA also maintains an incident reporting system (IRS), details of which are available at *www.iaea.or.at/worldatom/inforesource/other/iaeanea/iaeanea-irs.html*.

The Health and Safety Executive (HSE) in the UK produce a quarterly statement on incidents at nuclear installations, which is available from their information centre in London. Press releases about various nuclear incidents where the HSE has been called in to investigate are also made available on-line, at the HSE's web site (*www.hse.gov.uk*). These press releases are usually summaries of the full reports which can be purchased from the HSE.

The UK nuclear power industry have their own internal database of incidents, called the Nuclear Plant Event Reporting (NUPER) system. There is no information generally available on it, however. Scottish Nuclear also produce regular reports for public consumption on incidents that occur at their various power stations.

The Public Document Room (PDR) of the Nuclear Regulatory Authority in the USA is accessible at *www.nrc.gov/nrc/pdr/pdr1.htm*. The PDR has links to some incident reports, and allows indirect access to the PDR Bibliographic Retrieval System and the PDR Bulletin Board System.

3.7.2 Oil and chemical production

These two domains are considered together because they overlap to varying extents, most noticeably in the Petrochemical industry.

The Chemical Safety Newsbase covers information on health and safety in chemical and associated industries. It is a bibliographic database.

The Institution of Chemical Engineers in the UK produces an accident database, which was released in 1998. This database, which contains over 8000 records, includes details of some incidents that have not previously been made available. The database has been extended and upgraded to include more data since its initial release. The database is available on CD-ROM; details are available at *www.icheme.org/she/accident_db.html.*

The American Institute of Chemical Engineers has also recently developed its own database, called the Process Safety Incident Database (PSID). The PSID is only available to subscribers for a one-off payment fee. Details are available at *www.aiche.org/ccps/lldb.htm.*

The Health and Safety Executive office in Bootle in the UK maintains a database of Offshore Accidents which is available on CD-ROM from Silverplatter. The HSE also publish press releases which are available on-line, and are usually summaries of the HSE's report into incidents involving organisations from the chemical industries.

A search conducted on request by CSERIAC uncovered some potential sources of information in the United States of America. There were some doubts expressed about the availability of the data, however, since much of it is of a proprietary nature.

3.7.3 Aviation

Aviation is probably the most enlightened domain when it comes to making accident and incident data available. The net effect is that there are several sources of aviation incident reports available, most of which have details available on the Internet. It needs to be stressed, however, that the way in which the aviation system works differs across the world, which makes it difficult to directly integrate incident report data from different regions. As a result, the available databases tend to be restricted to individual geographical regions.

The Air Accident Investigation Branch (AAIB) in the UK produces a monthly bulletin of accident reports which is accessible at *www.gtnet.gov.uk/aaib/aaibhome.htm.* The bulletin contains textual reports of incidents that have been investigated by AAIB.

Also within the UK there is the Confidential Human Factors Incident Reporting Programme System (CHIRPS; Green, 1990), which was originally set up by the RAF Institute for Aviation Medicine. It is now maintained by a separate charitable trust. Although the data is not currently directly available, they may run queries on the data upon request. A quarterly bulletin, *Feedback*, which contains a summary of reports received, is available at *www.users.dircon.co.uk/~chirp*.

In the United States there are a number of incident reporting systems available. Probably the best known is the Aviation Safety Reporting System (ASRS; Reynard, Billings, Cheaney, & Hardy, 1986) which was set up by the FAA's Office of the Assistant Administrator for System Safety and is administered by NASA. The *Callback* bulletin, which includes reporting figures, is produced monthly, whilst *ASRS Directline*, which has in-depth reports on a number of reported issues, is published infrequently. Both of these bulletins are available at *asrs.arc.nasa.gov*; the ASRS CD-ROM is available separately for purchase.

Three other databases are accessible on-line through the Office of System Safety's web site. The first of these is the FAA Incident Data System (FIDS) which is accessible at *nasdac.faa.gov/asp/asy_fids.asp*. The second is the National Transportation Safety Board's Aviation Accident/Incident Data Base, which can be accessed at *nasdac.faa.gov/asp/asy_ntsb.asp*. The third is specifically for near mid air collisions (NMACs) and can be accessed at *nasdac.faa.gov/asp/asy_nmacs.asp*.

The International Civil Aviation Organisation (ICAO) have a database called the Aircraft Accident/Incident Reporting System (ADREP). This is maintained by the Accident Investigation and Prevention Section of the Air Transport Bureau of the ICAO. Information from the database is available in printed form. ICAO can be found at *www.iaco.org*.

The Joint Research Centre of the European Commission has been setting up a database to overcome compatibility problems of different individual databases. The European Co-ordination Centre for Aircraft Incident Reporting Systems (ECC-AIRS) is based on the ICAO standard for accident and data reports. The details of ECC-AIRS were originally at *www.jrc.org/isi/atia/activities/eccairs.asp* although this location no longer appears to be valid.

The European Confidential Aviation Safety Reporting Network (EUCARE) was based in Germany. After a successful introductory phase, however, internal wranglings led to the closure of the system in June 1999. Details about EUCARE are still available at *www.eucare.de*.

In Australia, the Bureau of Air Safety Investigations (BASI) maintains an incident database, called the Confidential Aviation Incident Reporting System (CAIRS). Details on CAIRS can be found at *www.basi.gov.au/cair/cair1.htm*.

New Zealand also has its own incident reporting system. Details of the Independent Confidential Aviation Reporting System (ICARUS) can be found at *www.icarus.co.nz/icarus/*.

The Transportation Safety Board of Canada has an integrated system for reporting incidents in the marine, rail and air transport domains. The details of the system, which is called SECURITAS, can be found at *www.bst-tsb.fc.ca/eng/about/securitas/securitase.html.*

In addition to the national initiatives reported above, South Africa collects incident reports via the Aviation Safety Council (SAASCo), and Russia has its own Voluntary Aviation Reporting System. Neither of these are directly available to the public.

A more general resource for safety issues in the aviation domain is the Aviation Safety Network. It includes descriptive reports of some accidents, and can be found at *aviation-safety.net.* Perhaps the most significant development over recent years, however, has been the establishment of the Global Aviation Information Network (GAIN). The basic aim of GAIN is to share aviation information, including incident reports, world wide. The membership of GAIN encompasses people from aviation regulatory authorities from across the world, as well as from individual companies. GAIN has its own web site at *www.gainweb.org.*

3.7.4 Shipping

The US Coast Guard keeps data on shipping accidents in US waters. Lloyds of London maintains a database for world wide operations. Arthur Mackenzie, director of Tanker Advisory Center, Inc, collects some of this data.

The Marine Accident Investigation Branch in the UK maintains a database of shipping incidents. Details are available on-line at *www.open.gov.uk/maib/maibhome.htm.*

3.7.5 Medicine

In healthcare the Adverse Incident Centre of the Medical Device Agency in the UK keeps a record of adverse incident reports. In 1995 over 4000 incidents were reported, covering all aspects of the delivery of health services. The reports range from technical aspects such as the problems of a syringe pump providing excessive amounts of medication, to seemingly more mundane matters, such as drawers of supplies which are jammed closed and cannot be opened. Although superficially these two incidents would appear to have widely differing effects on patient safety, both turned out to be potentially life threatening. In the latter case, the drawer in question was used to store life saving drugs and equipment on an emergency trolley.

The Department of Anaesthesia at the University of Basel in Switzerland maintain the Critical Incidents Reporting System for Anaesthesia Incidents (CIRS). CIRS is accessible at *www.medana.unibas.ch/cirs/default.asp.*

3.7.6 Miscellaneous

Contact was also made with a number of individual companies in the Chemical, Nuclear and Water industries. Each of them reported that they maintained an incident reporting system, but none of them make the details of the reports available for general public consumption. This feedback confirmed that CSERIAC's concerns about the accessibility of the data, due its proprietary nature, are not confined to the USA.

In addition to the various sources listed here, CSERIAC also noted that the Defence Technical Information Center in the USA produced a two volume document of summary information in 1989, entitled *Accident/incident data analysis database summaries* (Murphy & Levendoski, 1989a; 1989b). These are available for purchase through the National Technical Information Service in the USA.

3.8 Selecting The Data Source

The selection process was conducted using the following criteria in order of importance:

1. Availability—the database needs to be currently available.
2. Accessibility—the data recorded in the database needs to be publicly accessible.
3. Size—given that state misinterpretation is a comparatively rare event, the database needs to be large enough to yield of the order of 100 instances of state misinterpretation.
4. History—the database should ideally have a history which will be reflected in the quality (and quantity) of the data.

The three main candidates were the Accident Database from the Institute of Chemical Engineers, the database for the process industries which is produced by Silverplatter, and the NASA Aviation Safety Reporting System (ASRS) database. The IChemE's Accident Database was eventually rejected because it was not available in the required time frame, after its initial release was postponed in 1997.

The Silverplatter database was evaluated and rejected because it is a bibliographic database. In other words, it provides references to sources of information about incidents, rather than details about the incidents themselves. Therefore, by a process of elimination, the selected database was the NASA ASRS database (Reynard, et al., 1986). The ASRS database also proved to satisfy all of the selection criteria as follows:

1. The ASRS database was already in use, being updated quarterly, and available on a CD-ROM.
2. The ASRS database can be purchased directly by the public, and has been used in several studies in aviation research.
3. The ASRS database at the time of the selection was receiving over 2,000 new reports each month, and contained over 50,000 incident reports.
4. The ASRS database has been in existence in its current form for over 10 years.

The advantages and disadvantages of using incident data, which were noted earlier, all apply to the ASRS database.

3.9 Summary

In this chapter the collection of human error data has been discussed, with particular reference to state misinterpretation by human operators in complex systems. After due consideration of the particular requirements of this research, it was decided to use archive incident data. A set of selection criteria was then established to determine what sort of archive data to use. The chosen data source to be used for investigating state misinterpretation was the Aviation Safety Reporting System which holds incident reports from the aviation domain. Having delineated state misinterpretation, in the next chapter the working definition is translated into a set of queries that can be used to retrieve a set of incident reports in which state misinterpretation is implicated from the ASRS database.

4 Selecting State Misinterpretation Incident Reports

The definition of state misinterpretation that was developed in Chapter 2 is exploited in this chapter in the process of selecting an appropriate set of incident reports—those in which state misinterpretation is implicated—from the ASRS database. This definition is translated into a format that can be exploited in an initial search of the ASRS database. The implementation of the ASRS database imposes some constraints on the way that queries can be formulated, and the extraction of data, however. The problems raised by these limitations and the solutions that were used to overcome them in the initial search of the database are described. The initial search is then refined in light of the overly general nature of the information retrieved from the database, adopting a heuristic for identifying expert pilots and air traffic controllers. The particular evolutionary nature of the aviation domain is then taken into account, and the search is refined again in order to select a more homogeneous set of reports for further processing. Finally, the selected set of reports are retrieved and manually processed to remove false positive reports—reports which are retrieved because they syntactically match the database queries, even though they do not relate to incidents in which state misinterpretation is implicated.

4.1 Using The ASRS Database

The CD-ROM (version 97-2) of the ASRS database operates in two basic modes: search mode and display mode, and it is possible to readily switch between the two. As the names imply, search mode is used to query the database, and display mode is used to display the contents of the reports identified by a search.

In search mode, the ASRS database utilises a forms based interface. Interrogating the database is simply matter of filling in the appropriate fields on the form—either manually or by reloading a query which has been previously saved–with the details on which a match is to be performed. So, for example, to find all the reports for incidents which happened on a Monday, the term "Monday" would be entered in the *Day Of The Week* field on the search form. The query language is simple, although it does provide wildcards, and logical operators (*and*, *or*, and *not*) for use in text matching, and provides proximity operators (*adj*, *same*, *sames*, *near*, and *nears*) which can be used to search for terms that are located in the proximity of other terms. Guidance on how to formulate queries is provided in the ASRS CD ROM User's Manual (Aeroknowledge, 1994). Once a query has been formulated, it can be saved for future use.

The contents of the reports identified by a search can be exported out of the ASRS database for subsequent processing. The data in the reports can be exported in several formats, including text, which allows the reports to be imported and processed by other software applications. In addition, each of the fields in the reports can be selectively excluded from the exported files.

4.2 Search 1: Incidents reported for the different types of perceived state

The goal of the first search is to identify two sets of reports: one for pilot behaviour, and one for air traffic controller behaviour. Each of these sets of reports is divided into eight subsets: one for each of the different types of perceived state. In other words, there will be sixteen sets of reports, which means that sixteen database queries are required, one for each set of reports. The search fields used in the database queries are the FAA Category, and the Narrative Section.

The incident reports stored in the ASRS database relate to a multitude of facets of the aviation domain, including flying commercial aircraft, and air traffic management. Each report in the database has several fields, which are used to describe the circumstances in which the incident occurred. When the reports are processed prior to being added to the database, they are categorised using the scheme defined in Chapter 5 of the FAA Facility Operation and Administration Handbook (FAA, 1998). Although there are several different categories of incident, the only ones that are directly relevant here are *Pilot Deviation*, *Operational Error*, and *Operational Deviation*; the FAA definitions of the categories are given in Appendix A. Incidents that are attributable to pilot behaviour will normally be classified as Pilot Deviations, whilst those incidents attributable to air traffic management will be classified as Operational Error, or Operational Deviation. These categories are not mutually exclusive, so any particular incident can belong to several of the FAA defined categories.

The FAA Category field is directly used to divide the reports into the two top level sets. Since the FAA Categories of incident types are not mutually exclusive, this field is checked to make sure that the report only belongs to one of the domains—piloting an aircraft, or air traffic management—by only selecting reports which are categorised as Pilot Deviation on the one hand, or as Operational Error or Operational Deviation on the other. In other words, a logical exclusive or condition is used.

The key to the initial search, however, lies in working out how to identify the different types of perceived state from the textual narrative description of the incidents. A simple search of the narrative fields is insufficient, because, as noted in Chapter 2, the perceived state is the result of an interaction between three dimensions. Consequently, there is no simple term that can be associated with each of the perceived states. Instead, the perceived state has to be analysed into its individual dimensions, and a search formulated using terms which can be used to describe each dimension individually.

4.2.1 Describing perceived state types using the ASRS database

The determination of the number of reported incidents belonging to each of the eight different types of perceived state of the system relies on being able to classify each incident as belonging to one of those types. There is no section of the report which directly describes the operator's perceived state, however, so the classification has to be inferred from the narrative description of the incident. In other words, the classification of the reports is determined by inspecting the narrative section of the reports for a textual description which fits into one of the categories of perceived state identified in Chapter 2.

An example of a textual narrative from an ASRS incident report is shown in Table 4.1 with the Accession Number (the incident report number) appended to the end of the text. There are several important points to note about the content and structure of the narrative sections in the database, which have a potential impact on the way that the text is extracted and subsequently analysed:

- All of the text appears in upper case.
- The text is free form.
- Each line of the narrative is terminated by a paragraph marker.
- An informal abbreviated form is used in several places to substitute for some words, such as *apch* for *approach.*
- The narrative includes the use of standard aviation abbreviations, such as ILS for instrument landing system. (An extensive glossary of commonly used aviation

Table 4.1. Textual narrative section for ASRS report 363880.
(The report is shown verbatim.)

WE WERE BEING VECTORED BY CHICAGO APCH FOR AN ILS IN VMC FOR RWY 4 AT MDW. FO WAS FLYING. WE WERE AT 4000 FT ON 280 DEG HDG. ATTEMPTING TO STAY AHEAD I SET MDW TWR FREQ IN THE RADIO THAT HAD APCH FREQ SET ON IT. I HEARD A XMISSION FROM AN ACFT ON THE GND AT MDW AND REALIZED THAT WE WERE ON THE WRONG FREQ. THINKING THAT I HAD ACCIDENTALLY HIT THE SWITCH TO THE OTHER RADIO I SWITCHED IT BACK. THIS PUT US ON ZAU WHO WE HAD BEEN TALKING TO BEFORE APCH. WE REALIZED MY MISTAKE AS WE FLEW THROUGH THE RWY 4 LOC AND SAW OTHER TFC TURNING OUT OF OUR WAY. THE FO REMEMBERED THE APCH FREQ, WHICH I QUICKLY TUNED IN AND CHKED ON. THE CTLR WELCOMED US BACK AND TURNED US FOR THE APCH TO MDW. THE FO AND I HAD BOTH WOKE UP EARLY THAT MORNING. BY THIS TIME I WAS HAVING A GREAT DEAL OF TROUBLE STAYING AWAKE. IT WAS NOT A REALLY HARD DAY (MKE-RSW-MDW) BUT JUST ONE OF THOSE DAYS WHEN YOU'RE JUST NOT COMPLETELY UP TO SPD. IN SITS LIKE THIS IT WOULD SEEM AS IF THE SMALL MISTAKES ARE THE ONES THAT SLIP BY UNNOTICED. (Accession Number 363880) terms and abbreviations can be found in Bruford, 1994.)

The classification of the incidents by the type of the operator's perceived state of the system is founded on the use of a textual description of each of the three previously identified dimensions of that state (knowledge, access, and interpretation). In particular, a set of key phrases which can be associated with a *failure* in these dimensions is required, because the narrative sections of the reports describe the incident in terms of what happened. In other words, the reports focus on what went wrong, so a failure in any of the three dimensions should be reflected in the narrative section of the incident report. A failure to access all of the defining state variables, for example, might be described in the narrative section by the phrases "did not check", "failed to check" or "did not read", followed by the item(s) which had not been checked. In order to ensure that all the alternative descriptions are found during a search of the database, what is ideally required is a complete set of synonymous textual phrases that could be used to describe a failure in each of the dimensions.

The number of incidents reported for each of the types of perceived state is determined by searching the database using a separate query for each of the types of perceived state, for each of the two domains. Each query has to include a textual description which encapsulates the relevant combination of the three dimensions which determine the perceived state. In other words, the description must take account of the terms that the incident reporters use in the narrative section to describe whether or not they successfully exploit each of the dimensions during state interpretation. The textual descriptions for each of the individual dimensions are therefore all phrased in the negative, as noted above. The corollary of this is that the negation of the textual descriptions of the individual dimensions corresponds to the successful exploitation of that dimension.

The first step in establishing the queries to the database is to identify the sets of textual phrases which could be associated with the breakdowns in each of the dimensions of the perceived state. The narrative sections of the incident reports are written in the reporter's own words. As noted above, this means that a set of synonymous phrases is needed for each of the dimensions. Separate phrases were generated using knowledge about the individual dimensions, and how they could be described. Each of the phrases was used to query the ASRS database to determine the number of reports which had a narrative section containing that particular phrase. Any phrases which returned zero reports matching the query were eliminated from further consideration. In this way three sets of phrases were constructed, one for each of the different dimensions. The full sets of phrases are shown in Table 4.2. Each of the columns represents one of the dimensions, and closely related phrases are shown grouped together. It is worth emphasising at this juncture that the descriptions are domain independent, and would apply equally well to any domain involving interaction between humans and complex systems.

The length of each of the fields on the ASRS search form is fixed, which means that the overall length of queries used to retrieve selected reports is correspondingly limited. In general, these restrictions are unlikely to be a problem, particularly when the possible range of values for the fields included in the search query occupies a small, fixed set. When searching the textual narrative field, however, the restrictions start to have a significant impact on the structure of the query. The main problem, noted earlier, is that the narrative section is comprised of free form text, so different reporters may use different terms to describe the same situation or sequence of

Table 4.2. Text phrases used to characterise failures in the dimensions of state interpretation.

State variable identities	State variable numbers	State variable values
((did not) OR (didn't)) know	incomplete	(misread OR mis-read OR (mis-read)) OR (misinterpret OR mis-interpret OR (mis-interpret)) or (misheard OR mis-heard OR (mis- heard)
((was unaware) OR (((was not) OR wasn't) aware)	(failed OR failure) to complete	misunderstood
(lack of experience) OR (lacking experience) OR (lacking in experience) OR (inexperienced) OR (inexperience)	((did not) OR didn't) complete	((did not) OR didn't) clearly
(never OR not) (previously OR before) (encountered OR experienced OR seen OR noticed OR observed)	"not" (fully OR totally OR entirely) complete(d)	incorrect reading
without (knowing OR knowledge)	not completely	incorrectly heard
(rare OR rarely) OR (infrequent OR infrequently)	((did not) OR didn't) (see OR hear OR notice OR note OR heed OR observe)	
cannot (remember OR recall)	(failed OR failure) to (see OR hear OR notice OR note OR observe)	
((was not) or wasn't) familiar	(omitted OR forgot OR neglected) to	
unfamiliar	inadequate	
	(attention OR attentive) OR (inattention OR inattentive)	
	insufficient OR (not enough) OR (lack)	
	omission	

events. When this problem is coupled with the need to search the text field across three different dimensions (knowledge, access and interpretation), each of which has its own set of associated textual phrase, the size of the search query increases correspondingly, because the set of text phrases (or their negation) for each of the dimensions have to be included in the query.

The search field on the query form for the textual narrative section of the report can only hold a maximum of 400 characters. The ASRS interface supports the use of wildcards in searches, but even after the desired queries for each of the types of perceived state were hand optimised to maximise the use of wildcards, they were still too long for the search field on the form.

Rather than simply trying to work out the combination of phrases which returned the largest number of reports by trial and error—pruning the query by hand and re-running it—it was decided to utilise the export facility of the database to effectively generate the numbers of reported incidents in each of the different categories of perceived state (including the two categories corresponding to state misinterpretation identified in Chapter 2). The number of reported incidents in each category could then be used to check the qualitative prediction (made in Chapter 2) that the expected number of reported incidents for each of the categories where the operator does not know all of the state defining variables should be low. The numbers of reported incidents would also give an indication of the relative frequency of state misinterpretation.

Having established a method for identifying the perceived state from the narrative sections of the incident reports the next step is to incorporate this into the search process. This method is extended below to determine the number of reported incidents that can be allocated to each of the categories of perceived state for pilots and for air traffic controllers.

4.2.2 Formulating the search

The database is interrogated using one set of queries for each of the two domains of interest, with a separate query for each dimension of the perceived state and its negation. In total, twelve queries are generated, six for each domain. Each of the queries is hand optimised, and the results of the queries are saved in external files using the ASRS database export facility, as plain text. The contents of the fields which appear in each of the reports can be selectively excluded from the exported file. In this instance, only the accession number field is selected for inclusion.

Strictly speaking, the negation of the queries is really only the negation of the search terms used for the narrative section of the report—which correspond to the situation where the operator exploited the appropriate dimension. The negation of the queries can therefore be considered to describe the situations where there was *not* a failure in that dimension. There is a fundamental assumption here that the omission of a reference to a failure to successfully exploit a particular dimension is equivalent to that dimension being successfully exploited; in Boolean logic terms it is the negation of a negation which is the same as the original positive value. This assumption is a simplification, and it does have the limitation that incidents where the operator did notice the breakdown (as opposed to the breakdown not happening) are treated in the same manner as those incidents where the breakdown did not happen. On the other hand, however, the narrative sections of the reports *do* describe situations in which something went wrong. Also, the reports describe incidents, rather than accidents, so, in general, the operators noticed what went wrong and corrected it. In other words, as a rule, the operators are observant, and *do* notice when things go wrong, even though it may require some retrospective analysis to discover what actually happened.

The CD-ROM version of the ASRS database runs on an IBM compatible personal computer under the MS-DOS operating system. The MS-DOS limitations on the naming of files therefore apply, such that any data files exported from the database can only have a file name with a maximum length of eight characters, and a file extension (or file type) with a maximum length of three characters. To make it easier to keep track of the exported files, each filename is given a two letter prefix to identify the type of data held in the file: *PD* for pilot deviations; *OE* for operator (ATC) errors; and *OD* for operator (ATC) deviations. The remaining six characters identify the factor to which the query or its negation related. The files are always exported as text, so a file extension of "txt" is used.

The process of formulating the search within the ASRS for each dimension of the perceived state is summarised in Table 4.3. Once all of the queries (and their negations) have been run, there are six files of accession numbers corresponding to the binary dimensions (true and false) for each of the individual dimensions of state interpretation. The accession numbers in the

Table 4.3. Summary of the steps required to formulate a database search for the accession numbers of the reports for each of the dimensions of the perceived state and their negations.

1. *Generate ASRS database query for the next dimension of the perceived state (i.e. knowledge of the state defining variables; accessing those variables; and interpreting their values correctly).*
2. *Run the query.*
3. *Export the accession numbers of the retrieved reports.*
4. *Run the query using the negation of the text part of the query.*
5. *Export the accession numbers of the retrieved reports.*

exported files are held in ascending order by default. The next step involves combining the results of the queries by performing a matching process across the individual dimensions outside of the ASRS database.

4.2.3 Combining query results outside the ASRS database

A software application was implemented using the Inprise Delphi software development tool (Version 3.0) to perform the matching across the three input files. This application generates a new file for each of the eight possible combinations of dimensions corresponding to the eight types of perceived state. The matching process is performed by sequentially comparing the contents of the three files. Accession numbers that appear in all three files are written to the output file. The output file is thus formed from the logical conjunction of the three input files.

The matching process is performed on the data files that are exported from the ASRS database, which only contain the accession number field. The exported files could have been set up in such a way as to include the narrative section, which superficially would appear to make sense, since it is the narratives that are to be subsequently analysed. On closer inspection, however, it becomes apparent that doing so merely adds an unnecessary overhead at this stage. The matching process is character based, because each of the accession numbers is exported as a text string. In this way, the process is kept fairly simple. If the narrative sections were included in the exported files, the matching process would also have to read all of the narrative sections in all of the files, discarding the contents of those which did not match. This would simply make the implementation more complex (albeit to a small extent) but would significantly increase the time taken to perform the matching process, because of all of the extra read operations that would be required for each of the input files.

The output files produced by the matching process are named using a predefined convention. This naming convention was designed to make it possible to determine the general nature of the file contents without having to inspect the file's contents or refer to the queries used to generate it. Each file has the same prefix as noted above (*PD*, *OD*, or *OE*). The remaining six characters of the file name are divided into three pairs. The first pair of characters is set to "Id" when the query is searching for a failure in the knowledge of the state defining variables; the second pair is set to "No" when the query is searching for a failure to access all of the state variables; the third pair is set to "Vl" when the query is searching for a failure to interpret all the individual values correctly. For the successful exploitation of a dimension of the perceived state—a negation of the original query—the corresponding pair of characters in the file name is set to "xx". So, for example, a file with a name of *PDxxIdVl* contains the accession numbers of those reports classified as pilot deviations for the case where the state defining variables are known—

remembering that the standard form of the queries are searching for a failure along a particular dimension—but all of these variables are not accessed, and the values are not interpreted correctly.

There is some potential for confusion in the way that the files are named, in that the use of the letters *Id*, *No*, and *Vl* are used to represent a failure along a dimension. It would seem more appropriate to use *xx* to represent the failure. The main reason for the difficulty stems from the fact that the research involves actively seeking out failures for subsequent processing. In other words every instance of a failure is regarded in a positive light, because it means that it is *true* that there has been a failure in one or more dimensions of the perceived state.

4.2.4 Results and discussion

The results of the search are shown in Table 4.4. Row one shows that almost half of the reported incidents occurred when the perceived state was optimally correct. This simply shows that half of the reported incidents cannot be attributed to the operator's perceived state being incorrect, or to failures in state interpretation. The inherently biased nature of voluntarily reported incident data makes it difficult to quantify the incident rate directly attributable to state misinterpretation, because there is no way of knowing exactly how many incidents *actually* occurred. At worst, however, the figures in Table 4.4, show the minimum number of incidents for each type of perceived state that occurred.

The most striking result is that more than one third of the reported incidents arise when the perceived state does not include all of the values of those variables which define the system state (row three in Table 4.4). It was suggested earlier that this type of perceived state can appear to be correct, when it is really only coincidentally correct. If operators only ever encounter conditions in which a subset of the state defining variables can be used to determine the real state of the system, they will learn to rely on that subset of variables, adopting what Rasmussen (1986) refers to as the *line of least effort*. Rather than using an isomorphic mental model of the system when controlling the system, the operators use a homomorphic mental model of the system in which only a subset of the real state is represented (Ashby, 1956; Moray, 1987). In situations where the unread variables play a critical role in distinguishing a normal system state from an abnormal one, the net effect of the reliance on an incomplete mental model will be that incidents (or even accidents) will be more likely to occur.

Table 4.4. Reported Incidents for Each Type of Perceived State.

	Perceived System State Factors			Pilots	ATC
	Knowledge	Access	Interpretation		Operators
1	True	True	True	16,531 (47.8%)	2,719 (49.2%)
2	True	True	False	1,423 (4.11%)	121(2.19%)
3	True	False	True	12,119 (35.0%)	2,173 (39.3%)
4	True	False	False	1,053 (3.05%)	156 (2.82%)
5	False	True	True	1,392 (4.03%)	133 (2.41%)
6	False	True	False	206 (0.60%)	5 (0.09%)
7	False	False	True	1,609 (4.66%)	200 (3.62%)
8	False	False	False	223 (0.65%)	15 (0.27%)
			Reported Incidents	34,556 (100 %)	5,522 (100 %)

The figures in row three of Table 4.4 show that the use of homomorphic mental models was detected in up to 35.0 % of reported incidents attributed to pilots, and 39.3 % of those attributed to air traffic controllers. These percentage values are the upper limits for reported incidents, according to the definition of state misinterpretation that was proposed in Chapter 2. The real figures can only be determined by a more detailed analysis of the incident reports to identify whether each incident occurred because an abnormal system state was wrongly identified as a normal one, or there was some other reason.

Based on the relationship between state misinterpretation and the perceived state that was described in Chapter 2, the percentage of possible instances of state misinterpretation is 38.05% for pilots, and 42.12% for air traffic controllers. These values are the upper limits for reported incidents, for the reasons noted above.

One other important result is that the number of incidents reported where the state defining variables are unknown are all low, as predicted in Chapter 2. The figures range between 0.60% and 4.66% for pilots, and between 0.09% and 3.62% for air traffic controllers. These figures lend some credence to the appropriateness of the dimensions which have been used to divide the perceived state into eight different types.

The seemingly paradoxical figures in line one of Table 4.4 require further explanation. Superficially they appear to suggest that almost half the reported incidents occurred when the operator's perceived state was completely correct. What needs to be remembered, however, is that state misinterpretation only arises during the first stage of the control task (as shown in

Figure 2.2 in Chapter 2). In other words, even though the operator may have interpreted the system state correctly, there are several ensuing stages where there are opportunities for things to go wrong which could give rise to an incident.

The results are based on a coarse search of the ASRS database, and naturally include a number of false positive reports—incidents satisfying the search criteria, but not involving state misinterpretation. These reports will be removed during a detailed analysis of the incidents. The problem of identifying false negative reports—incidents involving state misinterpretation which do not satisfy the search criteria—is more difficult to resolve, however, without explicitly reading the narrative section of all of the reports for pilots and air traffic controllers.

As noted earlier, the ASRS database covers many aspects of the aviation domain. In effect, the ASRS database can be considered to be composed of several different homogeneous populations of pilots (general aviation, military, commercial air carriers and so on), and air traffic controllers (ground, departure, tower, and so on) amongst others. The results of the initial search show the overall picture for pilots and air traffic controllers, but do not take into account the level of expertise of the operators, whereas the focus of this research is on expert performance.

In order to investigate the problem of state misinterpretation in expert performance, the heterogeneous data of the ASRS database needs to be appropriately partitioned into homogeneous sets and re-interrogated using the same search criteria. A second database search was therefore conducted, using the same method for generating the queries, but with extra constraints added relating to the level of expertise of the operators.

4.3 Search 2: Incidents reported for expert pilots and controllers

The scope of the second search of the ASRS database is restricted to just one homogeneous group of pilots: experienced commercial air carrier pilots, with over 1000 hours of flying experience, and flying passenger missions. This group was singled out on the following basis:

- 1000 hours is often used as a heuristic for determining an expert pilot, especially when considering that this figure is on top of the amount of training that pilots have to undergo.

- Commercial air carriers are all subject to the same set of regulations (Federal Aviation Regulations Part 121).

- Passenger missions generally have the same structure to the crew (flight crew and cabin crew), and the crew will all be trained to work together to attain the same goals; these goals will be different for freight missions, and air ambulances, for example.

Although the ASRS reporting forms include fields for reporting the level of experience of the people involved, there is no mandatory requirement to complete all of the fields. The application of search criteria to a particular field, such as experience, can therefore result in the reports for several incidents being omitted from the results, simply because the level of experience was not reported, rather than because the reporter did not have the requisite level of experience. This limitation needs be borne in mind when trying to interpret the results of any search of the database in which constraints are placed on any of the fields which are not required to be completed.

The criteria for determining an ATC expert is slightly less clear. The ASRS database logs air traffic controller experience in terms of years, rather than hours. In addition, the controller experience figures recorded in the ASRS may be shown under three different categories: general, military, and radar. If the 1000 hour heuristic for expertise is adopted, then assuming an effective working year of 44 weeks (allowing eight weeks for vacations and public holidays) and an effective 30 hour working week (allowing for a maximum of 15 minutes break every hour on average in order to maintain effectiveness), this means that one year's controller experience should be sufficient. It is not clear whether the ASRS figures include on-the-job training time, and the fact that controllers do not initially work alone. There is no simple way to take these factors into account, since the amount of time spent under supervision will vary between individuals. A pessimistic figure of 50% efficiency is therefore assumed for the first two years, so allowing for overheads of training and so forth, a figure of three years is taken as a guideline to expertise as a controller.

4.3.1 Results and discussion

The results of this search are shown in Table 4.5. The overall shape of the distribution of reports is very similar to that for all types of pilots, and all types of air traffic controllers previously shown in Table 4.4. There is no simple way to compare the figures for the two populations, since there is not any way of equating one hour's flying experience to a number of hours experience in ATC. This fact has to be borne in mind when trying to interpret the increase in the ratio of pilot reports to air traffic controller reports from approximately 6.5:1 in Table 4.4 to just over 8:1 in Table 4.5. So, for example, the change could be the result of an imbalance between the two types of experience, or it could be due to the fact that experienced pilots are more inclined to report incidents than experienced controllers. The key point is that it is not possible to discriminate between the various possible explanations just by looking at the numbers.

Rows three and four of the table show that the size of the problem of state misinterpretation for experts is still potentially a large one. The combinations of factors in these two rows comprise 40.3% of the reported incidents for expert pilots (compared with 38.1% for all pilots, as reported in Search 1), and 43.6% of the reported incidents for expert air traffic controllers (compared with 42.1% for all controllers).

The figures for pilots shown in Table 4.5 offer a clearer view of the true state of affairs than those in Table 4.4, because they focus on a more homogeneous population. Interpreting the situation regarding the figures for air traffic controllers is less straightforward, however. Within ATC there are several different types of controller (ground controller, en route controller, and so on). Each of these different positions has its own particular set of tasks, and requires a corresponding set of position dependent skills. Although it is possible to delineate the figures for air traffic controllers using the information supplied in the ASRS incident reports, the total number of reported incidents for all air traffic controllers is quite low (1,277). There are more than twelve identifiable different types of controller roles in the ASRS database, which suggests that the number of reports for each type of controller will be close to the minimum threshold limit of 100 reports established earlier.

The reporting period utilised in Table 4.5 covers almost ten years (1988-1997), so any

Table 4.5. Reported Incidents for Each Type of Perceived State for Expert Pilots and Air Traffic Controllers.

	Perceived System State Factors			Pilots	ATC
	Knowledge	Access	Interpretation		Operators
1	True	True	True	4,908 (47.6%)	607 (47.5%)
2	True	True	False	468 (4.54%)	26 (2.04%)
3	True	False	True	3,782 (36.7%)	520 (40.7%)
4	True	False	False	370 (3.59%)	37 (2.90%)
5	False	True	True	299 (2.90%)	29 (2.27%)
6	False	True	False	43 (0.42%)	1 (0.08%)
7	False	False	True	382 (3.70%)	53 (4.15%)
8	False	False	False	63 (0.61%)	4 (0.31%)
			Reported Incidents	10,315 (100 %)	1,277 (100 %)

analysis of the reported numbers is still relatively coarse grained. Over such an extended period, the regulations governing the aviation industry are liable to change somewhat, especially given that the aviation domain is a highly regulated one. Indeed, some of the changes have resulted from trends that have been identified by the FAA in the analysis of incidents reported in the ASRS database. Furthermore, there have been some particularly significant changes in aircraft in the last decade. The increased use of computers leading to the emergence of the glass cockpit, in particular, has changed the way that pilots perform the task of flying the aircraft. The mandating of particular items of safety equipment, such as the Traffic Alert and Collision Avoidance System (TCAS), has also had a significant effect.

To address the combined effects of changes in equipment and regulations over time another search of the ASRS database was conducted, using the same method for generating the queries as above. The additional constraints of a limited time frame were included as a way of attempting to ensure a greater uniformity in the tasks performed by the operators, because they should be constrained by the same regulations and having to use the same sort of equipment to perform these tasks. Given the comparatively low numbers of reported incidents for air traffic controllers over the period 1988-1997, and the fact that there are so many different types of controllers, it was decided not to further analyse the controller data because the number of reported incidents for each category fell some way below the 100 report threshold.

4.4 Search 3: Incidents reported for expert pilots between January 1996 and April 1997

In order to make sure that the selected reports all related to pilots flying similarly equipped aircraft, it was decided to only select incidents that were reported between January 1996 and April 1997. This time frame was chosen on the grounds that all pilots would have had a chance to become familiar with TCAS, since TCAS II became a mandatory requirement in aircraft having over 30 seats in 1994, whilst TCAS I became mandatory for aircraft with between 10 and 30 seats in February 1995.

4.4.1 Results and discussion

The numbers of reported incidents for the selected period are shown in Table 4.6. The most striking aspect of the results is the remarkable similarity between the percentages for each of the different types of perceived state and the corresponding figures for pilots in both Table 4.5 and Table 4.4. Of particular interest here are the figures in rows three and four, which show that

Table 4.6. Reported Incidents for Each Type of Perceived State for Expert Pilots in the Period January 1996 - April 1997.

	Perceived System State Factors			Pilots
	Knowledge	Access	Interpretation	
1	True	True	True	732 (47.4%)
2	True	True	False	77 (4.99%)
3	True	False	True	584 (37.8%)
4	True	False	False	59 (3.82%)
5	False	True	True	28 (1.81%)
6	False	True	False	7 (0.45%)
7	False	False	True	49 (3.17%)
8	False	False	False	8 (0.52%)
			Reported Incidents	1,544 (100%)

there were 643 reported incidents—which includes possible false positive reports—where state misinterpretation was potentially implicated. In total these reports comprise some 41.6% of the total number of reported incidents for expert pilots in the period between January 1996 and April 1997. These reported incidents will provide the initial data set that will be subjected to a more detailed causal analysis.

4.5 Extracting The Reports From The ASRS Database

Having identified a set of incident reports, the next step is to extract the reports from the database into a format that supports a more detailed level of analysis. The ASRS database provides a facility which allows reports to be exported to files in a number of different formats, including plain text. The resultant files can then be imported by other applications for further processing and manipulation.

The accession numbers of the selected incident reports have been generated outside of the ASRS database, however, and need to be used to select the reports for extraction from the database. One way of doing this is to feed the accession numbers back into the database using a query which simply selects the required reports directly using the accession numbers. The problem with this approach is that the maximum number of six digit accession numbers that will fit into the accession number field on the database query form is just over 20—the field is 200 characters long, but the report numbers have to be separated by a comma and a space. If the number of reports to be retrieved is relatively small, this limitation is not too much of an inconvenience, but for larger numbers of reports it means having to manually generate several queries, which can be time consuming. After the reports have been identified and retrieved, the resultant exported files have to be concatenated together at the end to generate a unified coherent data set for subsequent analysis. Fortunately, the size of the problem is ameliorated somewhat by the fact that the only reports which need to be retrieved in full are those for the perceived states reported in lines three and four of Table 4.6.

The simplest way to retrieve the identified reports, though, is to exploit one of the features of the matching process. The matching process performs a conjunction of the contents of three files, in such a way that a report will only appear in the output file if it present in all three of the input files. As a result, the size of the output file, in terms of the number of reports it contains can only, at most, be the same size as the input file which contains the least number of reports. So, the task of retrieving the identified reports is made easiest by determining which of the three input files—which of the dimensions—contains the smallest number of reports for each of the two required types of perceived state. The query associated with that dimension is then re-run, with the full contents of the reports being exported to file. The full contents are exported, because there may be data in some of the fields other than the narrative which can be used in later analyses to help partition the data using different criteria, such as the aircraft type, or the phase of the flight when the incident occurred.

The resulting file then simply has to pruned by hand, removing any reports which do not appear in the list of report numbers for the appropriate type of perceived state. This pruning can be performed in conjunction with the manual removal of false positive reports.

The removal of false positives from the initial set of reports requires that the narrative section of each report be manually scrutinised. Removing false positives is a relatively straightforward, if potentially time consuming process. The length of time required reflects the quality of fit between the search terms used to identify the concept of interest and the way in which that concept is described by the incident reporters. The bigger problem, however, is that there is no way of easily identifying false negatives without reading through all the possible reports for the selected population of expert pilots over the chosen time period. In general it is simply accepted as a fact that there may be some false negatives which have been omitted from the retrieval, and this is explicitly noted when interpreting the results of any data analysis.

A total of 99 false positive reports were removed out of 643 reports. These false positives were categorised on the basis of why they were removed; 24 different categories were identified. The main source of false positives was TCAS related reports, of which there were 18. These reports describe incidents in which the reporter was flying the route assigned by ATC when TCAS detected a conflict—usually a loss of horizontal or vertical separation—with another aircraft. In TCAS related reports there is usually little or no information about which aircraft initially breached the separation limits.

None of the other categories had a frequency greater than 10. It is worth noting, however, that only 9 of the false positives were removed because they were not instances of state misinterpretation. Most of the other reports that were removed were for reasons that were nothing to do with pilot behaviour, such as maintenance related issues, in-flight encounters with bad weather which the flight crew could do nothing about, and even in one instance because the plane was on fire!

In addition to removing false positive reports, any reports which lacking in causal information, or in which the causal information was unclear were also removed. These reports were excluded on the grounds that no further useful analysis could be conducted to provide details on what happened. A total of 51 such reports were removed, leaving a final set of 493 reports for detailed analysis.

4.6 Summary

This chapter described how the ASRS database was used to identify and extract a set of reports about incidents in which state misinterpretation was implicated. The first stage was to establish a set of terms associated with each of the dimensions of the perceived state (which is inherently linked to state interpretation and hence state misinterpretation). These terms formed the basis of a set of queries to the database, which were used to retrieve the accession numbers of those reports in which the pilots and the air traffic controllers had either successfully exploited, or failed to exploit the dimensions of the perceived state. For each group of personnel (pilots and air traffic controllers) the accession numbers were then cross-matched to produce a set of accession numbers for the reports that belonged to each of the eight different types of perceived state. The search was then refined to make the two sets of accession numbers more homogeneous, by restricting the search to those reports in which the reporter could be defined as an expert. The final refinement of the search was to concentrate solely on pilot data reported over a limited time period in order to minimise the effects of changes in aviation regulations and aircraft technology. The reports identified by this search were exported from the database, and manually processed to remove false positive reports, and reports which were lacking in causal information. The next chapter describes the causal analysis which is carried out on the remaining set of reports.

5 An Overview of The Cognitive Reliability and Error Analysis Method

In this chapter the analysis of the set of ASRS incident reports that were selected in Chapter 4 is described in detail. The chapter opens with a discussion of the existing methods and frameworks that can be deployed in the analysis of incident data. The shortcomings of some of the existing methods are addressed through the selection of the Cognitive Reliability and Error Analysis Method (CREAM; Hollnagel, 1998a) to analyse the ASRS data. An overview of the generic form of the CREAM is provided, before discussing how the CREAM can be exploited to perform a sequential data analysis of the ASRS data. Taking into consideration the combined requirements of sequential data analysis, and the structure and content of the ASRS incident reports, the need for a new version of the CREAM is identified.

5.1 Existing Methods for Analysing Incident Report Data

The ASRS database is regularly used as a source of data by the aviation psychology community, in the investigation of operator behaviour in the aviation domain. The results of such research tend to appear mainly in specialist publications, such as *The International Journal of Aviation Psychology* or in the proceedings of the biennial *International Symposium on Aviation Psychology*. Although the topic of the investigations varies widely, there is a generally accepted method of conducting the initial search to find reports that are relevant to that particular topic (Mangold, Morrison, & Frank, 1995). The research studies tend to focus on the relationship between a single aspect of behaviour (usually an event or an action) and a particular class of problems, where this relationship is not directly represented in the ASRS database. Typically, these studies consist of two stages. The first stage is a search through the database to select a set of incident reports which relate to the selected class of problems. The second stage involves the processing of the selected reports. In general, this stage focuses on the narrative section of the incident reports, which is essentially a selective serial retrospective account of what happened during the evolution of the incident. The narrative section of each of the selected reports is manually processed, and there is often some recasting or reanalysis of the data into a format that is better suited to the way in which it is to be used.

In contrast, this research is somewhat different in that it is not a search for a simple relationship between a specified event or action and a problem or class of problems. Instead, the goal here is to find patterns of behaviour—sequences of actions and events—which relate to the particular class of problems that have been defined here as state misinterpretation. The level of abstraction used here is therefore slightly different to that which is generally used.

The first stage of the method of processing the reports here is the same as the general method, in that the class of state misinterpretation problems is used to select the initial set of incident reports (as described in the previous chapter). The second stage (also described in the previous chapter), however, divides the selected set of reports into a number of subsets, with one subset for each identifiable type of state misinterpretation. The third stage is to identify and categorise the erroneous actions that occur during each incident where state misinterpretation is implicated, and then to identify and categorise each of the elements in all of the causal sequences of events and actions which precede the erroneous actions. The overall process thus involves more than the two stage search that is more typically employed in the analysis of incident report data.

The reports of accidents and incidents which are recorded in databases such as the ASRS are categorised using a scheme which is often specific to the intended purposes of the stored data. In some instances, this categorisation scheme does not relate directly to human error. With a little effort, however, it is still possible to analyse the data using existing models of human error. The U.S. naval aviation accident database, for example, in which aviation mishaps are classified according to the severity of the accident, was reclassified by Wiegmann and Shappell (1997) using three existing taxonomies of error that were developed from the following relatively well known models:

1. A four stage model of information processing (e.g. Wickens & Flach, 1988).

2. A model of internal malfunction (O'Hare, Wiggins, Batt, & Morrison, 1994) which is based on Rasmussen's (1982) method for categorising information processing failures.

3. A model of active failures (Reason, 1990).

Wiegmann and Shappell found that whilst using the three frameworks helped to reveal meaningful trends in the types of human errors associated with each accident, the shortcomings in the frameworks meant that there were several human factors which could not be classified. They concluded that a more comprehensive model was required to take appropriate account of human factors such as air crew co-ordination, and physiological factors, as well as wider organisational and environment related issues. Their ultimate aim is to utilise a single model as the foundation for categorising the reported data on the basis of the human factors involved. Whilst their goal is laudable, there are two fundamental problems with trying to achieve it.

The first problem is how to deal with the existing contents of the database. Reclassifying the existing data using the new model is unlikely to be a viable exercise. The reclassification of the contents of ASRS database, for example, would be a massive undertaking, because the version used here (97-2) already includes around 66,000 reports, and new reports are received at an average rate of more than 2,000 per month. Although Wiegmann and Shappell (1997) explicitly recognised the infeasibility of abandoning existing reporting systems and starting again from scratch, they failed to offer any suggestions about how the existing data should be handled.

The second problem is that any model will necessarily be a reflection of the current state of human factors research. As a consequence, the model is unlikely to be permanently fixed, and will be subject to change over time as more is learned and understood about human behaviour in the operation of complex systems. This state of flux in the model leads to a dilemma over whether the classification scheme should be updated every time the model changes due to some new findings, with all the existing data being appropriately recategorised, or whether it should simply be accepted that the model is out of date.

The net result is that there may be some advantage in using a classification model which is not based on current best practice, as long as the data is available in a format that can be appropriately recast to reflect the latest developments in the field. In other words, the present state of affairs, whereby data which has been categorised using one taxonomy can be recategorised using another taxonomy which is better suited to the work at hand, seems to offer the best compromise solution.

It is worth noting at this juncture that Wiegmann and Shappell (1997) make explicit reference to Senders and Moray's (1991) comments about there appearing to be as many taxonomies of error as there are people interested in the topic. They have overlooked Senders and Moray's (1991) broader argument, however, which was that the different taxonomies are usually defined with different purposes in mind. Each of the different taxonomies is developed or used with the intention of focusing on an aspect of the data that is of particular importance to the research being carried out. So, for example, the taxonomy of state misinterpretation, which is later developed and used to group the data for subsequent analysis, is specific to this research. Furthermore, a taxonomy of state misinterpretation is not directly applicable to the level of observable behaviour—state misinterpretation is not a directly observable action or event—and hence provides another example of a taxonomy at a different level of abstraction of human behaviour (Toft and Reynolds, 1994) as previously discussed in Chapter 3.

The problems noted above are indicative of a wider problem: the difficulty of integrating existing theories of cognition. It is not possible to simply combine all of the existing theories into one, nor is it possible to make an eclectic model by selecting the best parts of the existing theories and assemble these into a single theory, because of differences in the assumptions and the levels of abstraction used by the various theories. At present there is no single framework which successfully combines the power and richness of expression that is provided by the separate use of multiple frameworks and theories. The idea of moving towards a single framework does ultimately make sense, however, and is consonant with the notion of developing unified theories of cognition (Newell, 1990).

The research presented here makes no attempt to develop a novel framework for analysing data, since this would be a major research project in itself. The shortcomings of the existing frameworks, identified by Wiegmann and Shappell (1997) and noted above, are not ignored, however. Instead, these shortcomings are addressed by employing an existing, but relatively novel method of analysing human performance data, called the Cognitive Reliability and Error Analysis Method (CREAM; Hollnagel, 1998a). The CREAM is a general method, however, so it needs to be appropriately tailored to suit the task domain.

The next section provides an introductory overview of the generic version of the CREAM. The CREAM inherently supports both predictive and retrospective analysis of human performance. Only the latter will be emphasised here, however, because this research is only concerned with the retrospective analysis of incidents which have already occurred, as described in the selected ASRS incident reports.

5.2 The Cognitive Reliability and Error Analysis Method

Many of the early methods for analysing human performance were developed without the use of an underlying conceptual framework, and simplistically categorised actions as either successes or failures. The majority of these methods were developed by the engineering community under the auspices of the field of human reliability analysis, and were geared towards the practical needs of that field. Most of the methods were developed in the early 1980s, before more detailed error classification schemes had been established. The inadequacies of existing methods of human reliability analysis were noted by Dougherty (1990) who highlighted the need for newer methods which took account of the latest best practice in the fields of human reliability analysis and the cognitive sciences.

The Cognitive Reliability and Error Analysis Method (Hollnagel, 1998a) was specifically designed to address some of the points raised by Dougherty (1990). In general terms, the CREAM attempts to unify research efforts from the field of human reliability analysis and the cognitive sciences. In the field of human reliability analysis the focus usually falls on the ability to be able to predict behaviour, often quantitatively through generating figures for the probability of the occurrence of particular types of error. In contrast, the cognitive sciences have tended to focus more on the need to understand and explain human behaviour, with a view to being able to predict it under given circumstances.

In the CREAM, performance takes place in a context which is defined by the interaction between three high level factors:

1. The human operator.

2. The technology (usually the system being operated by the human).

3. The wider organisation (including the environment in which the system is located).

Each of these factors is explicitly accounted for in the CREAM which includes a scheme for classifying actions and events. The CREAM also includes a relatively simple functional model of human cognition called the Contextual Control Model (COCOM; Hollnagel, 1993b). The COCOM asserts that human performance is the result of a combination of planning and reaction, or in Hollnagel's (1998a) terms, "... human performance is an outcome of the **controlled** use of competence adapted to the requirements of the situation..." (p. 154, original emphasis). In other words, cognition is regarded as the result of a combination of competencies—human cognitive functions, together with skills and knowledge which may have been assembled into compiled procedures—and the level of control exercised by the human during performance. The net result is that cognition is not considered as a fixed set of steps which must always be executed in sequence.

The CREAM also takes account of the current consensus view on the nature of erroneous behaviour. This view accepts that there is not a separate uniquely identifiable part of the human physiology which can be conveniently labelled *error generator*. Indeed, in many cases erroneous behaviour is simply the result of applying what appears to be the appropriate action in conditions which are similar to, but not quite the same as what the operator believes them to be. Once this notion of how erroneous behaviour arises is accepted, it becomes increasingly obvious that erroneous behaviour needs to be considered as an inherent facet of normal human behaviour. The margin separating correct behaviour and erroneous behaviour is often a very thin one: erroneous behaviour simply occurs as the result of a mismatch between the situation and the selected action(s).

Any erroneous actions that can be detected can be categorised as belonging to one or more of the eight possible error modes or effects—Hollnagel (1993b) also refers to these as the logical phenotypes of human error—each of which can manifest itself in several ways (shown in italics):

- Timing: The action is *too early/too late/omitted.*
- Duration: The action is *too long/too short.*
- Force: The action uses *too much/too little* force.
- Distance: The action is carried on *too far/too short.*
- Speed: The action is *too fast/too slow.*
- Direction: The action is performed in the *wrong direction* or involves *wrong type of movement.*
- Object: The action was carried out on a *proximal object/similar object/unrelated object* rather than the required object.
- Sequence: In a sequence of actions there was an *omission/skip forward/skip backward/repetition/reversal* or the *wrong action* was performed.

In addition, the CREAM makes provision for the inclusion of correct actions. Hollnagel (1998a) suggests that a category labelled *no erroneous action* should be incorporated either by adding it to each of the classification groups, or by keeping it as a separate group. This suggestion, which appears almost as an afterthought, is discussed in the description of the use of the classification scheme later in this section.

The category of error mode to which a particular instance of an erroneous action belongs may not always be immediately obvious. In such cases, careful consideration of the particular situation is required before any judgement can be made. This need to make considered judgements is critical to the field of human error research, since the labelling of a particular action as being erroneous is invariably a judgement made with the benefit of hindsight (Woods, et al., 1994).

Although accidents are often described in terms of chains of causal events, the analogy is somewhat misleading, because a chain usually consists of a single strand of links. The CREAM, however, attempts to identify all the possible precursors (or causes) of each error mode, which can be considered equivalent to trying to identify all the possible sub-chains within the chain which begins at the error mode. In other words, viewing from the error mode towards the causes, the possible sequences of events resemble a multi-way branching tree (or even a network) rather than a chain.

When the CREAM is used to perform a retrospective error analysis of an accident the results can be used to construct a causal tree of actions and events. Each action in a causal tree can have any number of precursors. This feature distinguishes causal trees from event trees—often used in human reliability analysis—since event trees are binary trees which represent the combined success or failure of each of the events that make up a particular sequence.

Causal trees can be generated by using the CREAM to trace back along all the possible branches of causal actions or events for each of the error modes identified for the accident in question. The error modes, which are observable manifestations of behaviour, form the root nodes of the trees of actions. An accident may have more than one identifiable error mode, and hence more than one root node. Each instance of an error mode is the result or *consequent* of some preceding antecedent. By extension, this antecedent can also be considered as the consequent of some other previous antecedent. In this way, branches (and sub-branches) of consequents/antecedents are built up, with each branch terminating where no further antecedents can be found for a particular consequent. The antecedents which make up the tree of actions for an accident comprise the attributed causes of that particular accident. A retrospective analysis of an accident using the CREAM is thus a search for the attributed causes of that accident. The CREAM constrains the complexity of the search through the use of three techniques, described below, which can be used to effectively limit the size of the search space.

5.2.1 Bounding the search for causes of an accident

On one level, the retrospective analysis of an accident using the CREAM can be considered as a search through a problem space of all the possible consequent-antecedent pairs. The CREAM only allows consequents and antecedents to be connected under certain conditions. These conditions are defined by taking into account theoretical and practical considerations. Even so, the number of possible combinations is potentially very large, which means that the search process is correspondingly complex. The CREAM makes the search more tractable by providing three methods with which to bound the search process:

1. Common Performance Conditions (CPCs).

2. Identification of possible error modes.

3. Identification of probable error causes.

The CPCs listed below are used to describe the context in which the accident occurred:

- Adequacy of organisation.
- Working conditions.
- Adequacy of human machine interface and operational support.
- Availability of procedures and plans.
- Number of simultaneous goals.
- Available time.
- Adequacy of training and experience.
- Crew collaboration quality.

In an attempt to bound the size of the search space, the CREAM explicitly considers the relationship between the various CPCs. If an increase or improvement in one of the CPCs would lead to a change in any of the others, the direction of the change is noted. So, for example, if the quality of collaboration between the crew improved, this would lead to an increase in the amount of time available to do the task. The CPCs help to bound the search to the extent that they help determine the possible error modes and the probable causes. The use of the CPCs is utilised more when the CREAM is being used for performance prediction rather than for retrospective analysis.

The CREAM also gives consideration to the possible error modes as a way of further bounding the search space. The likelihood of occurrence of each of the error modes will depend on the task itself, and the context in which the task is performed. Given sufficient domain knowledge, it should be possible to highlight which error modes are the most likely to occur, and which, if any, will never occur. The focus of attention can then be concentrated initially on analysing the most likely error modes first, whilst those which never occur can be excluded from further consideration.

Once the possible error modes have been identified, the final step in bounding the search space is to identify the probable error causes. The probable error causes can be found by qualitatively considering the relationship between the main groups of causes or *genotypes* (Hollnagel, 1993b)—the factors that combine to produce performance: human, technology, and organisation—and the CPCs. These causes can then be used to direct the search to the most appropriate parts of the problem space. There is one category of causes for each of these interacting factors which combine to constitute performance as shown in the left hand column of Table 5.1. Each of these categories can be decomposed into several subcategories, which, in turn may be decomposed into one or more further levels of subcategory. There is one classification group of antecedents allocated to each of the lowest levels of subcategories shown in Table 5.1. In other words, there is one classification group of antecedents for the entries which appear at the right hand side of the table: *observation*, *interpretation*, *planning*, and so on.

The CREAM classification scheme can be viewed as a hierarchical structure at one level of abstraction, as shown in Table 5.1. It should be noted, however, that the table reflects a simplified, static view of the CREAM classification scheme. The real power of the CREAM lies in the way that the tables which describe the classification scheme are linked together. Rather than using a simple hierarchical manner, the tables are linked by antecedent-consequent relations across tables which reflect the dynamic 1-to-many relations that exist in the real world. This linking scheme reflects the dynamic nature of the operation of complex systems, and takes

Table 5.1. Hierarchical decomposition of the classification tables in the CREAM. (Adapted from Hollnagel, 1998a).

<table>
<tr><td rowspan="6">Human</td><td rowspan="4">Cognitive</td><td rowspan="2">Analysis</td><td>Observation</td></tr>
<tr><td>Interpretation</td></tr>
<tr><td rowspan="2">Synthesis</td><td>Planning</td></tr>
<tr><td>Execution</td></tr>
<tr><td rowspan="2">Person</td><td colspan="2">Temporary</td></tr>
<tr><td colspan="2">Permanent</td></tr>
<tr><td rowspan="4">Technology</td><td colspan="3">Equipment failure</td></tr>
<tr><td colspan="3">Procedures</td></tr>
<tr><td rowspan="2">Human-machine interface</td><td colspan="2">Temporary</td></tr>
<tr><td colspan="2">Permanent</td></tr>
<tr><td rowspan="5">Organisation</td><td colspan="3">Communication</td></tr>
<tr><td colspan="3">Organisation</td></tr>
<tr><td colspan="3">Training</td></tr>
<tr><td colspan="3">Ambient conditions</td></tr>
<tr><td colspan="3">Working conditions</td></tr>
</table>

appropriate account of the role played by context in task performance. These features cannot be adequately captured using a static, hierarchical (and sequential) mechanism for linking actions and events together.

Each classification group of antecedents shown in Table 5.1 comprises a set of general consequents, and each set of these general consequents may be associated with a number of specific consequents, which are instances of the general consequents. So, for example, in the classification group of antecedents which relates to the consequences of failed observations, which is shown in Table 5.2, the general consequent *Observation missed* is associated with the specific consequents of *Overlook cue/signal* and *Overlook measurement*.

5.2.2 Detailed retrospective analysis of accidents

Having narrowed the search space using the above techniques, the detailed analysis can then be performed. This analysis is an iterative process (although it can also be carried out recursively), involving several steps.

The first step involves categorising the instance of the action (error mode or general consequent), using the set of tables which list the categories of error modes and consequents. There is one such table for each of the error modes, and one for each of the classification groups of antecedents listed at the right hand side of Table 5.1. The categories of consequents for failures of observation is shown as an example in Table 5.2. The left hand column lists the general consequents, and can be considered to be an abstraction of the specific consequents

Table 5.2. Categories of Consequents for "Observation". (Adapted from Hollnagel, 1998a).

General Consequent	Specific Consequent	Definition/explanation
Observation missed	Overlook cue/signal	A signal or event that should have been the start of an action is missed.
	Overlook measurement	Some information/event is missed usually during a sequence of actions.
False observation	False reaction	A response is given to an incorrect stimulus.
	False recognition	Some information/event is incorrectly recognised or mistaken for something else.
Wrong observation	Mistaken cue	A signal or cue is mistaken for something else.
	Partial identification	The identification of some information/event is incomplete.
	Incorrect identification	The identification of some information/event is incorrect.

listed in the middle column of the table. The right hand column in the table gives a brief textual definition or explanation of the meaning of the specific consequent. The format of the table for each of the error modes and the consequents is almost identical. The only difference is that in the tables of error modes the left hand column has the heading *General Effect* (rather than General Consequent), whilst the middle column has the heading *Specific Effect* (rather than Specific Consequent).

The second step is to find an antecedent for the general consequent (or the error mode) using the tables of general and specific antecedents. As with the tables of categories of error modes and consequents there is one such table for each of the classification groups of antecedents listed in the right hand column of Table 5.1, although there is only a single table for all of the error modes. These tables of general and specific antecedents define the links between consequents and antecedents. The general and specific antecedents corresponding to the category of consequents for failures of observation is shown as an example in Table 5.3. The left hand column lists the general consequents, which are used to provide a link between the categories of consequents table (such as Table 5.2) and the immediate antecedents of the general consequent selected in the first step of each stage. The antecedents listed in the tables of links are divided into general antecedents and specific antecedents; the latter can be thought of as specific instances of a particular type of action.

The matter of the inclusion of correct actions in the CREAM classification scheme raises some

Table 5.3. General and Specific Antecedents for "Observation". (Adapted from Hollnagel, 1998a).

General consequent	General antecedent	Specific antecedent	
Observation missed	Equipment failure Faulty diagnosis Inadequate plan Functional impairment Inattention	Information overload Multiple signals	Noise Parallax
False observation	Fatigue Distraction	None defined	
Wrong identification	Distraction Missing information Faulty diagnosis Mislabelling	Ambiguous symbol set Ambiguous signals Erroneous information	Habit, expectancy Information overload

interesting and important issues. Hollnagel (1998a) suggests that a category (*No Erroneous Action*) could either be included in each of the classification groups, or kept as a separate group. In theory it should be possible for correct actions to precede any other action—correct or otherwise—and, in turn, for a correct action to be preceded by any other action. The corollary is that the set of antecedents for a correct action is the complete set of possible actions. In other words, every general consequent in the CREAM tables of antecedents would have to have *correct action* as a general antecedent, and *correct action* would have to have every other possible action as a general antecedent.

One way around the problem lies in the fact that the CREAM differentiates between direct antecedents and indirect antecedents. A direct antecedent is one which appears in the same row of the table of general and specific antecedents, whereas an indirect antecedent is one which appears in one of the other rows in the same table. So, if *correct action* was added to every table, with a general antecedent of *correct action*, all of the other actions in the table could be considered as indirect antecedents. Similarly, every consequent in the table could have *correct action* as an indirect antecedent.

The other alternative is to simply ignore correct actions. The only problem with this solution is that it makes the links between consequents and antecedents a causal one, which is slightly stronger than Hollnagel (1998a) intended. Unless the data is being analysed using a technique which investigates the number of intervening actions between pairs of particular actions of interest (such as Lag Sequential Analysis), however, there is no problem with ignoring correct actions.

The use of direct and indirect antecedents provides yet another mechanism for focusing on the most likely antecedents first, since antecedents which appear in the same row of the general and specific antecedents tables are more closely (causally) related than those which appear in different rows.

The next step depends on the outcome of the search for an antecedent in the previous step. If a general antecedent is found, the process starts again at the top, using this general antecedent as the general consequent. If, on the other hand, the antecedent is a specific antecedent, or no antecedents can be found, the process is terminated for the current branch of the tree.

The main part of the analysis can be summarised by the algorithm shown in Table 5.4. The algorithm has to be executed for each accident that is to be analysed.

Table 5.4. Summary of the the basic steps involved in processing the data for each accident.

1. *Determine next error mode or general consequent.*
2. *Identify specific consequent for that error mode/general consequent.*
3. *Look for next antecedent.*
4. *If a general antecedent is found then continue else terminate analysis for this current branch*

The process is repeated for each error mode for a given accident, and for all the antecedents of each of the consequents. In this way all of the possible branches of the tree of consequents/antecedents are identified. The construction of a particular branch of the tree ends when either a specific antecedent is found for a particular consequent, or when no antecedents can be found for a particular consequent.

Hollnagel's (1998a) version of the CREAM is intentionally generic. It does, however, show some biases that reflect the method's origins in the results of several years' research in the industrial process industries, particularly nuclear power generation. The CREAM was developed for use in accident and event analysis, which means that it should be suitable for use in the analysis of incidents too, since an incident can be thought of as an accident which has been avoided (often by timely intervention on the part of the operator). Although the CREAM is designed to support both predictive and retrospective analyses, in this research only those parts that apply to retrospective analysis shall be used.

5.3 Summary

The first part of this chapter discussed the methods and frameworks available for analysing incident report data. In particular, issues surrounding the classification of incidents were considered. Given that research into human performance is very much an ongoing subject, it was concluded that there is no definitive classification scheme, since each scheme will simply reflect the state of research at a particular point in time. The most important factor for any database of incident data is that the details of each of the incidents are kept available. As long as the details are stored, it should be possible for the data to be appropriately recategorised using any of the available methods.

In the second part of the chapter an overview of the Cognitive Reliability and Error Analysis Method (CREAM; Hollnagel, 1998a) was provided. The CREAM unifies previous work in the fields of human reliability analysis and the cognitive sciences to produce a single method of analysing behaviour either in a predictive or a retrospective manner. At the heart of the CREAM is a mechanism for identifying the precursors of the erroneous actions which can give rise to an accident. Although the CREAM is directed at the analysis of accidents (and events) most of the features of the CREAM are suited to the analysis of incident data too. The next chapter describes how the CREAM can be adapted to facilitate the analysis of the selected ASRS incident reports.

6 Analysing the Data Using the CREAM-AIR

The chapter begins with a recapitulation of ASRS incident reports in order to identify the overall structure of the raw data. Due to the sequential nature of the data, the use of sequential data analysis techniques is proposed. Since the data analysis is to be automated where possible, the use of appropriate software tools is discussed. The ASRS database is supplied in read only format (on CD-ROM), so a method for extracting the reports from the database is then described. Finally, the method for analysing the data is described, and applied to identify the type of state misinterpretation and the error modes involved in each of the incidents. During the data analysis the CREAM-AIR method is developed as an extension of the CREAM appropriately tailored to the aviation domain and the particular goals of the research.

6.1 ASRS Incident Reports

The incident reports stored in the ASRS are provided voluntarily by flight crew members (or air traffic controllers) who experienced the incident firsthand. The reporters—incidents may be reported by several members of the flight crew, and, where more than one aircraft is involved, by members of different flight crews—fill in the details of the incident on the ASRS reporting pro formas. These forms include a section for a narrative textual description of the incident. If there is a need for clarification of any of the incident details supplied, the ASRS analyst who logs the incident can telephone the reporter to ascertain the required information. The details of these so-called *callback* conversations appear as an addendum to the textual narratives in the report that gets stored in the ASRS database. Before the reports are stored they are sanitised by removing any details that could be used to identify the reporter.

Where an incident is reported by several protagonists, only one report is stored in the ASRS database. Extra details from the narrative section of the other reports of the same incident are appended to the end of the narrative section of the stored report, and explicitly marked as supplementary information. In addition, the stored report is explicitly flagged as containing multiple reports.

Although each incident report in the ASRS database includes a plethora of fields which can be used in categorising the incident, the main section of interest here is the textual narrative description. The level of detail provided in the narratives (which can be considered as retrospective written protocols) varies widely between reports. In the extreme case, the narrative consists of a single sentence. The length of the narrative section should not, however, be taken as a guide to the quality of the information contained in the report. In some cases longer reports make it harder to determine the correct causal sequence of actions and events that occurred during the evolution of the incident. The main reason for this is that sometimes reporters try to record everything that they can think of which may be related to the incident. In so doing, they report events in the order in which they recall them, rather than the order in which they occurred.

The narrative section of the reports is essentially a kind of sequential data, in that it generally describes the details of the incident in a time sequential manner. The causal trees which are produced during the analysis are also a time sequential representation of the (causal) details of the incident, due to the inherent temporal nature of the consequent-antecedent links that are part of the CREAM. The sequential nature of the data means that it can be analysed using exploratory sequential data analysis (ESDA; Sanderson & Fisher, 1994) techniques.

In general it makes more sense to use automated tools to support the analysis of the data. The use of appropriate software tools can help to impose an extra level of rigour (and repeatability) on the analysis. It is particularly important to make sure that the analysis is performed in the same way when the data consists of several hundred reports which have to be analysed and the results then compared. The next section considers the available options for software support of the data analysis.

6.2 Software Support for Data Analysis

The way in which the analysis is performed and recorded has to be carefully considered, since this will determine the type of tools and techniques that are required. Although it is possible to conduct the analysis using pencil and paper, there are now a range of software tools available to support ESDA. In general these tools make the analysis easier because they automate at least some of the fundamental operations. The choice of tool(s), however, requires the use of critical judgement which should take account of the following factors (Bainbridge & Sanderson, 1995):

- The tool(s) should support the data analysis, rather than having to excessively adapt the analysis to fit the tool(s).
- The fact that some other person may have found one particular tool useful in analysing sequential data should not be taken to mean that the tool in question will be suitable for all types of sequential data analysis.
- Similarly, the simple fact that a tool is described as a sequential data analysis tool, should not be taken to mean that the tool is ideally suited to all types of sequential data analysis.

- A track record of success for a particular tool does not automatically make it the best tool for all types of analysis.
- The lack of suitability of a tool for a particular type of analysis simply indicates that the tool in question is not best suited to that type of analysis, rather than indicating some failure in the way that the analysis was carried out.

The list highlights the tension between using existing tools for analysing data and developing purpose built tools to do the analysis (Sanderson & Fisher, 1997). It is worthwhile briefly exploring the structure of the data and its analysis to guide the decision about which tool(s) to use.

The data to be analysed consists of 493 incident reports after the false positives have been removed as described in Chapter 4. Each incident report is a relatively short, almost episodic, sequence of actions and events. All of the incidents will eventually be represented by a causal tree. The input data—the narratives from the ASRS incident reports—is free form text, which in most cases incorporates some element of temporal ordering. There is no explicit timing information associated with the data, however, beyond the time of day when the incident happened, which is recorded to the nearest quarter of a day. This lack of detailed timing information should not cause any problems. The causal analysis of the data will necessarily include a qualitative temporal element, because the cause of an action must necessarily occur before that action. In addition, the use of causal analysis provides for multiple paths through the same data, in that there may be several causal branches which lead to the error mode(s) at the root of the causal tree for a particular incident.

Flying an aircraft is a complex skill and, like many complex skills, is highly proceduralised. The procedures, which consist of sequences of actions and events, are often long and complicated, and in some cases may have to be performed concurrently. The analysis of the data has to be able to uncover the structure of these procedures from the incident reports. The use of a domain specific toolset helps to constrain the selection of actions required as part of the analysis, whilst still providing the potential for future expansion. In other words, the analysis becomes easier because the tools help to bound the search for, and the selection of actions. The analysis is also made more consistent, because the encapsulation of the constraints on possible sequences of actions and events imposed by the CREAM-AIR in the tool helps to improve the reliability of coding across multiple incidents.

Ostensibly it appears that the MacShapa sequential data analysis tool (Sanderson, Scott, Maintzer, Johnston & James, 1994) could be used here to analyse the ASRS incident data, because it can support the following functions, which are closely related:

- Use of the CREAM-AIR's unique classification scheme.
- Retrospective analysis of the ASRS incident data, using the CREAM-AIR.
- Classification (using the CREAM-AIR) of individual instances of actions and events, that are often context dependent, which relies on human judgement.

MacShapa does not support the following specific requirements of this research, however:

- Generation and analysis of causal trees.
- Support for the CREAM-AIR dynamic linkage mechanism which only allows valid antecedents to be selected for a particular consequent.
- Execution on an IBM PC compatible machine, because this is the only platform on which the ASRS CD-ROM can be used. If multiple types of machines are used, time can be unnecessarily wasted on converting between file formats.

It was therefore decided that the best compromise was to use the Microsoft Excel spreadsheet as the basic tool to support the analysis. Although Excel is not the ideal tool, it does provide a set of basic facilities for manipulating data. In particular, Excel can perform calculations, and sort data, as well as supporting the development and use of purpose written macros. Equally important is the fact that Excel can import and export data in a text format. This feature means that the data can be converted into a form that is readable by other analysis tools, since most of them can read textual data. Excel was supplemented by purpose built software tools (described in Appendix B) to perform the following functions:

- Generating a concordance of individual actions across all the incidents.
- Generating a frequency table for sequences of actions across all incidents.
- Quantitative comparing of causal trees.
- Calculating relational metric values for pairs of actions.

The software tools have been developed specifically to support the use of the CREAM-AIR to perform a retrospective analysis of ASRS incident reports. These tools implement the CREAM classification scheme and linkage mechanism in such a way that only valid combinations of consequents and antecedents can be selected.

Having decided to use Excel as the basis for the analysis of the data, the next step is to extract the data from the ASRS reports so that it can be stored and manipulated. This process is described in the next section.

6.3 Extracting the Data from ASRS Database

The ASRS database is supplied as a CD-ROM, which means the contents of the database cannot be manipulated *in situ*. Once the required incident reports have been selected using queries to the database, these reports have to be extracted from the database before their contents can be manipulated. The ASRS database provides a mechanism for exporting incident reports to file, which allows the various fields in the report to be selectively included in the export file. Although the number of different data formats the ASRS can generate is limited, it can export data in a text form which can be imported by Excel, with the constraint that any data fields which extend beyond the maximum cell size supported by Excel (255 characters) are truncated during the exporting process. This constraint makes it impossible to directly export the narrative sections of the reports into Excel, because the narratives usually occupy more than 1,000 characters. It is possible to circumvent the length limitation problem by first decomposing the narrative text into single sentences using the following process:

1. Export each report from the ASRS database as text.
2. Open up each report using a text editor and extract the narrative section.
3. Remove the paragraph markers at the end of each line of the narrative.
4. Convert the text in the narrative to sentence case.
5. Reformat the narrative into paragraphs by adding a paragraph marker at the end of each sentence.
6. Use cut and paste to transfer the data from the text editor into a Microsoft Excel spreadsheet.

The choice of text editor in step 2 is an important one. If a word processor, such as Microsoft Word, is selected, then when the report that is exported from the ASRS database (in step 1) is opened, much of the rest of the process can be automated by developing appropriate macros.

There may also be some minor adjustments required in step 6. The sentences used in the incident report narrative sections occasionally exceed the maximum cell size supported by Excel of 255 characters. In such cases the sentence is manually split at any grammatical conjunction which occurs in the sentence (usually the word "and"), and is stored in two vertically adjacent Excel spreadsheet cells.

The data from the narrative sections of the ASRS incident reports is transcribed verbatim into Excel. Normally during the transcription of data, meaningless parts of that data can be removed. In this instance, however, all of the data has to be transcribed because the analysis is intended to uncover all of the possible causal sequences of actions. It is therefore not possible to selectively eliminate any parts of the narrative, because it is not always possible to make an informed judgement about which parts of the data are important without doing the causal analysis.

Once the data has been imported into Excel it is ready to be analysed in detail. Since the data is of a sequential nature it can be analysed using ESDA techniques as described in the rest of the chapter.

6.4 Analysing the ASRS Incident Data Using the CREAM-AIR

The process of retrospectively analysing an accident using the CREAM is a search for the attributed causes of that accident. In order to make this search more tractable, the CREAM constrains the search space by focusing on the particular contextual aspects of the accident using the CPCs, and concentrating attention and effort on the most common error modes and causes. These constraining techniques are less applicable to incident analysis, because the level of detail available in incident reports is usually much less than that which appears in accident reports. The latter are usually the product of a lengthy major inquiry, in which a team of experts examines all aspects of the accident in minute detail. As such, accident reports typically run to hundreds of pages, whereas incident reports, such as those held in the ASRS database, rarely occupy more than a few pages, and generally only represent the perspective of the protagonists.

The needs to constrain the space of possible actions and events are therefore excluded here. Removing the need to generate an explicit contextual description of the incident (using the CPCs), and the lists of the possible error modes, and probable error causes, however, leads to an increase in complexity of the analysis. This increase in complexity is offset through the use of software tools to support the data analysis, however.

The analysis of the data proceeds in parallel with the adaptation of the CREAM to the aviation domain. As noted earlier, the extended method is referred to as the CREAM-AIR—the CREAM for ASRS Incident Reports—in order to distinguish it from the generic version of the CREAM. Although the AIR suffix is intended to stand for ASRS Incident Reports, in a broader sense it could also be used to stand for aviation incident reports.

One of the goals of this research is the identification of common sequences of causally linked actions across multiple incidents in which state misinterpretation is implicated. The CREAM-AIR method of analysing the narrative sections of the incident reports has been developed to meet the particular goals of this research, whilst still adhering to the basic philosophy of the CREAM. Where possible, the software tools that have been developed to support the method of analysis, have been designed in such a way as to introduce an element of rigour and repeatability into the analysis.

The selection of data for analysis and the removal of false positive reports have already been addressed in Chapter 4. The remainder of this section therefore focuses on the detailed analysis of the narrative sections of the selected reports. The basic operations of ESDA are introduced first. The reports are then analysed to determine the type of state misinterpretation, and the error modes involved in each of the incidents.

6.4.1 The fundamental operations of ESDA

The fundamental operations or data transformations that need to be considered during ESDA were identified in a broadly based review of existing techniques and practices by Sanderson and Fisher (1994). These operations essentially combine data smoothing techniques with the philosophy engendered by Tukey's (1977) principles of Exploratory Data Analysis (EDA). The operations support ESDA in three ways:

1. By assisting in the choice of the right level of detail to include in the analysis;
2. By establishing a scheme for encoding the data; and
3. By uncovering any regularities that may exist in the data.

Sanderson and Fisher (1994) refer to the fundamental operations of ESDA as *the eight Cs*. These operations can be summarised as:

- Commenting: annotating the raw data.
- Chunking: grouping together contiguous actions or events.
- Coding: redescribing the data into smaller, theoretically meaningful sets of terms.

- Connecting: identifying the relations between non-contiguous events or actions.
- Comparing: testing for similarities between sets of events or actions.
- Constraining: analysing a portion of the data in more detail to look for threads or higher order structures.
- Converting: representing the data in different forms, such as graphical notations to identify any non-immediate patterns.
- Computing: manipulating the data using statistics, or symbolic inferencing.

In general, at least some of these operations will be performed when using ESDA, rather than all of them (Sanderson & Fisher, 1994). The operations are shown as discrete entities but this does not mean that they have to be discretely performed in the order in which they are shown here. Indeed, there will often be a degree of overlap between operations, and some of them may be performed as a side effect of the use of other operations. The remainder of this section describes how the analysis of the ASRS data proceeds, and points out where the various fundamental ESDA operations are performed during the analysis.

6.4.2 Classification of state misinterpretation

The first step in the analysis is to identify the type of state misinterpretation that occurred during each of the incidents, using the narrative sections of the reports. Once all of the incidents have been processed at this level it should be possible to establish a taxonomy of the different types of the reported instances of state misinterpretation.

The concept of state misinterpretation was defined in Chapter 3, and a mechanism for detecting the occurrence of state misinterpretation was developed in Chapter 4. This detection mechanism is based on a combination of symptoms, whereas the identification of the type of state misinterpretation requires a more detailed consideration of the context in which the incident occurred.

State misinterpretation arises when the system operators—the flight crew—take some action in the belief that the system is in a normal state (when it is not) and that action is not directed at diagnosing the system state or moving the system to a normal or safe state. It is also a state misinterpretation if the operator fails to take some action to diagnose the system state or move the system to a normal or safe state, when the system is in an abnormal state but the operator believes it is in a normal state. The state misinterpretation itself is not an observable action, and hence has to be identified using the narrative section of the incident report. In some cases this will require that one or more inferences have to be drawn. So, for example, if an aircraft fails to level off at a specified altitude and this is not noticed until some time after the failure to level off, then this is an instance of state misinterpretation, because it can be inferred that the crew thought that the aircraft was (still) climbing to the specified altitude. Each instance of state misinterpretation may be of a transitory nature, only lasting until it is detected. If it is not detected, then the situation could develop into an incident or accident.

The crucial factor in determining the occurrence of state misinterpretation is that the believed state and the actual state must be identifiable (or at least directly inferable) from the incident report. The difference between the two states may be subtle and may even be impossible to directly observe. If the believed and actual states cannot be identified, then it is not possible to determine that a state misinterpretation occurred.

The different types of system state will naturally be highly domain specific. It is therefore to be expected that there will be several different types of system state misinterpretation for a particular domain. In addition, the fundamental underlying assumption that the system is a socio-technical one means that the system state at any particular point extends far beyond the computer system which is used here to control the aircraft.

Defining a taxonomy of types of state misinterpretation for the aviation domain is a problem which has two possible solutions. The first solution is to identify the potential types of state misinterpretation in advance of the analysis of the data. The two main drawbacks of this solution are that it is not clear when the taxonomy can be regarded as complete—and hence when the data analysis should begin—and what should be done if an incident highlights an instance of state misinterpretation which has not been defined. There is a tension between the two drawbacks, since the simplest solution to the second drawback is to allow new types of state misinterpretation to be defined as and when appropriate during the analysis of the incident reports.

The second possible solution derives directly from the solution to the second drawback of the first solution, and simply allows for the taxonomy to evolve as the incidents are analysed. The main drawback of this solution is that it could be seen as an ad hoc approach, which, in the worst case, would allocate a separate type of state misinterpretation for each of the incidents that is analysed. The simplest way around this drawback is to iterate over the taxonomy, using an appropriate level of abstraction in order to ensure that similar instances of state misinterpretation can be grouped together.

The latter method is used here, and the identification of the type of state misinterpretation for the selected incidents is conducted as an iterative process. Each of the reports is initially labelled with a textual comment which describes the state misinterpretation in a precise way. Once all of the reports have been processed in this way, a second pass over the data concentrates on the state misinterpretation labels, and groups them accordingly. So, for example, in the first instance the type of state misinterpretation identified for report number 339770 is:

climbing to flight level 340 (when the aircraft had flown through that flight level)

then, on the second pass through the data that instance of state misinterpretation would be generalised, and end up being categorised as belonging to the state misinterpretation type of:

climbing to altitude <x>

where *<x>* is a placeholder for any altitude.

The abstraction process used in the second pass through the data mostly generalises over numbers. This notion of abstraction is one that has been discussed earlier. The difficulty lies in finding the right level of abstraction to use. Initially a relatively low level of abstraction is selected, on the grounds that after the data has been analysed at that level, further abstractions can be made in an incremental manner as required. Using too high a level of abstraction on the first pass through the data may remove some important low (or even intermediate) level detail. The salient details which are removed can only be recovered by reintroducing them back into the data, but the difficulty lies in determining how much of the lower level detail to reintroduce.

The results, after two passes through the data are shown in Table 6.1. The most striking feature of the data is the relatively low frequency of reports that fall into each of the different types of state misinterpretation. From a starting point of 493 reports, the biggest single category (*Cleared to Altitude <x>*) only contains 50 incident reports, and in total there are only 11 categories with frequencies greater than 10.

Table 6.1. Summary of Frequencies of Occurrence of State Misinterpretation.

Type of State Misinterpretation	Frequency
Air conditioning switched on	2
Airport was correct destination airport	2
Altimeter setting OK	25
Altitude alerter set for cleared altitude	7
Calculated weight value OK	2
Clearance for other aircraft	9
Clearance obtained before take-off	4
Cleared for <x> arrival route	2
Cleared on Standard Instrument Departure route	3
Cleared on standard/normal route	2
Cleared onto runway <x>	6
Cleared route as filed	3
Cleared to altitude <x>	50
Cleared to cross active runway	3
Cleared to heading/turn <x>	6
Cleared to land	19
Cleared to land on runway <x>	4
Cleared to push by ground crew	12
Cleared to turn at <x>	2
Climbing to altitude <x>	28
Cruising at altitude <x>	2
Descending to cleared altitude <x>	35
Descent to crossing restriction	8
Flight Management System was correct	2
Flight Management System was executing programmed route	3
Flying cleared route	36
Fuel level OK	9
Landing gear pins removed	4
Landing gear was up	2
Landing gear/tyres OK	2
Level at altitude <x>	9
Levelling off at <x>	4
Maintenance check not required before take-off	2
No crossing restriction on descent	8
No speed restriction	4
OK to start engines	2
On glideslope	2
Position in air short of action point	13
Position of crossing restriction was at <x>	3
Position on ground at <x>	2
Position on ground short of action point	20
Rate of descent OK	14
RNAV frequency setting OK	16
Runway <x> was runway <y>	3
Selected communications frequency OK	6
Slats/flaps set OK	6
Speed was <x>	2
Take-off data was for runway <x>	2
Taxiing along cleared route	3
VNAV mode engaged	2

The other notable feature is the number of different types of state misinterpretation shown in the table. In all, there are 50 different types shown. This is not the full picture, because only those types of state misinterpretation which occurred more than once are shown. In addition to the types of state misinterpretation shown in the table, there were also 76 types of state misinterpretation which only occurred once.

The taxonomy of types of state misinterpretation highlights that there are instances of state misinterpretation which can be grouped together on the basis of their common features or characteristics. This taxonomy will be used when analysing the causal sequence data for the ASRS incidents to identify common patterns of actions. Having established that there are different types of state misinterpretation, the next step is to identify the different error modes for each of the incidents.

6.4.3 Identifying the error modes

In parallel with the identification of the type of state misinterpretation for each of the reports, the identifiable error modes for each of the incidents are also identified; there may be more than one for each incident These error modes are the observable manifestations of error, and form the starting point for the retrospective causal analysis of the incident. In other words, it is the error modes which form the roots of the causal trees that are generated during the analysis.

For each of the error modes the following information is recorded against the transcribed data:

- The type of CREAM-AIR error mode.
- The general and specific effects.
- A brief textual comment describing the error in English.

So, for example, for the incident report with accession number 357140, in which the pilot failed to meet a crossing restriction assigned by ATC, the information that is recorded is:

Error Mode	Action at Wrong Time
General Effect: Specific Effect	Timing: Too Late
Description	Late start to descent to cross waypoint OJAAY

In this example, OJAAY is the name of the way point which the aircraft was cleared to cross at a specific height (10,000 feet, in this instance) by ATC.

In most cases, the identification of the error mode(s) for a particular incident is a relatively straightforward process. There are some cases, however, where the issue is less clear cut. The main problem arises because of the physical relationship between time, distance, and speed, each of which appears in the classification scheme as an error mode category. A detailed consideration of the available facts for the incident will usually help to pinpoint the most appropriate error mode to use, as the following example taken from Hollnagel (1998a) will show.

An accident occurred on the New York subway in 1995 when a train passed through a red light and ran into the back of another train, killing the driver and injuring 54 passengers. The trains are fitted with an automatic braking system which is supposed to stop the train when it passes through a red light. In this case, the braking system appeared not to work. Ostensibly the two possible combinations of general and specific effects of the error modes are *distance: too far* and *speed: too fast*. A more detailed consideration of the facts suggests that the latter pairing is less appropriate in this instance, however, because it was not the speed which caused the direct consequence—the collision between the trains—but the fact that the train travelled the full distance between the red light and the train in front of it on the same line. Further analysis of the accident showed that the problem ultimately lay in the fact that the placement of the signals had been based on stopping distances that were calculated in 1918, and had never been updated to take appropriate account of changes in train technology and volumes of traffic.

During the process of identifying the error modes for the selected incidents, it soon became apparent that the CREAM-AIR error mode categories would have to be extended. The aviation domain requires that the aircraft—the system that is being controlled by the operators—be controlled in three dimensions. In order to take appropriate account of erroneous actions in the vertical dimension, the definition of the error mode category of *Action of wrong type* was extended to incorporate descending and climbing too far and too short.

The identification of the error modes is a lengthy process, which requires that all of the transcribed data from the narrative section of all of the incident reports be read. The quality of the reports varies somewhat, as has already been noted, so comments are added in a worksheet cell adjacent to the sentence to which the comment applies. In general these comments are limited to points of clarification, and are used for reference purposes during the coding of the data.

Table 6.2 shows the frequency of occurrence for each of the possible error modes. The figures are shown based on the raw data, rather than sub-divided into the figures for each of the different

Table 6.2 Summary of Frequencies of Error Modes.

Error Mode	Frequency
Timing/Duration	128
Sequence	97
Force	0
Distance/Magnitude	169
Speed	17
Direction	30
Wrong Object	69
Total	510

types of state misinterpretation. The total number of error modes is 510 whereas the number of incident reports is only 493. This apparent discrepancy arises because there are 15 incidents in which two separate error modes can be identified, and one incident in which three separate error modes can be identified.

The most striking result is that there are no occurrences recorded for the error mode *Force*. Although this seems strange, there are few occasions where the use of too much or too little force would be reported when flying an aircraft. If too much (or too little) force was being used this would usually become apparent via feedback in terms of the aircraft's performance, and an appropriate adjustment made to immediately correct the problem.

The other point that is worthy of note is that the highest number of occurrences are logged against the error mode of *Distance*. In the light of the CREAM-AIR extension to include climbing and descending too far in the *Distance* category, this result is in line with expectations, since a large proportion of the incidents recorded in the ASRS database are categorised as altitude deviations. where the aircraft either climbed or descended too far.

Having identified the types of state misinterpretation and the error modes for each of the incidents, the next step is to identify the sequences of actions and events which precede the error modes, and hence surround the state misinterpretation. By so doing, it should become possible to identify any common sequences of actions and events, and hence suggest how the different types of state misinterpretation can either be prevented or managed to mitigate any potential adverse consequences.

6.4.4 Encoding using the CREAM-AIR

The identification of all the possible sequences of antecedents for each of the error modes involves encoding the data from the narrative sections of the incident reports. The encoding of the data takes place on two levels simultaneously. The first level is a description of the antecedents of the error modes involved in the incident in a standardised textual form. This textual form is based on the use of the definitions and explanations that are provided in the CREAM-AIR tables used to categorise actions and events. The coding is performed using a purpose built software tool, which allows the user to select an appropriate definition or explanation for each action (or event). These definitions are saved to file, and effectively provide a standardised textual account of the sequence of causal events for each incident.

The second level, which is an encoding of the antecedents of the error modes using a shorthand notation, is somewhat closer to the methods of coding used in ESDA (Sanderson & Fisher, 1994). This encoding is carried out automatically by the software tool. When the user selects an error mode or an antecedent, the software tool determines the appropriate code to be used. These codes are saved to file, such that there is a sequence of (causally linked) codes for each of the paths in the causal tree for each of the incidents.

The CREAM-AIR coding support tool (described in Appendix B) is an encapsulation of the CREAM-AIR method. As such, it includes an implementation of the CREAM-AIR tables, and the mechanism which allows consequents and antecedents from different tables to be dynamically linked together in a causal manner. For each selected consequent, the coding support tool determines all of the possible antecedents by searching the appropriate CREAM-AIR tables. The results are then presented as a list, from which the most appropriate antecedent is selected. The CREAM-AIR support tool thus eliminates the need to search for all the possible alternatives before the selection of an antecedent can be made. If a novel consequent-antecedent relation is discovered during the analysis of an incident, this can be dynamically added to the CREAM-AIR tables, which can be reloaded into the coding support tool on the fly. In this way, the coding support tool is refined in parallel with the tables as the analysis proceeds.

The choice of coding scheme is an important decision, which again relates to the issue of selecting the appropriate level of abstraction for analysing the data. The CREAM-AIR follows the lead of the CREAM, and hence commits to a coding scheme which uses observable actions as the level of abstraction. In other words, all of the error modes which appear in the CREAM-AIR tables are directly observable actions, and all of the antecedents in the CREAM-AIR tables are connected to these observable actions either directly, or indirectly, through a causal relationship.

There are two possible approaches to generating the CREAM-AIR's coding scheme. The first is to use the contents of the incident reports to construct a coding scheme from scratch. The main advantage to this approach is that there will be a close fit between the coding scheme and the data, since the former is directly derived from the latter. The main disadvantage is that, unless great care is exercised, the coding scheme will be too specific, in that each incident could give rise to a specific code (or set of codes), which would make it difficult to compare and contrast the data across several incidents. In other words, once again, the choice of the appropriate level of abstraction is crucial. The other big disadvantage is that there is potentially a large duplication of effort. Developing the CREAM-AIR coding scheme from scratch would result in the rediscovery of consequents and antecedents that already exist in the CREAM coding scheme, because some of them will be generally applicable across domains.

The second approach is to base the CREAM-AIR's coding scheme on the one that underpins generic version of the CREAM (Hollnagel, 1998a). The generic version simply gets adapted and extended according to the specifics of the ASRS incident data. There are two main advantages to this approach. The first is that the data can be processed in a single pass, rather than having to generate the coding scheme in one pass through the data, and then apply it during a second pass. The second is the tacit recognition that there are certain aspects of human behaviour which are applicable across domains. The main disadvantage of this approach is that the coding scheme will potentially carry some redundant items in it, since the generic CREAM was developed largely on the basis of experience in the nuclear power domain.

The latter approach is adopted here, principally because it is more consistent with the idea that human behaviour is the same, to some degree, across domains and applications. In addition, this approach requires less time because the coding of the data can be done in a single pass. If the coding scheme was developed from scratch, there would be a potential need to recode the previously coded data every time the coding scheme changed. The overhead of having to carry the redundant codes is also judged to be less of a problem than that of the multiple passes over the data (and coding scheme) required by the first approach.

The classification scheme at the heart of the CREAM-AIR covers issues of both coding and connecting in ESDA terminology. The generic CREAM coding scheme (Hollnagel, 1998a), which is largely domain independent, was originally developed in such a way as to strike a balance between the need for a level of abstraction that is appropriate to the genotypes in the classification scheme, and the need for pragmatism so that use of the method remains tractable. Any biases which are apparent in the scheme are an artefact of the classification scheme's genesis in the nuclear power domain.

One particular shortcoming of the generic CREAM coding scheme identified by Hollnagel (1998a) is that the number of specific antecedents is comparatively low, and should be extended as part of the adaptation to reflect the particular aspects of the aviation domain in this case. The development of the CREAM-AIR takes appropriate account of these principles in the way that it allows extensions to be added to the generic method.

Many of the links which exist in the CREAM classification scheme hold true for a number of domains. In order to develop the CREAM-AIR classification scheme which is more appropriate to the aviation domain, the main area which needs to be addressed is the number of specific antecedents which are included in the classification scheme. The specific antecedents are the area of the classification scheme which most strongly reflects the domain. In addition, however, there may be a need to add extra links to the classification scheme. It is vital that any changes to the classification scheme do not violate the integrity of the scheme. The following criteria are used to ensure that integrity is maintained:

- Any consequents which are added to the classification scheme must have at least one associated antecedent (even if this antecedent is "none").
- Specific antecedents can only be added to the classification scheme if they occur in an incident report, or in the aviation literature. This criterion is used to ensure that the specific antecedents reflect the true nature of the domain rather than a normative description of what the nature of the domain should be.
- Additional links should only be added when the scheme does not provide an appropriate link. This criterion requires the careful exercising of judgement in order to avoid inferring the occurrence of actions or events where the narrative description of the incident does not provide enough information to support such an inference.

At the completion of the coding of the selected incident reports the CREAM-AIR contained four new general consequents (making a total of 60), and 36 new specific consequents (making a total of 160). In addition, several new domain specific descriptions were added to the tables of general consequents. These changes to the CREAM tables reflect the specialisation of the CREAM to the aviation domain.

The CREAM-AIR coding scheme relies on the basic table structure of the CREAM, although the contents of the tables include some elements which are specific to the aviation domain. A shorthand notation has been adopted, whereby each table in the CREAM-AIR is allocated a two letter prefix as shown in Table 6.3.

The general consequents in the table are allocated a numeric identifier in the range 1 to *n*, where *n* is the number of general antecedents in a particular table. Each of the general antecedents is allocated an identifier of the form *Px*, where *P* is the two letter prefix for the relevant table of antecedents, and *x* is a number in the range 1 to *m*, where *m* is the number of general antecedents in that table. Similarly, each of the specific antecedents is allocated an identifier of the form S*y*, where *y* is a number in the range 1 to *p*, and *p* is the number of specific antecedents for a particular general consequent in that table.

The distinction between direct and indirect general antecedents which is present in the generic CREAM of Hollnagel (1998a) is not preserved here; all of the general antecedents—direct and indirect—are treated in the same way, so no detail is lost. The removal of this distinction simplifies the search process, by removing duplicates of general antecedents from the search process.

Coding of actions using the CREAM-AIR is a search process in which the classification scheme provides the framework for describing the structure of all types of incident. The process involves a high degree of judgement, requiring the selection between alternative antecedents for a particular consequent. This selection cannot always be done in a completely deterministic

Table 6.3. CREAM-AIR Prefix Codes.

CREAM Table Title	Prefix
Error Modes	EM
Observation	OB
Interpretation	IN
Planning	PL
Temporary Person	TP
Permanent Person	PP
Equipment Failure	EF
Procedures	PR
Temporary Interface	TI
Permanent Interface	PI
Communication	CO
Organisation	OR
Training	TR
Ambient Conditions	AC
Working Conditions	WC

way, because novel sequences of events will not be present in the table. In many domains, systems evolve over time, so it is possible that the sequence of events which precede a particular incident may also change over time.

The search process is terminated for a particular sub-branch of the tree when either a specific antecedent is found, or when no further antecedents for the current (general) antecedent can be found. In the former case, the last event in the sub-branch of the tree is encoded as a pairing of the form *GG.Sn*, where *GG* is the two letter code corresponding to the category of the last general antecedent in the sub-branch—as shown in Table 6.3—and *Sn* is the code for the specific antecedent taken from the table of specific antecedents corresponding to the last general antecedent in the sub-branch. In the latter case, the last event in the sub-branch of the tree is simply encoded in the form *GGn*, where the meaning of *GG* is the same as before, and the *n* corresponds to the number of the appropriate general antecedent in the table.

The basic process used to encode the individual ASRS incidents using the CREAM-AIR is summarised by the flowchart shown in Figure 6.1. This process is followed for every possible causal sequence that can be identified for a particular incident. The use of the process is illustrated below by means of a simple example.

6.4.5 A worked example of coding an incident using the CREAM-AIR

What follows is a simplified example of the coding of an incident. It is only intended to provide a general indication of how coding can be manually performed using the CREAM-AIR. In reality, all of the coding is performed using a software tool (described in Appendix B).

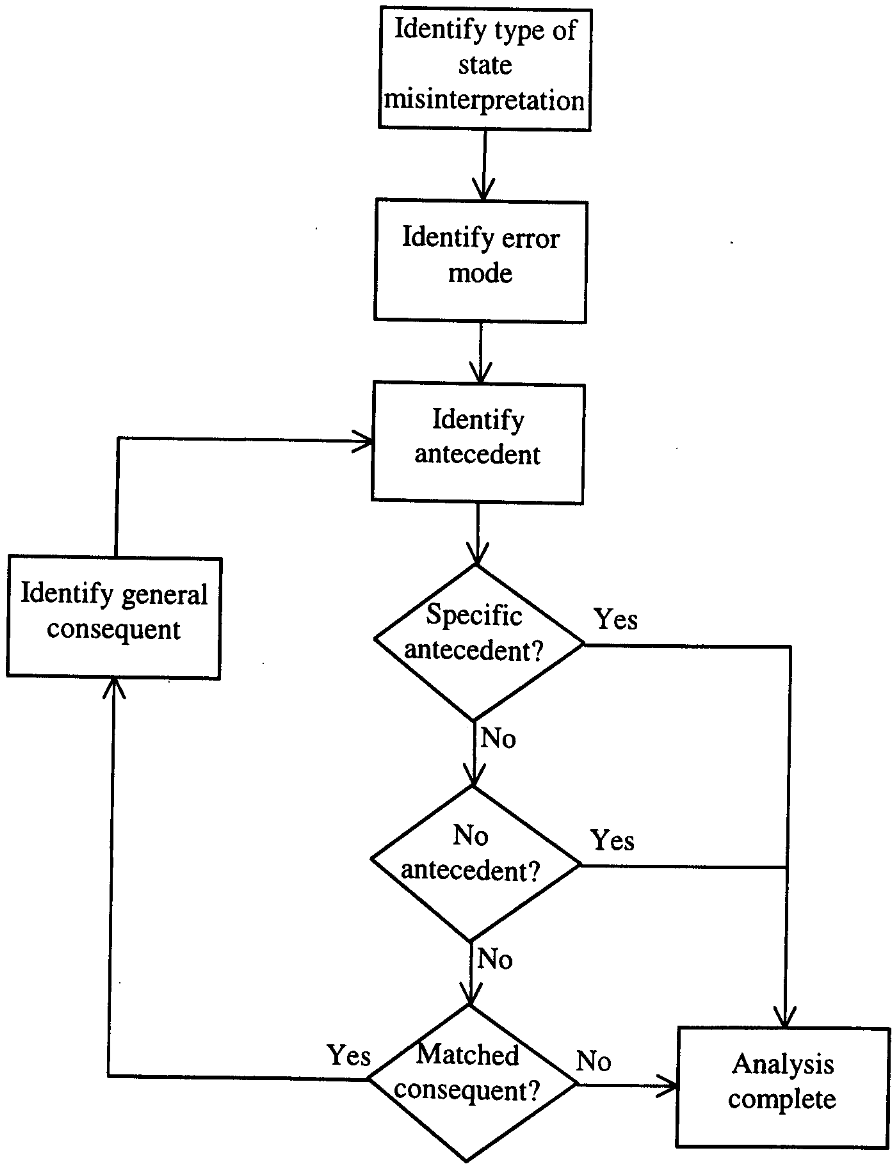

Figure 6.1 General method for encoding ASRS incidents using the CREAM-AIR.

Table 6.4 shows the narrative section of ASRS incident report 326750. The first step in the coding is to identify the type of state misinterpretation. In this case, the pilots appear to have had the implicit belief that they were still short of the point at which they should have started their descent. So, the type of state misinterpretation is *Position in air short of action point*.

The next step is to identify the error mode. In this incident, the report notes that "WE HAD STARTED OUR DSCNT" at the time that the flight crew were asked to begin the descent by ATC. In other words, the crew had started the descent too late, and ATC had noticed this, and

Table 6.4. Textual narrative section for ASRS report 326750. (The report is shown verbatim.)

ON HAR.BUCKS 3 ARR TO PHL OVER JST WE WERE GIVEN A DSCNT TO CROSS COFAX AT FL250, FROM OUR CRUISE OF FL290. WE CALCULATED THE DISTANCE NEEDED TO MAKE THE DSCNT AND THEN BECAME PREOCCUPIED WITH COCKPIT PROCS AND DUTIES. JUST AS WE NOTICED OUR DISTANCE TO THE DSCNT POINT, ATC CALLED TO ASK US TO START OUR DSCNT (THIS WAS ABOUT 36 NM FROM JST AND 45 NM FROM HAR). WE HAD STARTED OUR DSCNT AND AS WE FIGURED THAT WE WOULD MISS OUR XING RESTR BY A MI OR TWO ATC CALLED TO GIVE US A NEW XING RESTR OF 30 NM W OF LRP AT FL180. WITH THIS NEW RESTR WE REDUCED OUR DSCNT AND CONTINUED WITH NO FURTHER INCIDENTS. (Accession Number 326750)

were asking them to expedite the descent. So, the error mode in this case is Error Mode 1: *Action at wrong time: Too late.*

Having identified the error mode, the final step is to find all of its antecedents. The reporter notes that the flight crew "BECAME PREOCCUPIED WITH COCKPIT PROCS AND DUTIES". In other words, the flight crew were distracted from flying the plane. This is an example of *Distraction* which is a temporary person related cause in the CREAM-AIR tables, and has the code: *TP3.*

Distraction is now used as a general consequent in order to search for its antecedents. As noted above, the flight crew were busy doing other tasks, rather than focusing on flying the aircraft and starting the descent. The matching antecedent from the CREAM-AIR tables for this consequent is *Competing Task*, which is a specific antecedent with the CREAM-AIR code of *TP3.S4.*

In this instance, there are no other antecedents that can be found, so the analysis is complete. So, the full causal sequence for this incident is that an action was started at the wrong time because the flight crew were distracted by a competing task, which has the CREAM-AIR sequence *EM1-TP3-TP3.S4.*

6.4.6 The CREAM-AIR coding and fundamental ESDA operations

The encoding process encapsulated in the CREAM-AIR involves several of the fundamental ESDA operations, which are performed in tandem. In particular, the process involves issues of coding, chunking, and constraining. The coding issues are of a general nature, and hence covered by the description of the general ESDA operations (Sanderson & Fisher, 1994), so only chunking and constraining are described here.

During the coding process, causal trees of actions which occurred during the incident are identified. There may be more than one causal tree for each incident. Each of these trees consists of one or more branches. Each of these branches, in turn, consists of one or more sub-branches, and so on. There can be several levels of sub-branches, and each sub-branch can be considered as a chunk in the ESDA sense, in that the elements in each sub-branch are connected together in a causal manner. The nodes—actions and events—in the causal tree are also connected in a temporal sense in that the order of occurrence of the actions or events follows a time line that runs backwards from the roots of the causal trees—the error modes—towards the tips of the branches.

The constraining of the data is simply a selection of part of the data for further processing whilst temporarily excluding the rest of the data. In one sense, the data that is processed using the CREAM-AIR has already been constrained and condensed during transcription. The ASRS incident reports contain a wealth of information, but only the narrative section which describes the incident in the reporter's own words is actually transcribed.

The data is also constrained by categorising it according to the type of state misinterpretation which occurred in each of the incidents. It is also inherently constrained by the CREAM-AIR, because the threads between the actions and events identified during coding, are connected in a causal sequence, rather than just a temporal one. Note however, that a causal sequence is necessarily temporal, in that "*a* causes *b*" means that *a* must occur before *b*, because *b* is the consequent or outcome of doing *a*. The only exception is when *b* is not only a consequent of *a* but can also occur independently of *a*.

Although CREAM-AIR (when used for retrospective analysis) is fundamentally concerned with establishing the possible chains of actions and events linking effects to causes, it is important not to overlook the role played by the other actions and events which occurred within the same episode. For this reason, the sequential nature of the data is preserved, with the cause and effect chains simply being one of the coding schemes that is applied to the data.

6.4.7 Reliability of the coding scheme

A coding scheme is only effective if it can be used reliably by other people. The CREAM-AIR coding scheme is no different in this respect, although it makes particular demands on multi-disciplinary knowledge. In a domain such as aviation, the set of potential actions and events is very large, because the research is concerned with all aspects of pilot behaviour when flying an aircraft. The CREAM-AIR tables, for example, contain over 150 different specific effects and specific consequents.

Although the reports that are stored in the ASRS incident database are available to the public, the database itself cannot be queried on-line. The performing of a database search requires the purchase of the CD-ROM version of the database, or making a request to NASA to perform the search on your behalf.

There are four basic types of knowledge that are required in order to select and retrieve the incident reports from the database, and then encode them using the CREAM-AIR. These are:

- Software tools knowledge: Some knowledge of how databases, and in particular the ASRS database, work is needed in order to retrieve the data in the first instance. In addition some knowledge of how the bespoke software tools work is required, so that the relevant incident reports can be retrieved from the database, before they are analysed.
- Domain knowledge. A reasonable level of knowledge of the aviation domain is required so that the details of the incidents can be understood before they are appropriately encoded. This includes knowledge of the language of Air Traffic Control (e.g. how altitude clearances are communicated to the flight crew), knowledge of how the cockpit instrumentation works (e.g. knowledge of how the instrument landing system, or ILS, works), and knowledge of how the runway system operates at airports (e.g. runway 21 is the same physical runway as runway 3, it is just being approached from the other direction). The reports are written using a combination of English and a shorthand that is specific to aviation (e.g. using "APCH" instead of the word *approach*).
- Human error knowledge. In particular some knowledge of how, why, and under what circumstances people can perform erroneous action.
- CREAM-AIR knowledge: The CREAM-AIR, and the CREAM method on which it is based relies on the idea that incidents (and accidents) are described at the level of observable actions.

Although these different types of knowledge can be acquired or taught, the amount of effort involved needs to be measured in terms of weeks, rather than hours or days. This makes it somewhat difficult to obtain suitable inter-reliability ratings within a reasonable time frame. There are no specific measures available for determining intra-rater reliability, however. In order to obtain some general indication of the reliability of the coding scheme, a sample of 25 reports was recoded approximately 12 months after the data was initially encoded, and an inter-rater reliability measure (Cohen's Kappa) was calculated using the two sets of data, and a figure of 94.5% agreement between the two sets of coding was achieved. This is very much a compromise solution, and given a longer time frame, or the availability of greater local expertise in one or more of the areas of knowledge, an inter-rater reliability would have been performed. The high value of the reliability measure should therefore be considered as an *indication* that the coding scheme can be used consistently. The actual magnitude of the figure must be interpreted with some caution.

6.5 Summary

In this chapter the process of analysing the data has been described. The particular requirements of this research meant that purpose built software tools had to be developed to support the process. The method used to process the data is a development of the CREAM (described in chapter 5), which has been tailored to the retrospective analysis of incident reports from the aviation domain. Each of the incident reports was initially processed to determine the type of state misinterpretation that occurred, and the results were used to develop a taxonomy on state misinterpretation. The error modes for each of the incidents were identified, together with the sequences of actions and events which precede these error modes. The issue of reliability of the coding scheme was also addressed.

The next chapter describes the next stage of the analysis of the sequences of actions and events for each of the incidents. Specifically, the data is analysed at the level of individual actions for each of the types of state misinterpretation.

7 Data Analysis: Concordances

In this chapter the causal data generated in Chapter 6 are analysed to identify the manifestations of the error which occurred, and the attributed causes. For each type of state misinterpretation concordance indexes of the frequencies of occurrence of error modes and attributed causes were automatically generated. The results for each type of state misinterpretation are preceded by a brief description of the sort of task that the flight crew were performing. This is followed by a qualitative prediction of the error modes and attributed causes that could be expected to occur based on a simple analysis of the flight crew's task. The data have been deliberately limited to those instances of state misinterpretation which give rise to a minimum of 20 possible causal sequences of actions, and is presented in the order in which the different types of state misinterpretation could be expected to occur during a flight (from push back to landing). Figure 7.1 shows a typical altitude profile for a flight, with the different flight phases shown in time sequential order.

7.1 Cleared to Push by Ground Crew

When the passengers have boarded the aircraft, the luggage has been loaded and the doors are all locked, the flight crew normally have to wait for the ground crew to signal that it is safe for the aircraft to push (back) from the boarding gate. The reason that one of the ground crew gives the signal is that there may be one or more members of the ground crew who are still underneath the fuselage of the aircraft, out of sight of the flight crew. Communication with the ground crew takes place via an intercom system initially—using a jack socket in the plane's fuselage—and then by means of hand signals when the ground crew disconnects from the jack socket.

Some qualitative predictions can be made about which error modes could be expected during the push back process, and what sort of actions could be expected to be among the attributed causes. The main potential problem is one of starting to push back before the appropriate hand signal

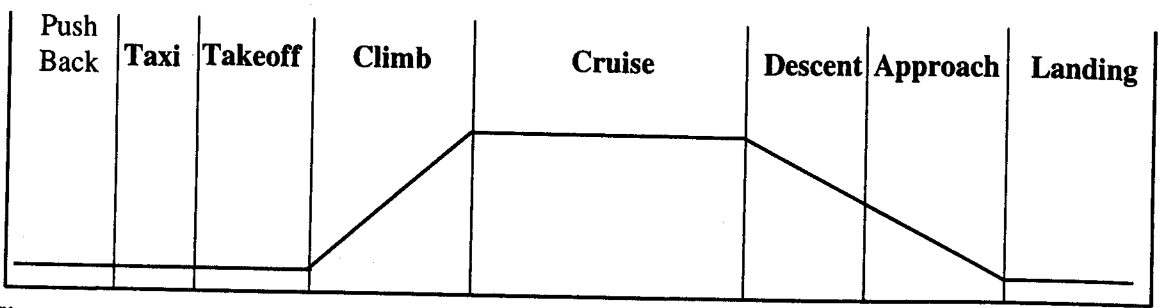

Figure 7.1. Aircraft altitude profile across the phases of a flight. (Not shown to scale.)

has been given by the ground crew. The error mode *Action at wrong time: Timing/duration: too early* (*EM1*) should therefore be expected to feature among the error modes. Similarly, because the problem is a failure of (non-verbal) communication between the flight crew and the ground crew, it should be expected that actions from the CREAM-AIR communication category (prefix *CO*) will feature among the attributed causes for these incidents.

The frequencies of occurrence for the different error modes detected when the flight crew believed that they had been cleared to push by the ground crew are shown in Table 7.1. The results are dominated by *Action at wrong time: Timing/duration (EM1)*, which happened in 20 out of the 21 detected cases (95.2%). In every one of these instances the action occurred *too early*: the flight crew started to push back before getting the proper signal from the ground crew.

The frequencies of occurrence for the antecedents of the error modes shown in Table 7.1 are listed in Table 7.2. The most frequently occurring antecedent is *Communication failure* (*CO1*) which accounted for 11 out of the 43 detected antecedents (25.6%) and for 11 out of the 31 general antecedents (35.4%).

Specific antecedents were identified for 13 of the 21 possible causal sequences (57.1%), which includes one instance where *Cognitive bias* (*PP3*) occurred, for which there are no antecedents defined in the CREAM-AIR. This is an anomaly in the basic CREAM tables, in that there are several instances where a general consequent does not have any antecedents defined. In some cases, this is because it is difficult, if not impossible, to find a meaningful antecedent for consequents such as *Adverse weather* (*AC7*), whilst in others it is because the antecedents have not been included. This shortcoming is acknowledged by Hollnagel (1998a). In order to discriminate between the two, general consequents for which no meaningful antecedents can be defined are treated in the same way as specific antecedents. So, in this instance, *Cognitive bias* is treated like a specific antecedent.

There is no obvious pattern to the occurrences of the specific antecedents: four of them occurred twice, and five occurred just once. Eight of the specific antecedents (38.1% of possible causal sequences) had some sort of communication failure as a root cause, however. This result is in

Table 7.1 Frequency of occurrence of error modes for state misinterpretation type of *Cleared to push by ground crew*.

Error mode	Frequency of occurrence
Action at wrong time: Timing/duration (EM1)	20
Action of wrong type: Distance/magnitude (EM4)	1
Total	21

Table 7.2 Frequency of occurrence of antecedents for state misinterpretation type of *Cleared to push by ground crew*.
(Specific antecedents are shown in italics.)

Code for antecedent	Frequency of occurrence
Communication failure (CO1)	11
Procedure violation (PR2)	4
Distraction (TP3)	3
Competing task (TP3.S4)	2
CRM failure (CO1.S4)	2
Expectation (CO1.S6)	2
Faulty diagnosis (IN1)	2
Observation missed (OB1)	2
Trapping error (EM1.S2)	2
Cognitive bias (PP3)	1
Excessive demand (WC1)	1
Habit or expectancy (OB3.S4)	1
Hearing failure (CO1.S10)	1
Inadequate procedure (PR1)	1
Inadequate quality control (OR2)	1
Inattention (TP6)	1
Lack of information (CO1.S5)	1
Memory failure (TP1)	1
Other priority (TP1.S3)	1
Radio frequency congestion (CO1.S7)	1
Temporary incapacitation (CO1.S3)	1
Wrong identification (OB3)	1
Total	43

line with the general prediction made earlier, and shows that for the state misinterpretation type of *Cleared To Push By Ground Crew* the net outcome is usually that the push back starts before clearance is received from the ground crew.

The analysis of the data for this first type of state misinterpretation gives an initial indication of some of the problems of identifying the system state. The fact that the flight crew have (or have not) been cleared to push back by the ground crew is an inherent part of the system state—it determines when the push back operation can commence—but there is no indicator that the flight crew can consult to determine whether the clearance has been issued. The flight crew are looking for a signal from the ground crew, and problems can arise if the (transient) signal is unclear or ambiguous.

7.2 Position on Ground Short of Action Point

During the taxiing of the aircraft to or from the assigned runway, the flight crew will be told which route to follow by ATC. The directions will often require the flight crew to perform some action at particular points along the route (described here as *action points*). The action point is normally indicated on the ground, usually taking the form of hold short lines—where the aircraft is supposed to halt short of a runway—or taxiway or runway signs at intersections. All of these provide visual cues which can be used to determine the aircraft's current position.

The flight crew may pass the action point without realising it, so it can therefore be predicted that *Action of wrong type: Distance: Too far* (*EM4*) should feature strongly among the error modes for this type of state misinterpretation. In addition, since the flight crew need to be looking outside the aircraft for the visual cues which indicate the location of the action point, it can also be predicted that some of these errors will be attributable to *Observation missed* (*OB1*).

In Table 7.3 the frequencies of occurrence for the different error modes detected when an incident included a state misinterpretation of the type *Position on Ground Short Of Action Point* are shown. The aircraft travelled too far (*Action of wrong type: Distance/magnitude* (*EM4*)) in 36 out of the 37 detected cases (97.3%).

The frequency of occurrence of the antecedents for the error modes shown in Table 7.3 are listed in Table 7.4. The most frequently occurring antecedent is *Observation missed* (*OB1*) which accounted for 17 of the 93 antecedents (18.3%), with *Inadequate workplace layout* (*WC2*) the second most common, occurring eight times (8.6%). If the specific antecedents are ignored, then *Observation missed* accounted for 23.9% of the 71 general antecedents.

In the 37 possible causal sequences, specific antecedents were identified in 25 cases (67.6%). The most common of these root causes were *Competing task* (*TP3.S4*), identified in five of the sequences (13.5%), and *Habit or expectancy* (*OB3.S4*), leading to a *Wrong identification* (*OB3*), with three occurrences (8.1%).

Table 7.3 Frequency of occurrence of error modes for state misinterpretation type of *Position on ground short of action point*.

Error Mode	Frequency of occurrence
Action of wrong type: Distance/magnitude (EM4)	36
Action at wrong time: Timing/Duration (EM1)	1
Total	37

Table 7.4 Frequency of occurrence of antecedents for state misinterpretation type of *Position on ground short of action point.* (Specific antecedents are shown in italics.)

Antecedent (Code)	Frequency of occurrence
Observation missed (OB1)	17
Inadequate workplace layout (WC2)	8
Distraction (TP3)	7
Access problems (PI1)	6
Wrong identification (OB3)	6
Competing task (TP3.S4)	5
Communication failure (CO1)	3
Excessive demand (WC1)	3
Fatigue (TP4)	3
Habit or expectancy (OB3.S4)	3
Mislabelling (PI2)	3
Access limitations (TI1)	2
Adverse weather (AC7)	2
Design failure (OR4)	2
Early morning start (TP4.S2)	2
Faulty diagnosis (IN1)	2
Lack of lighting (TI1.S7)	2
Missing information (CO2)	2
Adverse ambient conditions (AC6)	1
Ambiguous information (TI2)	1
Ambiguous label (EM4.S1)	1
Ambiguous signals (OB3.S1)	1
ATC call to meet time restriction (WC1.S5)	1
Exhaustion (TP4.S1)	1
Habit or expectancy (CO1.S11)	1
Hearing failure (CO1.S10)	1
Inadequate plan (PL1)	1
Inattention (TP6)	1
Misleading symptoms (IN1.S4)	1
Multiple signals (OB1.S2)	1
Overlooking precondition (PL1.S4)	1
Procedure violation (PR2)	1
Short taxi (WC1.S9)	1
Total	93

The general picture for *Position On Ground Short of Action Point* indicates that the aircraft ultimately travels the wrong distance, as predicted. In the cases analysed here the aircraft travelled too far, often because the flight crew missed seeing the hold short lines, or the signs indicating taxiway or runway information. Both of these are instances of *Observation missed*, as previously predicted. In several cases the flight crew also wrongly identified the location of the action point. In three instances this was attributed to *Habit or expectancy*: the flight crew had been anticipating a particular runway or taxiway sign at some point. There were also a total of

14 instances where it was difficult to locate or read the appropriate signs, which were poorly situated.

7.3 Flying Cleared Route

Before any flight, the crew file a flight plan of their intended route with ATC. If the flight plan has to be modified, ATC inform the flight crew as appropriate. Any changes will normally be included when the intended route is discussed as part of the pre-flight briefing. Often, the flight crew will be required to follow published departure routes—Standard Instrument Departures, or SIDs—and arrival routes—Standard Terminal Arrival Routes, or STARs. These SIDS and STARs are described on charts carried by the flight crew.

If the flight plan is altered, this creates the possibility for the flight crew to think they are flying the route they had been cleared to use by ATC, when they are flying the route that they originally filed. Another possible error that may arise is if the flight crew mistakenly follow the wrong SID or STAR; there are often several such routes with similar names all described on one chart. In both cases, the situation can be predicted to lead to an error mode of *Action at wrong object* (*EM7*). One of the attributed causes that can be predicted is that the flight crew fail to check the routing details closely enough, which is an instance of *Observation missed* (*OB1*).

Table 7.5 shows the frequencies of occurrence for the different error modes that occurred when the flight crew mistakenly believed that they were flying the correct route. The frequencies of occurrence are distributed fairly widely distributed: the most common error mode was *Action at wrong time: Timing/duration* (*EM1*; 23 out of 76 detected possible causal sequences, or 30.3% of cases), followed by *Action of wrong type: Direction* (*EM6*; 18 occurrences; 23.7%). In addition there were 14 instances of *Action at wrong object: Wrong object* (*EM7*; 18.4%)—previously predicted—and 13 of *Action of wrong type: Distance/magnitude* (*EM4*; 17.1%).

The frequencies of occurrence for the antecedents of the error modes shown in Table 7.5 are

Table 7.5 Frequency of occurrence of error modes for state misinterpretation type of *Flying cleared route*.

Error Mode	Frequency of occurrence
Action at wrong time: Timing/duration (EM1)	23
Action of wrong type: Direction (EM6)	18
Action at wrong object: Wrong object (EM7)	14
Action of wrong type: Distance/magnitude (EM4)	13
Action in wrong place: Sequence (EM2)	8
Total	76

listed in Table 7.6. The most frequently occurring antecedent is *Observation missed* (OB1; 22 out of 170 antecedents, or 12.9%)—in line with predictions—closely followed by *Inadequate plan* (*PL1*; 20 out of 170; 11.8%) and *Distraction* (*TP3*; 17 out of 170; 10.0%).

If the specific antecedents are ignored, then *Observation missed* accounted for 21.8% of the 101 general antecedents, with *Inadequate plan* accounting for 19.8% and *Distraction* for a further 16.8%.

In 60 of the 76 possible causal sequences (78.9%) a specific antecedent was identified. In 13 of the sequences the specific antecedent was *Competing task* (*TP3.S4*; 17.1%), whilst an *Overlooked precondition* (*PL1.S4*) accounted for 15.8%.

There appear to be two main strands of problem associated with the state misinterpretation type of *Flying Cleared Route*. The first relates to a failure to spot that the aircraft was following the wrong route. The second relates to issues of route planning. Before departure the flight crew normally hold a briefing where they discuss the take off and departure route. Any flaws or omissions in the planned route which are not detected during the briefing will manifest themselves at a later stage. ATC may note that the aircraft is not being flown along the route that the flight crew were expected (and cleared) to follow, for example. Such errors will tend to be detected when the aircraft flies too far in one particular direction, or turns at the wrong point.

This type of state misinterpretation further highlights possible problems in identifying the system state. If the flight plan lodged with ATC prior to departure is not accepted as submitted, any amendments to that plan are only known to the flight crew: there is no system display that shows the current status of the flight plan. In particular, there is no way to determine whether the route that the crew are currently following includes the latest set of changes to the flight plan.

Table 7.6 Frequency of occurrence of antecedents for state misinterpretation type of *Flying cleared route*.
(Specific antecedents are shown in italics.)

Antecedent (Code)	Frequency of occurrence
Observation missed (OB1)	22
Inadequate plan (PL1)	20
Distraction (TP3)	17
Communication failure (CO1)	13
Competing task (TP3.S4)	13
Overlook precondition (PL1.S4)	12
Equipment failure (EF1)	7
Wrong identification (OB3)	6
CRM failure (CO1.S4)	4
Excessive demand (WC1)	4
Fatigue (TP4)	4
Habit or expectancy (CO1.S11)	3
Inadequate quality control (OR2)	3
Wrong reasoning (IN2)	3
Adverse ambient conditions (AC6)	2
Ambiguous symbol set (OB3.S2)	2
Cognitive bias (PP3)	2
Earlier omission (EM1.S1)	2
Exhaustion (TP4.S1)	2
Habit or expectancy (OB3.S4)	2
Inadequate procedure(PR1)	2
Inattention (TP6)	2
Parallel tasks (WC1.S1)	2
Readback failure (CO1.S12)	2
Access limitations (TI1)	1
Ambiguous label (EM7.S1)	1
Ambiguous signals (OB3.S1)	1
Faulty diagnosis (IN1)	1
Glare (TI1.S6)	1
Hearing failure (CO1.S10)	1
Incorrect language (CO2.S2)	1
Information overload (OB1.S1)	1
Irregular working hours (WC4)	1
Memory failure (TP1)	1
Missing information (CO2)	1
Model error (PL1.S3)	1
Overgeneralisation (IN2.S3)	1
Procedure violation (PR2)	1
Rushed pre-flight routine (WC1.S8)	1
Time zone change (WC4.S3)	1
Unexpected tasks (WC1.S2)	1
Violation (PL1.S7)	1
Total	170

7.4 Climbing to Altitude <x>

After take-off, the flight crew will begin to climb to cruise altitude. The clearance to climb will be issued by ATC taking into account the need to ensure the safety of the aircraft in that traffic sector. The net result is that the flight crew will often receive a sequence of clearances, which will ultimately result in the aircraft attaining the cruise altitude. During the climb phase, the flight crew will be busy performing normal actions, such as retracting the flaps from their take-off settings, performing the After Take-off Checklist, and informing the company of the departure information (time, weight and so on). The climb should normally be monitored by the flight crew, and when the aircraft gets to 1,000 feet short of the altitude to which they have been cleared the pilot not flying will often make an announcement to this effect. In addition, the flight crew will generally utilise the altitude alerter—where present—which lights up or sounds a horn when the set altitude is exceeded.

It can be predicted that in several cases the aircraft would fly through the cleared altitude if the flight crew fail to detect that they have reached it, an instance of the error mode *Action of wrong type: Distance: Too far* (*EM4*). It can further be predicted that the failure to level off the aircraft at the required altitude may be due to a monitoring failure, which is an instance of *Observation missed* (*OB1*).

Table 7.7 shows the frequencies of occurrence for the different error modes detected when an incident arose in which the flight crew believed that they were climbing to the altitude to which they had been cleared. The most obvious feature is the marked predominance of occurrence of *Action of wrong type: Distance/magnitude* (*EM4*), predicted above. The aircraft travelled the wrong distance in 43 out of the 50 possible causal sequences (86.0% of cases).

In Table 7.8 the frequencies of occurrence for the antecedents of those error modes shown in Table 7.7 are listed. Four types of antecedent—*Distraction* (*TP3*); *Observation missed* (*OB1*); *Competing task* (*TP3.S4*); and *Equipment failure* (*EF1*)—account for a total of 99 of the 125 detected antecedents (79.2%). If the specific antecedents are ignored, *Distraction* and

Table 7.7 Frequency of occurrence of error modes for state misinterpretation type of *Climbing to altitude <x>*.

Error Modes	Frequency of occurrence
Action of wrong type: Distance/magnitude (EM4)	43
Action at wrong time: Timing/duration (EM1)	6
Action in wrong place: Sequence (EM2)	1
Total	50

Table 7.8 Frequency of occurrence of antecedents for state misinterpretation type of *Climbing to altitude <x>*.
(Specific antecedents are shown in italics.)

Antecedent (Code)	Frequency of occurrence
Distraction (TP3)	33
Observation missed (OB1)	32
Competing task (TP3.S4)	23
Equipment failure (EF1)	11
Communication failure (CO1)	6
Ambiguous label (EM4.S1)	2
CRM failure (CO1.S4)	2
Fatigue (TP4)	2
Inattention (TP6)	2
Adverse Weather (AC7)	1
Changing schedule (WC4.S1)	1
Commotion (TP3.S3)	1
Early morning start (TP4.S2)	1
Exhaustion (TP4.S1)	1
Extraneous conversation (TP3.S6)	1
Flight crew sight-seeing (TP6.S4)	1
Hearing failure (CO1.S10)	1
Irregular working hours (WC4)	1
Priority error (PL2)	1
Procedure violation (PR2)	1
Radio call interruption (TP3.S8)	1
Total	125

Observation missed account for 65 of the 90 general antecedents (72.2%).

A total of 36 specific antecedents were detected in the 50 possible causal sequences (72%). In 23 (46%) of these sequences, the specific antecedent was a *Competing task.*

The overall picture for the possible causal sequences of actions for this type of state misinterpretation is one in which the net result was that the aircraft either climbed or descended to the wrong altitude. This finding is in line with expectations, because if the flight crew do not believe that they have reached their final altitude, they will continue climbing (or descending). The fact that monitoring failures, which are categorised as *Observation Missed* in the CREAM-AIR, feature heavily in the attributed causes is also in line with predictions. The finding can be relatively easily explained: the flight crew did not notice that the altimeter reading indicated that they were either approaching or passing through the cleared altitude.

7.5 Altimeter Setting OK

During the climb phase, the flight crew read the aircraft's current altitude from the altimeter which is calibrated using an associated atmospheric pressure setting. The pressure varies from location to location and hour to hour, so each geographical area has its own standardised setting which all aircraft in that area should use. Using a single pressure setting within a particular area ensures that the altitude readings of all aircraft in that area are calibrated to the same standard. The vertical separation between two aircraft will therefore be the same as the difference in their altitude readings. In addition, there is a standard pressure setting used for all aircraft flying above the altitude known as the transition altitude, which varies depending on the airspace. In the USA, the transition altitude is 18,000 feet; this value applies to the incident reports in the ASRS. Providing all aircraft appropriately reset their altimeter pressure settings when they pass through the transition altitude, the mandated vertical separation—1,000 feet up to Flight Level 290 (an altitude of 29,000 feet)—should be maintained as long as the aircraft fly at their assigned altitudes. If an aircraft has its pressure setting too low, however, its altitude reading will be too high, and hence it will not maintain vertical separation with any aircraft that should be flying at the next flight level (1,000 feet above it).

The way in which the resetting of the altimeter is performed allows some qualitative predictions to be made. The first is that the net outcome will usually be that the aircraft climbs or descends to the wrong altitude. So, the most common error mode should be *Action of wrong type: Distance: Too Far/Too Short* (*EM4*). In addition, it should also be expected that monitoring failures (*Observation failure*; *OB1*), whereby the flight crew did not notice that they were passing through the transition altitude, will feature among the attributed causes.

The frequency of occurrence for the different error modes detected when the flight crew incorrectly believed that the altimeter setting was correct are shown in Table 7.9. The most frequently occurring error mode, as predicted, is *Action of wrong type: Distance/magnitude (EM4)* which occurred in 28 of the 38 possible causal sequences (or 73.7% of cases). In other words, in each of these instances, where one or more members of the flight crew incorrectly believed that the current altimeter setting was correct, the end result was that the aircraft travelled the wrong distance—either too far, or too short. A closer inspections of the incident reports shows that in most cases the incident was classified by NASA as an altitude deviation,

Table 7.9 Frequency of occurrence of error modes for State misinterpretation type of *Altimeter setting OK*

Error Modes	Frequency of occurrence
Action of wrong type: Distance/magnitude (EM4)	28
Action in wrong place: Sequence (EM2)	7
Action at wrong object: Wrong object(EM7)	3
Total	38

and the flight crew returned the aircraft to level flight at an altitude that was different to the one to which it had been cleared by ATC.

The frequencies of occurrence for the various antecedents—both general and specific—for the detected error modes shown in Table 7.9 are listed in Table 7.10. Three antecedents (*Observation missed* (*OB1*), *Distraction* (*TP3*), and *Competing task* (*TP3.S4*)) account for 43 of the 81 listed antecedents (53.1%). If the specific antecedents are ignored, then the general antecedents *Observation missed*, and *Distraction* account for 35 of the 59 general antecedents (59.3%).

There were a total of 38 possible causal sequences detected. Out of these sequences, specific antecedents were identified in 26 cases (68.4% of sequences). The most frequently occurring specific antecedent was *Competing task* which appeared in 8 of the sequences (21.1%).

Table 7.10 Frequency of occurrence of antecedents for State misinterpretation type of *Altimeter setting OK*.
(Specific antecedents are shown in italics.)

Code for Antecedent	Frequency of occurrence
Observation missed (OB1)	20
Distraction (TP3)	15
Competing task (TP3.S4)	8
Communication failure (CO1)	6
Adverse weather (AC7)	4
Inadequate plan (PL1)	4
Fatigue (TP4)	3
Ambiguous label (EM4.S1)	2
CRM failure (CO1.S4)	2
Early morning start (TP4.S2)	2
Overlook precondition (PL1.S4)	2
Access limitations (TI1)	1
ATIS information being changed (CO1.S13)	1
Equipment failure (EF1)	1
Exhaustion (TP4.S1)	1
Extraneous conversation(TP3.S6)	1
Inadequate procedure (PR1)	1
Inattention (TP6)	1
Lack of information (CO1.S5)	1
Lack of lighting (TI1.S7)	1
Memory failure (TP1)	1
Missing Information (CO2)	1
Radio call interruption (TP3.S8)	1
Wrong identification (OB3)	1
Total	81

For those incidents in which the flight crew mistakenly believed that the altimeter setting was correct, the overall picture is in line with predictions. In most cases the aircraft either climbed or descended to the wrong altitude. Although the value shown on the altimeter may have appeared to be correct, the incorrect pressure setting means that the altitude reading would actually be wrong. The most frequently attributed cause was a monitoring failure: the flight crew did not notice that the aircraft was passing through the transition altitude, where the altimeter should have been reset.

7.6 Cleared to Altitude <x>

During the flight, the flight crew will climb (and descend) in accordance with clearances issued by ATC. The process of climbing to the final cruise altitude for the flight may involve several separate clearances to intermediate altitudes. The flight crew will be unaware in advance of these intermediate altitudes so, to make sure that they fully understand the clearances, communication between ATC and flight crews should follow a formal predefined protocol. Initially the air traffic controller will issue the verbal clearance, and the flight crew then reads them back. If the air traffic controller detects any difference between the issued and echoed instructions, the correct instructions are read again to the flight crew.

It can be predicted that the error modes for this type of state misinterpretation will include the aircraft flying to the wrong altitude (*Action of wrong type: Distance: Too far/too short* (*EM4*)). In other words, the flight crew will be climbing or descending to what they believe is the correct altitude. Given that this belief is formed on the basis of communications with ATC, it should also be expected that *communication failure* (*CO1*) will feature among the attributed causes of the incidents.

Table 7.11 shows the frequencies of occurrence for the different error modes detected when an incident included a state misinterpretation of the type *Cleared to Altitude <x>*. The most frequently occurring error mode is *Action of wrong type: Distance/magnitude* (*EM4*), which

Table 7.11 Frequency of occurrence of error modes for state misinterpretation type of *Cleared to altitude <x>*.

Error Modes	Frequency of occurrence
Action of wrong type: Distance/magnitude (EM4)	61
Action at wrong time: Timing/duration (EM1)	37
Action at wrong object: Wrong object (EM7)	8
Action in wrong place: Sequence (EM2)	6
Total	112

accounted for 61 out of the 112 (54.5%) detected cases of this type of state misinterpretation.

In Table 7.12 the frequencies of occurrence for the antecedents of those error modes shown in

Table 7.12 Frequency of occurrence of antecedents for state misinterpretation type of *Cleared to altitude <x>*. (Specific antecedents are shown in italics.)

Antecedent (Code)	Frequency of Occurrence
Communication failure (CO1)	58
Distraction (TP3)	28
Observation missed (OB1)	26
Hearing failure (CO1.S10)	24
Competing task (TP3.S4)	19
Planning failure (PL1)	13
Readback failure (CO1.S12)	9
Fatigue (TP4)	8
Overlook precondition (PL1.S4)	6
CRM failure (CO1.S4)	5
Exhaustion (TP4.S1)	5
Habit or expectancy(CO1.S11)	5
Presentation failure (CO1.S2)	4
Adverse weather (AC7)	3
Excessive demand (WC1)	3
Inattention (TP6)	3
Memory failure (TP1)	3
Trapping error (EM1.S2)	3
Early morning start (TP4.S2)	2
Equipment failure (EF1)	2
Faulty diagnosis (IN1)	2
Flight attendant interruption (TP3.S7)	2
Radio frequency congestion (CO1.S7)	2
Wrong identification (OB3)	2
Access limitations (TI1)	1
Ambiguous information (TI2)	1
Extraneous conversation (TP3.S6)	1
Glare (TI1.S6)	1
Inadequate procedure (PR1)	1
Inadequate quality control (OR2)	1
Incomplete transcription (CO2.S5)	1
Insufficient skills (TR1)	1
Mislabelling (PI2)	1
Missing information (CO2)	1
Other priority (TP1.S3)	1
Parallel tasks (WC1.S1)	1
Procedure violation (PR2)	1
Speech rate (CO1.S8)	1
Temporary incapacitation (CO1.S3)	1
Total	252

Table 7.11 are listed. *Communication failure* (*CO1*) is the most frequently occurring antecedent, accounting for 58 out of 252 detected antecedents (23.0%). In combination with *Distraction* (*TP3*), *Observation missed* (*OB1*), *Hearing failure* (*CO1.S10*), *Competing task* (*TP3.S4*) and *Planning failure* (*PL1.S4*), it accounted for 168 antecedents (66.7%). If the specific antecedents are ignored, then *Communication failure*, *Distraction*, and *Observation missed* accounted for 112 out of 154 general antecedents (72.7%).

In all, a specific antecedent was identified in 98 of the 112 possible causal sequences (87.5%). The specific antecedents are dominated by *Hearing failure* which occurred 24 times (21.4%).

When a state misinterpretation of the type *Cleared To Altitude <x>* occurs, the resulting problems are often due to a communication breakdown. This is largely understandable, since the flight crew are dependent on instructions from ATC to clear them to change altitude. Any confusion arising during communication between the flight crew and ATC could lead to the flight crew believing that they have been cleared to change to a particular altitude which is different from that specified by ATC.

The analysis of incidents involving this type of state misinterpretation reiterates earlier problems of identifying precisely what constitutes system state. For the most part the state of system is reflected by the readings shown on the various instruments in the aircraft cockpit. In addition, however, the flight crew have to keep track of other information, such as verbal instructions issued by ATC, which are also an inherent part of the state of the system. Normally the aircraft should only change altitude when cleared to do so by ATC, which means that the flight crew have to keep track of the fact that they have been cleared to change altitude, and the altitude to which they have been cleared.

7.7 Position in Air Short of Action Point

In some instances, the flight crew are pre-cleared by ATC to perform various actions, such as climbing or descending to a specified altitude, or turn to a new heading. In other words, they are given the clearance before the aircraft reaches the point where the action must be performed. The flight crew therefore need to remember the details of the clearance, including the point at which the action has to be carried out, because this information may not be visible from the cockpit instrumentation. Given enough advance warning, and assuming that the aircraft is equipped with a FMS, the flight crew may be able to reprogram it such that the action can automatically be executed at the appropriate point.

The main difference between this type of state misinterpretation and *Position On Ground Short of Action Point* is to do with visual cues. On the ground the flight crew can look out of the cockpit window and see other traffic, runway signs, hold short lines and so on. In the air, such visual cues are not available, so the flight crew have to rely on the cockpit instrumentation and radio navigation aids to determine the aircraft's position.

If the flight crew are particularly busy performing other routine tasks, there is a possibility that they will forget to commence the action at the appropriate time. In other words, it can be predicted that among the error modes for this type of state misinterpretation will be *Action at wrong time: Too late* (*EM1*). Since the reason for not starting the action is that the flight crew were preoccupied with performing other duties it can also be predicted that *Distractions* (*TP3*) by *Competing tasks* (*TP3.S4*) will both feature among the attributed causes of the error modes.

In Table 7.13 the frequencies of occurrence for the error modes detected when an incident happened in which the pilots mistakenly believed that they were short of the action point are shown. The most commonly occurring error mode is *Action at wrong time: Timing/duration* (*EM1*). An action occurred at the wrong time or for the wrong duration, in 18 out of the 23 detected cases (78.3% of cases).

The figures for the frequency of occurrence for the antecedents of the error modes shown in Table 7.13 are listed in Table 7.14. The most frequently occurring general antecedent is *Distraction* (*TP3*; 10 out of 51 antecedents; or 19.6%), as predicted, followed by *Observation missed* (*OB1*; 7 out of 51 antecedents; 13.7%). If the specific antecedents are ignored, then distractions accounted for 10 of the 34 general antecedents (29.4%), whilst missed observations accounted for 7 (20.6%) of the general antecedents.

Of the 23 possible causal sequences, a specific antecedent was identified in 18 cases (78.3%). The most commonly occurring specific antecedent was *Competing task* (*TP3.S4*) which was detected in 5 of the sequences (21.7%).

Table 7.13 Frequency of occurrence of error modes for state misinterpretation type of *Position in air short of action point.*

Error Mode	Frequency of occurrence
Action at wrong time: Timing/duration (EM1)	18
Action of wrong type: Distance/magnitude (EM4)	3
Action in wrong place: Sequence (EM2)	2
Total	23

Table 7.14 Frequency of occurrence of antecedents for state misinterpretation type of *Position in air short of action point.* (Specific antecedents are shown in italics.)

Antecedent (Code)	Frequency of occurrence
Distraction (TP3)	10
Observation missed (OB1)	7
Competing task (TP3.S4)	5
Inadequate plan (PL1)	3
Fatigue (TP4)	3
Communication failure (CO1)	2
CRM failure (CO1.S4)	2
Equipment failure (EF1)	2
Memory failure (TP1)	2
Extraneous conversation (TP3.S6)	2
Exhaustion (TP4.S1)	2
Adverse weather (AC7)	1
Missing information (CO2)	1
Expectation (CO2.S6)	1
Multiple signals (OB1.S2)	1
Overlooking precondition (PL1.S4)	1
Other priority (TP1.S3)	1
Flight attendant interruption (TP3.S7)	1
Inattention (TP6)	1
Insufficient knowledge (TR2)	1
Excessive demand (WC1)	1
Parallel tasks (WC1.S1)	1
Total	51

When an incident involves a state misinterpretation of the type *Position In Air Short Of Action Point* the net result is often a failure to start executing an action in time to complete it before reaching some target location. In several of the incidents, the problem is actually detected by ATC, when they notice that the aircraft has failed to commence execution of the appropriate action. They report the problem to the flight crew, who then take corrective action in order to alleviate the problem.

The analysis of this type of state misinterpretation highlights another example of the difficulties of trying to track the system state correctly. There are three particularly relevant points worth mentioning at this juncture. The first is that the need to initiate an action at some point in time or space is part of the system state. The second is that there may not be—depending on the cockpit equipment—any way of representing the clearance on the cockpit instrumentation. The third, which is a corollary of the second, is that the flight crew may have to remember the details of the clearance in their heads, which means that the information will be subject to human memory limitations.

7.8 Radio Navigation Frequency Setting OK

Over the course of a flight the flight crew will follow a route which requires them to make contact with various radio navigation beacons at different times in order to plot their position. Correct navigation depends on the timely retuning of the cockpit radio navigation system (RNAV) to the appropriate beacons. If the flight crew tune into a VHF Omni-directional Radio Range beacon (VOR), for example, bearing information is displayed on the main compass system of the aircraft. If the VOR is co-located with a Distance Measuring Equipment (DME; the combination is called a VOR/DME), as is often the case, the flight crew can also utilise the distance information supplied by the DME. The navigation radios, however, only show the frequency to which they are tuned, rather than the name of the beacon to which that frequency belongs.

There are two types of erroneous action that can be predicted here. The first is that the flight crew could tune the navigation radio to the wrong frequency (*Action at wrong object: Similar object* (*EM7*)), and the second is that the flight crew could fail to switch radio frequencies (*Action in wrong place: Sequence* (*EM2*)). It is also possible to predict that monitoring failures (*Observation missed: OB1*) will feature among the attributed causes, reflecting those instances where the flight crew did not notice that they were navigating using the wrong radio navigation station.

The frequencies of occurrence for the different error modes detected for this type of state misinterpretation are shown in Table 7.15. The detected error modes are relatively evenly distributed across *Action at wrong object: Similar object* (*EM7*; 11 out of 33; 33.3%), *Action in wrong place: Sequence* (*EM2*; 9 out of 33 possible causal sequences or 27.3% of cases) and *Action of wrong type: Direction* (*EM6*; 7 out of 33; 21.2%).

Table 7.16 lists the frequencies of occurrence for the antecedents of the error modes shown in Table 7.15. The most commonly occurring antecedent is *Observation missed* (*OB1*) which

Table 7.15 Frequency of occurrence of error modes for state misinterpretation type of *RNAV frequency setting OK.*

Error Mode	Frequency of occurrence
Action at wrong object: Similar object (EM7)	11
Action in wrong place: Sequence (EM2)	9
Action of wrong type: Direction (EM6)	7
Action at wrong time: Timing/duration (EM1)	5
Action of wrong type: Distance (EM4)	1
Total	33

Table 7.16 Frequency of occurrence of antecedents for state misinterpretation type of *RNAV frequency setting OK*.
(Specific antecedents are shown in italics.)

Antecedent (Code)	Frequency of occurrence
Observation missed (OB1)	15
Distraction (TP3)	14
Competing task (TP3.S4)	7
Communication failure (CO1)	5
Wrong Identification (OB3)	4
Extraneous conversation (TP3.S6)	3
Fatigue (TP4)	3
CRM failure (CO1.S4)	2
Equipment failure (EF1)	2
Erroneous information (OB3.S3)	2
Faulty diagnosis (IN1)	2
Inadequate plan (PL1)	2
Access limitations (TI1)	1
Access problems (PI2)	1
Ambiguous label (EM6.S1)	1
Exhaustion (TP4.S1)	1
Flight attendant interruption (TP3.S7)	1
Habit or expectancy (OB3.S4)	1
Hidden information (CO2.S1)	1
Inadequate quality control (OR2)	1
Inattention (TP6)	1
Irregular working hours (WC4)	1
Lack of information (CO1.S5)	1
Maintenance failure (OR1)	1
Memory failure (TP1)	1
Missing information (CO2)	1
Obstruction (TI1.S4)	1
Overlooked precondition (PL1.S4)	1
Presentation failure (CO1.S2)	1
Radio call interruption (TP3.S8)	1
Time zone change (WC4.S3)	1
Total	80

accounted for 15 of the 80 antecedents (18.8% of all antecedents), whilst *Distraction* (*TP3*) accounted for 14 of the antecedents (17.5%). If the specific antecedents are ignored, then missed observations accounted for 27.3% of the 55 general antecedents, whilst distractions accounted for 25.5%.

Specific antecedents were identified in 25 of the 37 possible causal sequences (75.8% of sequences). The most frequently occurring specific antecedent was *Competing task* (*TP3.S4*) which was detected in 7 of the sequences (21.2%).

The general picture for the state misinterpretation type of *RNAV Frequency Setting OK* suggests that this type of state misinterpretation can lead to several different types of erroneous action. There are three main reasons why the navigation radio frequency settings can be wrong. The first is that the radios are not retuned at the appropriate point. The second is that the radios are retuned to the wrong frequency, which means that the aircraft is using the wrong radio navigation station. The third is that the radios are retuned too soon, which means the aircraft is using the right radio navigation station but at the wrong time.

Given that *Action of wrong type: Direction* (*EM6*), and *Action at wrong time: Timing/duration* (*EM1*) make up 40% of the error modes it is worthwhile considering why neither were predicted. The predictions that are made are all qualitative predictions based on a general view of the task(s) being performed by the flight crew. For this type of state misinterpretation the error modes really have two effects, the first is on the instrumentation—in this case the RNAV—whilst the second is on the aircraft itself. The former can usually be identified by a coarse informal task analysis of the flight crew's actions. The latter, however, require a more detailed task analysis which takes into account global effects on the aircraft. A more detailed, formalised task analysis may therefore have indicated the possibility that the two error modes could be predicted, although it still would not have enabled any quantitative predictions to be made about how often they would occur.

7.9 Descending To Cleared Altitude <x>

As the aircraft approaches its destination, the flight crew will be cleared to descend the aircraft to a lower altitude as part of the preparation for landing. The descent may have to be conducted in a number of steps depending on local traffic densities. Often, the flight crew will make use of the cockpit equipment to keep track of the latest cleared altitude. The altitude alerter can be set to sound an alarm or light up a warning lamp as the aircraft approaches the desired altitude, or the aircraft's autopilot can be set to automatically level off at the desired height.

Table 7.17 Frequency of occurrence of error modes for state misinterpretation type of *Descending to cleared altitude <x>*.

Error Mode	Frequency of occurrence
Action of wrong type: Distance/magnitude (EM4)	55
Action in wrong place: Sequence (EM2)	7
Action at wrong time: Timing/duration (EM1)	5
Total	67

Based on the preceding description, it can be predicted that one type of erroneous action which is likely to feature among the reported incidents is where the aircraft simply descends through the cleared altitude (*Action of wrong type: Distance: Too Far* (*EM4*)). It is also possible to predict that among the attributed causes will be the fact that the flight crew did not spot that the aircraft had reached its cleared altitude (*Observation missed, OB1*). In addition, because the descent phase of any flight is often a busy time for the flight crew, the possibility of the flight crew being distracted by any additional tasks which arise (*Competing Task; TP3.S4*) is a very real one.

The frequencies of occurrence for the different error modes detected in those incidents where the flight crew mistakenly believed that they were descending to the altitude to which they had been cleared by ATC are shown in Table 7.17. The error mode that was detected most frequently is *Action of wrong type: Distance/magnitude* (*EM4*), with the aircraft travelling the wrong distance in 55 out of the 67 reported incidents (82.1%).

The frequencies of occurrence for the antecedents of the error modes shown in Table 7.17 are listed in Table 7.18. The most frequently occurring antecedent is *Observation missed* (48 out of 176 antecedents, or 27.2%), which is closely followed by *Distraction* (45 out of 176, or 25.6%). Together with *Competing task*, these antecedents accounted for 116 antecedents (65.9%). If the specific antecedents are ignored, then *Observation missed* accounted for 36.9% of the 130 general antecedents, with *Distraction* accounting for another 34.6% of them.

In the 67 possible causal sequences, specific antecedents were detected 46 times (68.7% of sequences). *Competing task* (*TP3.S4*) occurred the most times, arising in 23 of the sequences (34.3%).

The overall picture for the state misinterpretation type of *Descending To Cleared To Altitude* <*x*> shows that the net result is most often an altitude deviation. The flight crew often failed to spot that they were approaching, or at the cleared altitude, and in many cases were distracted, and busy with other tasks that were competing for their attention.

Table 7.18 Frequency of occurrence of antecedents for state misinterpretation type of *Descending to cleared altitude <x>*.
(Specific antecedents are shown in italics.)

Antecedent (Code)	Frequency of occurrence
Observation missed (OB1)	48
Distraction (TP3)	45
Competing task (TP3.S4)	23
Adverse weather (AC7)	9
Communication failure (CO1)	6
Fatigue (TP4)	5
Radio call interruption (TP3.S8)	5
Excessive demand (WC1)	4
Exhaustion (TP4.S1)	3
CRM failure (CO1.S4)	2
Equipment failure (EF1)	2
Extraneous conversation (TP3.S6)	2
Flight attendant interruption (TP3.S7)	2
Inadequate plan (PL1)	2
Inattention (TP6)	2
Insufficient skills (TR1)	2
Parallel tasks (WC1.S1)	2
Unexpected tasks (WC1.S2)	2
Access problems (PI1)	1
Ambiguous signals (OB3.S1)	1
Changing schedule (WC4.S1)	1
Hearing failure (CO1.S10)	1
Insufficient knowledge (TR2)	1
Irregular working hours (WC4)	1
Overlook precondition (PL1.S4)	1
Procedure violation (PR2)	1
Readback failure (CO1.S12)	1
Wrong identification (OB3)	1
Total	161

7.10 Rate of Descent OK

During the descent phase, as the aircraft nears the destination airport, it will need to descend in accordance with clearances issued by ATC. Often these clearances will require that the aircraft be at a particular altitude by the time it reaches a specified location. In other words, the flight crew will have a crossing restriction to meet. The rate of descent required to meet the crossing restriction is left to the flight crew, and is calculated taking into account horizontal and vertical distances involved, and fuel consumption rates. In addition, the flight crew will calculate the optimum point at which to start the descent in order to make the crossing restriction.

Two types of erroneous actions can be predicted in incidents involving this type of state misinterpretation. The first is where the flight crew fly past the point where they should have started the descent (*Action of wrong type: Distance: Too far* (*EM4*)), and the second is where the flight crew start the descent too late to make the crossing restriction (*Action at wrong time: Timing/duration: Too late* (*EM1*)). It can be further predicted the attributed causes for the detected error modes will include monitoring failures (*Observation missed; OB1*), whereby the flight crew fail to detect that the aircraft has reached the point at which the descent was supposed to commence.

The frequencies of occurrence for the different error modes for this type of state misinterpretations are shown in Table 7.19. There is a fairly even distribution of the reported incidents between three detected error modes: *Action at wrong time: Timing/duration* (*EM1*; 7 out of the 20 detected possible causal sequences or 35.0% of cases), *Action of wrong type: Speed* (*EM5*; 6 out of 20, or 30.0%), and *Action of wrong type: Distance/magnitude* (*EM4*; 5 out of 20, or 25.0%).

In Table 7.20 the frequencies of occurrence for the antecedents of the detected error modes shown in Table 7.19 are listed. The antecedent which occurs most frequently is *Observation missed* (*OB1*) which accounted for 14 out of 51 antecedents, (27.5% of all antecedents) whilst *Distraction* (*TP3*) accounted for 12 of the antecedents (23.5%). If the specific antecedents are ignored, then *Observation missed* accounted for 36.8% of the 38 general antecedents, whilst *Distraction* accounted for a further 31.6%.

Specific antecedents were detected in 16 of the 20 possible causal sequences (80.0%). *Competing task* (*TP3.S4*) was the most common specific antecedent, appearing in 7 of the sequences (35.0%).

The results for the state misinterpretation type of Rate of Descent OK present a somewhat mixed picture. There appear to be three separate manifestations which follow such a state

Table 7.19 Frequency of occurrence of error modes for state misinterpretation type of *Rate of descent OK*.

Error Mode	Frequency of occurrence
Action at wrong time: Timing/duration (EM1)	7
Action of wrong type: Speed (EM5)	6
Action of wrong type: Distance/magnitude (EM4)	5
Action at wrong object: Wrong object (EM7)	2
Total	20

Table 7.20 Frequency of occurrence of antecedents for state misinterpretation type of *Rate of descent OK*.
(Specific antecedents are shown in italics.)

Antecedent (Code)	Frequency of occurrence
Observation missed (OB1)	14
Distraction (TP3)	12
Competing task (TP3.S4)	7
Equipment failure (EF1)	2
Faulty diagnosis (IN1)	2
Inadequate plan (PL1)	2
Flight attendant interruption (TP3.S7)	2
Fatigue (TP4)	2
Exhaustion (TP4.S1)	2
Adverse weather (AC7)	1
Communication failure (CO1)	1
Overlooked precondition (PL1.S4)	1
Cognitive style (PP2)	1
Cognitive bias (PP3)	1
Extraneous conversation (TP3.S6)	1
Total	51

misinterpretation. The first is where the flight crew did not start the descent until it was too late. The second is where the rate of descent of the aircraft is too slow to meet the imposed restriction. The flight crew may have to make some manual calculations to determine the appropriate rate of descent, start point for the descent, and where a time restriction is involved, the speed required to make a restriction imposed by ATC. The third is where the aircraft is prematurely descended through the altitude of the imposed crossing restriction.

The need to start the descent of the aircraft at a particular point is another example of an element of the system state which is not adequately represented by the cockpit instrumentation. Although it may be possible to program the flight management system to assist in the process, it depends on there being sufficient time available to do the programming. So this type of state misinterpretation is another one in which the system state relies on the flight crew's ability to remember information which is not readily available via the cockpit instruments.

7.11 Cleared to Land

As the aircraft approaches the destination airport, the flight crew generally become preoccupied with the tasks associated with preparing the aircraft for landing. During this period they will also be told by the approach controller to retune their communication radios to the frequency of the tower controller at the destination airport. It is the tower controller that issues the clearance to land the aircraft at the airport.

It can be predicted that the erroneous action that is most likely to occur is one in which the flight crew fail to retune the communication radio to the tower control frequency (*Action in wrong place: Sequence; EM2*). In addition, since this phase of the flight is invariably busy, it can be predicted that one of the attributed causes for failing to change frequency is that the flight crew are distracted by the need to perform other tasks at the point where they should retune their radios. It should therefore be expected that, *Distraction* (*TP3*) and *Competing task* (*TP3.S4*) should feature prominently among the attributed causes.

Table 7.21 shows the frequencies of occurrence for the different error modes detected when an incident included a state misinterpretation of *Cleared to Land*. The frequency of occurrence of the detected error modes is dominated by *Action in wrong place: Sequence* (*EM2*). An action occurred out of sequence (either in the wrong place, or was omitted) in 31 out of the 36 detected cases (88.9%).

The figures for the frequency of occurrence for the antecedents of the error modes shown in Table 7.21 are listed in Table 7.22. The most frequently occurring antecedents were *Distraction* (*TP3*), *Observation missed* (*OB1*), and *Competing task* (*TP3.S4*) which between them accounted for 58 out of the 92 detected antecedents (63.0%). If the specific antecedents (shown in italics) are ignored, then *Distraction*, and *Observation missed* accounted for 41 out of the 63 general antecedents (65.1%).

In 29 of the 36 possible causal sequences (80.1%) a specific antecedent was detected. *Competing task* was the most commonly occurring specific antecedent, arising in 17 of the sequences (47.2%).

In general, when a state misinterpretation of the type *Cleared To Land* occurs the net result is that an action occurs out of sequence. On closer inspection of the content of the incident reports, what often happens is that the flight crew simply do not make contact with ATC to obtain clearance to land, for a variety of reasons. The sequence of operations for landing generally involves the flight crew making contact with different air traffic controllers, each of

Table 7.21 Frequency of occurrence of error modes for state misinterpretation type of *Cleared to land*.

Error Modes	Frequency of Occurrence
Action in wrong place: Sequence (EM2)	31
Action at wrong time: Timing/Duration (EM1)	4
Action at wrong object: Wrong object (EM7)	1
Total	36

Table 7.22 Frequency of occurrence of antecedents for state misinterpretation type of *Cleared to land*.
(Specific antecedents are shown in italics.)

Code for antecedent	Frequency of occurrence
Distraction (TP3)	22
Observation missed (OB1)	19
Competing task (TP3.S4)	17
Communication failure (CO1)	4
Excessive demand (WC1)	4
Adverse weather (AC7)	3
Fatigue (TP4)	3
Parallel tasks (WC1.S1)	3
Wrong identification (OB3)	2
Exhaustion (TP4.S1)	2
Performance variability (TP5)	2
CRM failure (CO1.S4)	1
Radio frequency congestion (CO1.S7)	1
Missing information (CO2)	1
Noise (CO2.S3)	1
Earlier omission (EM1.S1)	1
Information overload (OB1.S1)	1
Inadequate plan (PL1)	1
Overlook precondition (PL1.S4)	1
Inattention (TP6)	1
Insufficient skills (TR1)	1
Unexpected tasks (WC1.S2)	1
Total	92

which involves a change of radio frequency. In some cases the change of frequency is tied in with a reference point on the approach to the airport. The flight crew therefore have to remember the reference point and the radio frequency in addition to carrying out all the other duties associated with preparing the aircraft for landing.

This type of state misinterpretation further highlights some of the problems of integrating verbal information received from ATC into the system state. There is no obvious indication that the flight crew can consult to determine whether they have been cleared to land or not, although some crews improvise by turning on a particular lamp which is not normally used at this point in the flight to act as a reminder that they have been cleared to land.

7.12 General Discussion

One of the problems with state misinterpretation is that it is not a directly observable phenomenon. It usually has to be inferred from the actions described in the narrative sections of the incident reports. The frequency of occurrence of error modes and actions for particular instances of the various types of state misinterpretation provides one method—albeit a static one—of characterising state misinterpretation.

The distribution of error modes across the different types of state misinterpretation is shown in Table 7.23. The main feature of the table is the dominance of *Error Mode 4* (*Action of wrong type: Distance/magnitude*) which is the most frequently occurring error mode in five out of the 11 types of state misinterpretation shown, and accounts for about half of the total number of error modes. Note that this is not quite the same as half the number of incidents, since there may be more than one error mode for some of the incidents. The relative dominance of *Error Mode 4* is in line with other findings which have shown for many years that altitude deviations—the aircraft fails to level off at the cleared altitude—form the main type of anomaly in ASRS incidents (e.g. Thomas & Rosenthal, 1982).

Out of the 11 different types of state misinterpretation analysed, eight are dominated by one type of error mode. In each of these cases, the dominant error mode accounts for more than two thirds of the total error modes for that type of state misinterpretation. For the other three types of state misinterpretation, the distribution is fairly evenly spread across two or three different error modes.

In general the results for the occurrences of error modes and antecedents were in line with the qualitative predictions that were made. There are two basic reasons for the correctness of the predictions. The first is the fact that the predictions are qualitative, rather than quantitative, and so deal in generalisations about which actions and error modes could be expected to appear frequently. The second is the fact that aviation is a highly proceduralised domain, which means that most tasks are composed of highly predictable sequences of actions, and it is relatively simple to identify which problems are likely to occur.

Part of the data analysis involves trying to find the root causes (specific antecedents) for the possible causal sequences that are generated for each of the incidents. There are some root causes, however, that are difficult, if not impossible, to prevent, such as adverse weather conditions which were identified in 22 of the reported incidents. Since root causes like these cannot be prevented, it becomes particularly important for the flight crew to manage the ensuing situation appropriately to mitigate any possible adverse consequences.

Table 7.23 Summary of Frequency of Occurrence of Error Modes For Each Type of State Misinterpretation

	Altimeter Setting OK	Climbing To Altitude <x>	Cleared To Altitude <x>	Cleared To Land	Cleared To Push By Ground Crew	Descending To Cleared Altitude <x>	Flying Cleared Route	Position In Air Short Of Action Point	Position On Ground Short of Action Point	Rate Of Descent OK	RNAV Frequency Setting OK	Total
Error Mode 1 (Timing/duration)		6	37	4	20	5	23	18	1	7	5	126
Error Mode 2 (Sequence)	7	1	6	31		7	8	2			9	71
Error Mode 3 (Force)												0
Error Mode 4 (Distance/magnitude)	28	43	61		1	55	13	3	36	5	1	246
Error Mode 5 (Speed)										6		6
Error Mode 6 (Direction)							18				7	25
Error Mode 7 (Wrong object)	3		8	1			14			2	11	39
	38	50	112	36	21	67	76	23	37	20	33	513

The same strategy of situation management also needs to be adopted for those detectable antecedents which the flight crew may be unable to prevent, such as some kinds of equipment failures. In incidents involving adverse weather conditions or equipment failures, part of the problem was that the flight crew devoted too many resources to addressing the situation, rather than first ensuring that enough resources were allocated to continuing to fly the aircraft.

Perhaps the most interesting feature to emerge from the data analysis is the difficulty in defining the system state. Although the cockpit instrumentation provides a fairly comprehensive view of the state of the aircraft, it does not adequately cover those aspects of the system that are provided by ATC. In particular, it does not provide a simple way of accessing the verbal clearances that were supplied by ATC. The general acceptance of the need to consider all parts of the system, including organisational aspects (e.g. Reason, 1997), and cultural influences (e.g. Helmreich & Merritt, 1998) makes it more difficult to completely define the system state. It therefore becomes more and more difficult to fully appreciate all the facts that the flight crew may be taking into account at any particular point in time whilst they are flying the aircraft.

7.13 Summary

The manifestation of erroneous actions are the last element in the possible causal sequence of actions for an incident. Although the manifestations are important, as is the recurrence of individual actions, it is the dynamic order of the of actions that give rise to the incident which is particularly important. One of the recognised ways of preventing the erroneous action from occurring at the end of the sequence is to prevent the sequence from running to completion. The way in which the actions are sequentially linked together for a particular incident can be determined using the CREAM-AIR. Once the sequences are identified, consideration can be given to how to break the sequences. The next chapter presents the results of the analysis of the dynamics of each of the different types of state misinterpretation, which is based on identifying the sequences of antecedents for each of the incidents.

8 Data Analysis: Sequences

In Chapter 7 the different types of state misinterpretation were analysed at the level of error modes and antecedents. The analysis essentially presented a way of statically characterising the different types of state misinterpretation. It is not practicable to use the results of such an analysis as the basis for preventing the occurrence of the various types of incidents, however, because any prevention mechanisms would be based on providing a defence against the individual error modes and antecedents. A more practicable method is to consider the various sequences of actions that lead up to the erroneous action (error mode), and try to identify ways in which the sequences can be broken, to prevent the erroneous action from occurring.

In this chapter the causal sequences for each of the different types of state misinterpretation are analysed. Strictly speaking these are the *possible* causal sequences, because there may be more than one possible explanation for a particular incident. Such a situation simply highlights that there may be more than one cause—more than one sequence of antecedents—for a particular incident.

Flying an aircraft is a highly procedural activity, with each flight consisting of several phases (take-off, climbing, cruising, descending and so on). Each of these phases has an associated set of tasks, and each task consists of a sequence of actions. The phases and tasks are broadly the same across different flights, so the sequence of actions should be similar for each of the tasks at some level of abstraction of behaviour. Given the close relationship between correct and erroneous behaviour, it should be expected that there would be similarities between the sequences of actions during which the state misinterpretation occurs. This expectation can be confirmed by analysing the sequences of actions to look for common (sub-)sequences of actions.

The sequences of actions and events were generated by analysing each of the incident reports individually using the CREAM-AIR to automatically create the coded sequences. Each sequence is normally only counted once when calculating the frequency of occurrence, as illustrated by the following hypothetical (but typical) example.

Suppose that an incident occurs when an aircraft travels too far. The error mode involved is an instance of the CREAM-AIR category *Action of Wrong Type* with the general effect of *Distance too far* (which has the code *EM4*). Assume that the analysis of the incident showed that this error mode arose when the flight crew missed the fact that the plane was approaching the desired altitude (an instance of the general consequent *Observation Missed* category in the *Observation* category, which has the code *OB1*), which was due to the flight crew being distracted

(*Temporary Person: Distraction*; code *TP3*) by the need to look for other traffic in the area (an instance of the Specific Antecedent *Competing Task*, which has the code *TP3.S4*). The coded sequence of actions and events for this incident is *EM4-OB1-TP3-TP3.S4*. The sub-sequences of that particular instance of the sequence (i.e. *EM4-OB1-TP3*, *EM4-OB1*, *OB1-TP3-TP3.S4* and so on) are not counted separately except where they occur in other sequences. So, for example, if there was also a sequence of actions and events *EM4-OB1-TP3-TP3.S6* for the same incident, where the error mode (*EM4*), the missed observation (*OB1*), and the distraction (*TP3*) were all the same as before, then this full sequence would be counted. If the sequences of error modes and antecedents are considered as a tree, where the error mode is the root node, then the set of possible causal sequences is simply the set of all the possible paths from the root node to the leaf nodes of that tree.

The results of the data analysis are presented below, using the same ordering of sections that was used in Chapter 7. The data is therefore organised using the same types of state misinterpretation, with the individual sections ordered according to the phase of the flight during which they could be expected to occur: those which are likely to occur earlier in a flight come first. The data has been deliberately limited to those instances of state misinterpretation which give rise to a minimum of 20 possible explanatory causal sequences of actions.

8.1 Cleared to Push by Ground Crew

Once all the passengers are on board the aircraft, and the doors have all been closed and locked, the flight is almost ready to depart from the boarding gate. The final act is for the ground crew to wave-off the flight crew once the all of the ground crew are clear of the aircraft. The wave-off clears the aircraft to push back from the gate.

The frequency of occurrence for sequences of actions that were detected in incidents in which the flight crew believed that they had been cleared to push back from the boarding gate by the ground crew are summarised in Table 8.1. Any sequences which only occurred once are amalgamated into a single group labelled *Other sequences which occurred only once* which is shown at the foot of the table. (This convention is also adopted in the presentation of the results for all the other types of state misinterpretation.)

The distribution of the frequencies of occurrence for the various sequences of actions is rather diverse, with only two sequences occurring more than once. The first of these consisted of an action at the wrong time which was attributed to a communication failure that was due to incorrect expectations on the part of the flight crew (*EM1-CO1-CO1.S6*). This sequence accounted for two out of the 21 possible causal sequences (9.5%). The second sequence, which

Table 8.1 **Frequency of occurrence of sequences for state misinterpretation type of *Cleared to push by ground crew*.**

Sequence Length	Description (Sequence Codes)	Frequency of Occurrence	%
3	Action occurred at the wrong time due to a communication failure arising from expectations by the flight crew (EM1-CO1-CO1.S6).	2	9.5%
2	Action occurred at the wrong time due to an earlier omission (EM1-EM1.S2).	2	9.5%
-	Other sequences which occurred only once.	17	81.0%
	Total	21	100.0%

also occurred twice (9.5%), consisted of an action at the wrong time due to the earlier omission of some other action (*EM1-EM1.S2*). Even if the error modes are ignored, there are only two other sub-sequences, apart from *CO1-CO1.S6* which occur more than once. The first was where a distraction was attributed to a competing task (*TP3-TP3.S4*), which occurred twice (9.5%).

The second, which also occurred twice (9.5%), was where a communication failure was precede by a breakdown in Crew Resource Management (CRM; *CO1-CO1.S4*). CRM was introduced as a concept to address the fact that flying an aircraft depended on good communication and co-ordination between all members of the crew. Initially the concept was called Cockpit Resource Management (Wiener, Kanki, & Helmreich, 1993), and was targeted at those people directly responsible for controlling the aircraft. The concept has been expanded to encompass the whole flight crew—including the flight attendants—hence the change of name to *Crew* Resource Management (see Orlady & Orlady, 1999 for a historical overview).

The overall picture for this type of incident is somewhat mixed. Although there are only two sequences of actions that occurred more than once, a communication failure (*CO1*) was reported in 11 of the 21 possible causal sequences (52.3%). Communication between the flight crew and the ground crew initially takes place using an intercom, with the ground crew using a headset connected to a jack socket on the aircraft's exterior, and ultimately is based on a series of hand signals once the ground crew have disconnected from the aircraft. Whilst there is a formal communication protocol which should be used, the data suggest that this protocol is not always strictly followed. Furthermore, when the communication consists of hand signals only, it becomes unidirectional, with the flight crew's responses to the ground crew's gestures consisting of actions which involve moving the aircraft.

8.2 Position on Ground Short of Action Point

When the aircraft is on the ground at an airport, the flight crew will be instructed by ATC how to proceed to the required location at the airport. The location will either be the departure runway for the flight, or the gate at which the passengers are to be disembarked. The instructions issued by ATC may include directions to hold short of a runway, until cleared to proceed onto (or across) that runway, or to follow a particular taxiway at an intersection; runways and taxiways are identified by signs placed on the ground.

In Table 8.2, the frequency of occurrence for sequences of actions in those incidents where the flight crew falsely believed that they were short of the action point specified by ATC are shown. The sequence of actions which occurred most often was where an aircraft travelled the wrong distance when the flight crew missed an observation due to a distraction which was attributed to a competing task (*EM4-OB1-TP3-TP3.S4*). This sequence comprised five of the 37 possible causal sequences (13.5%).

If the error modes are ignored, the longest, most frequently occurring sub-sequence of actions is where an observation was missed which was attributed to a distraction caused by a competing task (*OB1-TP3-TP3.S4*) which occurred in six sequences (16.2%). Apart from the sub-sequences of this sequence—*OB1-TP3,* and *TP3-TP3.S4*—only two other sub-sequences occurred more than twice. The first of these was where there were access problems which were attributed to an inadequate workplace layout (*PI1-WC2*) which occurred in six of the possible causal sequences (16.2%). The second was where a wrong identification was made by the flight crew, which was attributed to erroneous information (*OB3-OB3.S3*); this occurred in three of the

Table 8.2 Frequency of occurrence of sequences for state misinterpretation type of *Position on ground short of action point*.

Sequence Length	Description (Sequence Codes)	Frequency of Occurrence	%
4	Aircraft travelled the wrong distance because an observation was missed due to a distraction by a competing task (EM4-OB1-TP3-TP3.S4)	5	13.5%
3	Aircraft travelled the wrong distance because an observation was overlooked due to habit or expectancy (EM4-OB3-OB3.S4).	3	8.1%
3	Aircraft travelled the wrong distance due to access problems caused by inadequate workplace layout (EM4-PI1-WC2).	2	5.4%
-	Other sequences which occurred only once	27	73.0%
	Total	37	100.0%

possible causal sequences (8.1%).

For this type of incident, the picture is a fairly mixed one in terms of trying to identify any characteristic sub-sequences of actions. There appear to be three different tendencies, however. The first relates to missing the signs on the ground, which was attributed to a distraction, usually of a competing task which formed part of the normal sequence of operations, such as running standard checklists. The second is where the crew made a wrong identification, which in some cases was because the crew believed that the hold short lines were in one place, when they had actually been moved. In some cases the old lines are simply painted over, and hence may still be visible. The third tendency is where the layout of the signs has been poorly designed, such that they are difficult to read, or are not located in the same relative positions for each of the taxiways or runways.

8.3 Flying Cleared Route

The route that the aircraft is to follow during a flight is submitted to ATC beforehand in the form of a flight plan. The submitted flight plan may not be accepted as is, in which case ATC will inform the flight crew appropriately of any changes. In addition, due to flying restrictions around airports, standard routes are often used (SIDs for departures, and STARs for arrivals); the flight crew carry the published charts for these routes. The route that the flight will follow is normally discussed by the flight crew as part of the pre-flight preparations.

Table 8.3 shows the frequency of occurrence for those sequences of actions arising in incidents where the flight crew mistakenly believed that they were flying along the route for which they had received clearance from ATC. There were two most frequently occurring sequences. The first is one in which an action occurs at the wrong time when an observation is missed due to a planning failure which is attributed to overlooking one or more preconditions for the plan (*EM1-OB1-PL1-PL1.S4*). The second is one in which the aircraft flew in the wrong direction when an observation was missed due to a distraction by a competing task (*EM6-OB1-TP3-TP3.S4*). Both of these sequences occurred four times in the 76 possible causal sequences (5.3%).

The longest, most frequently occurring sub-sequence of actions was an observation that was missed due to a distraction by a competing task (*OB1-TP3-TP3.S4*). This sub-sequence actions occurred in nine of the 76 possible causal sequences (11.8%). The most frequently occurring sub-sequences of actions were a distraction which was attributed to a competing task (*TP3-TP3.S4*; 12 instances; 15.8%) and a failure in planning which was attributed to overlooking a precondition (*PL1-PL1.S4*; 11 instances; 14.5%).

Table 8.3 Frequency of occurrence of sequences for state misinterpretation type of *Flying cleared route*.

Sequence Length	Description (Sequence Codes)	Frequency of Occurrence	%
4	Action occurred at the wrong time due to a missed observation arising from an inadequate plan due to a precondition being overlooked (EM1-OB1-PL1-PL1.S4).	4	5.3%
4	Aircraft travelled in the wrong direction because of a missed observation due to a distraction by a competing task (EM6-OB1-TP3-TP3.S4).	4	5.3%
4	Action occurred at the wrong time because an observation was missed due to a distraction by a competing task (EM1-OB1-TP3-TP3.S4).	3	3.9%
3	Action occurred at the wrong time because of an inadequate plan which was due to a precondition being overlooked (EM1-PL1-PL1.S4).	3	3.9%
2	Aircraft travelled in the wrong direction due to an equipment failure (EM6-EF1).	3	3.9%
4	Aircraft travelled the wrong distance because an observation was missed due to a distraction by a competing task (EM4-OB1-TP3-TP3.S4)	2	2.6%
3	Aircraft travelled too far because of an inadequate plan which was due to a precondition being overlooked (EM4-PL1-PL1.S4).	2	2.6%
3	Action occurred at the wrong object due to a distraction by a competing task (EM7-TP3-TP3.S4).	2	2.6%
2	Action occurred out of sequence through inattention (EM2-TP6).	2	2.6%
-	Other sequences which occurred only once.	51	67.1%
	Total	76	100.0%

In those incidents where the pilots incorrectly believed that they were following the route along which they had been cleared by ATC is rather mixed. In 22 out of 76 (28.9%) of the possible causal sequences the error mode which resulted is directly attributed to the flight crew missing an observation. Similarly, in 23 (30.3%) of the possible causal sequences (which overlap with those in which an observation was missed), there is either a distraction by a competing task, or a planning failure due to a precondition being overlooked.

8.4 Climbing to Altitude <x>

During the initial climb phase of the flight ATC will tell the flight crew when they are cleared to climb, and to what altitude. The length of time it takes to climb from one altitude to another depends on the velocity of the aircraft, and the change in altitude. If the climb is going to take a matter of minutes the flight crew will often simultaneously perform other tasks, but should normally monitor the progress of the climb to ensure that it is halted at the appropriate altitude.

Generally, the flight crew will utilise the altitude alerter, setting it to generate an alert (audible or visible depending on the type of aircraft) when the aircraft exceeds the set altitude value. If the altitude alerter does go off, however, this indicates that the aircraft has already climbed too far.

Table 8.4 shows the frequency of occurrence for sequences of actions where the flight crew thought the aircraft was still climbing to the new altitude, when it had already reached or passed that altitude. The most frequent possible causal sequence was where the aircraft travelled the wrong distance, when an observation was missed by the flight crew, and this was attributed to a distraction by a competing task (coded sequence *EM4-OB1-TP3-TP3.S4*). This sequence comprised 17 out of the 50 possible causal sequences (34%). Ignoring the error modes, the longest sub-sequence of actions had a length of three, consisting of a missed observation attributed to distraction by a competing task (*OB1-TP3-TP3.S4*; 18 out of 50 possible causal sequences; 36%).

The sub-sequence of actions which occurred most often was a missed observation which was attributed to a distraction (*OB1-TP3*; 26 out of 50 possible causal sequences, or 52% of cases). Of the eight instances where the distraction was not attributed to a competing task, the only attributed cause that appears more than once is equipment failure (*EF1*), which occurred three times. Altogether, equipment failure appears as an attributed cause of a distraction in five of the possible causal sequences (10%).

Table 8.4 Frequency of occurrence of sequences of actions for state misinterpretation type of *Climbing to altitude <x>.*

Sequence Length	Description (Sequence Codes)	Frequency of occurrence	%
4	Aircraft travelled the wrong distance because an observation was missed due to a distraction by a competing task (EM4-OB1-TP3-TP3.S4).	17	34.0%
2	Aircraft travelled the wrong distance due to an equipment failure (EM4-EF1).	4	8.0%
4	Aircraft travelled the wrong distance because an observation was missed due to a distraction by an equipment failure (EM4-OB1-TP3-EF1).	3	6.0%
3	Aircraft travelled the wrong distance due to a distraction by a competing task (EM4-TP3-TP3.S4).	3	6.0%
4	Action occurred at the wrong time due to a communication failure arising from a distraction by a competing task (EM1-CO1-TP3-TP3.S4).	2	4.0%
2	Aircraft travelled the wrong distance due to an ambiguous label (EM4-EM4.S1).	2	4.0%
-	Other sequences which occurred only once.	19	38.0%
	Total	50	100.0%

The general picture of incidents where the flight crew (incorrectly) believed that the aircraft was climbing to the altitude to which they had been cleared by ATC is dominated by missed observations and distractions. The data suggest that the aircraft flies through the cleared altitude because the flight crew are attending to tasks such as reading navigation charts, looking for traffic in the vicinity, or complying with ATC clearances (e.g. turning to follow a new heading).

8.5 Altimeter Setting OK

The resetting of the pressure for the altimeter is a task that has to be performed by the flight crew when the aircraft passes through the transition altitude (18,000 feet above sea level in the USA). Above the transition altitude a standard pressure value setting is used on the altimeter (29.92 inches of mercury in the USA, or 1013.2 millibars) so that vertical separation between all aircraft in the same air space can be maintained. A one millibar error in the setting equates to an error of 30 feet in the altitude reading.

The frequency of occurrence for the sequences of actions which occurred in incidents where the flight crew believed that the altimeter setting was correct, are summarised in Table 8.5. The

Table 8.5 Frequency of occurrence of sequences of actions for state misinterpretation type of *Altimeter setting OK.*

Sequence Length	Description (Sequence Codes)	Frequency of occurrence	%
4	Aircraft travelled the wrong distance because an observation was missed due to a distraction by a competing task (EM4-OB1-TP3-TP3.S4).	6	15.8%
2	Aircraft travelled the wrong distance because an observation was missed (EM4-OB1).	4	10.5%
4	An action occurred out of sequence because an observation was missed due to a distraction by a competing task (EM2-OB1-TP3-TP3.S4).	2	5.3%
4	Aircraft travelled the wrong distance because an observation was missed due to a distraction by adverse weather conditions (EM4-OB1-TP3-AC7).	2	5.3%
4	Aircraft travelled the wrong distance because an observation was missed due to fatigue brought about by an early morning start (EM4-OB1-TP4-TP4.S2).	2	5.3%
3	An action occurred out of sequence due to a distraction by adverse weather conditions (EM2-TP3-AC7).	2	5.3%
2	Aircraft travelled the wrong distance due to an ambiguous label (EM4-EM4.S1).	2	5.3%
2	Aircraft travelled the wrong distance due to a communication failure (EM4-CO1).	2	5.3%
-	Other sequences which occurred only once.	16	42.1%
	Total	38	100.0%

sequence of actions that occurred most often (six out of 38 incidents; 15.8%) involved the plane travelling the wrong distance due to an observation being missed (that the aircraft was approaching or at the transition altitude) because of a distraction by a competing task (*EM4-OB1-TP3-TP3.S4*). If the error modes are ignored, however, then the longest, most frequently occurring sub-sequence of antecedent actions (eight out of 38 incidents; 21.1%) was a missed observation, which arose due to the flight crew being distracted by a competing task (*OB1-TP3-TP3.S4*).

The most frequently occurring sub-sequence of actions was a missed observation preceded by a distraction (*OB1-TP3*), which happened in 12 of the 38 possible causal sequences (31.6%). The distraction in seven of these sub-sequences was attributed to a competing task (as already noted), whilst in two instances it was attributed to adverse weather conditions (*OB1-TP3-AC7*). The table also shows that there were two other instances where the distraction was attributed to adverse weather conditions (*TP3-AC7*).

The other point of note is that there were three instances where a missed observation was attributed to fatigue (*OB1-TP4*). In two cases the fatigue was attributed to an early morning start (*OB1-TP4-TP4.S2*), whilst the other was simply attributed to exhaustion (*OB1-TP4-TP4.S1*).

The overall picture of incidents where the altimeter setting was (incorrectly) assumed to be correct is one in which distractions play a major role. The data provide some evidence for the fact that the altimeter is not reset due to some distraction, and the flight crew do not notice that the altimeter has not been reset or that the transition altitude has been passed. The nature of the distraction is generally either adverse weather (such as icing conditions or thunderstorms) which require that the flight crew pay extra attention to what is happening outside of the aircraft, or a competing task such as the need to talk to the passengers, or communicate with ATC.

8.6 Cleared to altitude <x>

Flight crews should only climb or descend to a new altitude when cleared to do so by ATC, to ensure that aircraft in the same traffic area maintain their mandated vertical and horizontal separations. The clearance will contain an altitude to which the aircraft can proceed, which may not always be the aircraft's final altitude. In general, an aircraft which is planned to fly at a cruise altitude of 35,000 feet, for example, will receive a number of clearances to intermediate altitude levels, before the final altitude is reached. The flight crew will often be unaware beforehand what these intermediate altitudes will be. The communication of the clearances between ATC and the flight crew should follow a strictly defined protocol which allows any discrepancies to be detected and repaired if part of the clearance gets blocked by noise, for example.

Table 8.6 Frequency of occurrence of sequences for state misinterpretation type of *Cleared to altitude <x>*.

Sequence Length	Description (Sequence Codes)	Freq. of occurrence	%
3	Aircraft travelled the wrong distance due to a communication failure arising through a hearing failure (EM4-CO1-CO1.S10).	14	12.5%
4	Aircraft travelled the wrong distance because an observation was missed due to a distraction by a competing task (EM4-OB1-TP3-TP3.S4).	11	9.8%
3	Action occurred at the wrong time due to a communication failure arising through a hearing failure (EM1-CO1-CO1.S10).	6	5.4%
3	Aircraft travelled the wrong distance due to a communication failure arising through a readback failure (EM4-CO1-CO1.S12).	6	5.4%
3	Aircraft travelled the wrong distance due to a communication failure arising through habit or expectancy (EM4-CO1-CO1.S11).	3	2.7%
3	Action occurred at the wrong time due to a communication failure arising through a readback failure (EM1-CO1-CO1.S12)	3	2.7%
3	Action occurred at the wrong time due to a trapping error (EM1-EM1.S2)	3	2.7%
3	Action occurred at the wrong object due to a communication failure arising through a hearing failure (EM7-CO1-CO1.S10).	3	2.7%
4	Action occurred out of sequence because an observation was missed due to a distraction by a competing task (EM2-OB1-TP3-TP3.S4).	2	1.8%
4	Aircraft travelled the wrong distance because an observation was missed due to fatigue brought about through exhaustion (EM4-OB1-TP4-TP4.S1).	2	1.8%
4	Action occurred at wrong object because an observation was missed due to a distraction by a competing task (EM7-OB1-TP3-TP3.S4).	2	1.8%
3	Action occurred at the wrong time due to a communication failure arising through a presentation failure (EM1-CO1-CO1.S2)	2	1.8%
3	Action occurred at the wrong time due to a communication failure because of radio congestion (EM1-CO1-CO1.S7)	2	1.8%
3	Action occurred at the wrong time through the use of an inadequate plan due to the overlooking of a precondition (EM1-PL1-PL1.S4).	2	1.8%
3	Aircraft travelled the wrong distance due to a communication failure arising from a CRM failure (EM4-CO1-CO1.S4).	2	1.8%
3	Aircraft travelled the wrong distance through the use of an inadequate plan due to the overlooking of a precondition (EM4-PL1-PL1.S4).	2	1.8%
2	Aircraft travelled the wrong distance due to a communication failure (EM4-CO1).	2	1.8%
-	Other sequences which occurred only once.	45	40.2%
	Total	112	100.0%

In Table 8.6, the frequency of occurrence for sequences of actions where the flight crew believed they were cleared to one particular altitude, when they were actually cleared to a different altitude are shown. Only two sequences occurred more than 10 times out of 112 possible causal sequences. The first sequence was where the aircraft travelled the wrong distance due to a breakdown in communication that was attributed to a hearing failure (*EM4-CO1-CO1.S10*), which occurred 14 times (12.5%). In other words when the flight crew did not hear the initial instructions from the air traffic controller correctly. The term *hearing failure* is used to distinguish it from a hearback failure which occurs when the flight crew do not hear the correct instructions when the controller repeats them. The second sequence was where the aircraft flew the wrong distance when an observation was missed, which was attributed to a distraction by a competing task (*EM4-OB1-TP3-TP3.S4*), which occurred 11 times (9.8%).

The longest, most frequently occurring sub-sequence of actions is one in which an observation was missed due to a distraction that was attributed to a competing task (*OB1-TP3-TP3.S4*). This sub-sequence occurred in 16 of the 112 possible causal sequences (14.3%). The most frequently occurring sub-sequence of actions, however, is a communication failure that was attributed to a hearing failure (*CO1-CO1.S10*) which occurred in 23 of the possible causal sequences (20.5%).

In those incidents where the flight crew believed that they were cleared to one altitude, whereas they were cleared to a different altitude, the general picture is mixed. The main problem is one of communication. If the communication of the clearance fails, then it is almost to be expected that the flight crew will transit to the incorrect altitude. The communication procedure between ATC and the aircraft should, theoretically, prevent such a problem from arising, but there is some evidence here to suggest that it does not. Although there are several possible explanations for the communication failure, such as radio distortion, the data here suggest that incorrect expectations also play a part, in that the flight crew hear the altitude that they want to hear—biased by previous experiences—rather than the altitude that the controller actually said. Since the controllers in these incidents did not correct the flight crew's reading back of the altitude value, this suggests that the controllers may also have been subject to the same bias, believing that the flight crew read back the altitude value that the controller originally read out.

The other particular point to note is that in those instances where a distraction was caused by a competing task, the nature of that task varied considerably. In some cases it was related to changing altitude:

- looking for traffic;
- reprogramming the FMS;

- checking navigation equipment

whilst in other cases it was related to communication:

- finding out the number of the gate to which the aircraft had to taxi before disembarking the passengers after landing;

- finding out the latest information for arrival or departure at a particular airport by listening to the recorded message, known as the Automatic Terminal Information Service or ATIS;

- or making a call to the company operating the flight to report any required information.

8.7 Position in Air Short of Action Point

When the plane is in flight, ATC will sometimes issue clearance instructions which have to be complied with at a specific location point, referred to here as action points. This location is normally specified in terms of waypoints, which appear on navigation charts, and are defined with reference to radio navigation stations. As such, these waypoints are not visible to the eye, although if a plane is equipped with an FMS, the flight crew will be able to include them on the programmed route.

The frequency of occurrence for sequences of actions in those incidents where the flight crew incorrectly thought that they still had not reached the action point are shown in Table 8.7. There are only three sequences of actions which occurred more than once. The first of these sequences is where an action happened at the wrong time when a distraction occurred which was attributed to a competing task (*EM1-TP3-TP3.S4*). The second is where an action happened at the wrong

Table 8.7 Frequency of occurrence of sequences for state misinterpretation type of *Position in air short of action point*.

Sequence Length	Description (Sequence Codes)	Frequency of Occurrence	%
4	Action occurred at the wrong time because an observation was missed due to a distraction by a competing task (EM1-OB1-TP3-TP3.S4).	2	8.7%
3	Action occurred at the wrong time due to a distraction by a competing task (EM1-TP3-TP3.S4).	2	8.7%
3	Action occurred at the wrong time due to a distraction by an extraneous conversation (EM1-TP3-TP3.S6).	2	8.7%
-	Other sequences which occurred only once	17	73.9%
	Total	23	100.0%

time because of a distraction which was attributed to an extraneous conversation going on in the cockpit (*EM1-TP3-TP3.S6*). The third is where an action happened at the wrong time after the flight crew missed an observation because of a distraction which was attributed to a competing task (*EM1-OB1-TP3-TP3.S4*). Each of these three different sequences occurred twice in the 22 possible causal sequences (9.1% each).

If the error modes are ignored, the only sub-sequence of more than two actions which occurred more than once is where an observation was missed due to a distraction which was caused by a competing task (*OB1-TP3-TP3.S4*). This sub-sequence occurred in three of the 22 possible causal sequences (13.6%).

The most frequently occurring sub-sequence of actions is a distraction which was attributed to a competing task (*TP3-TP3.S4*) which appeared in five out of 22 possible causal sequences (22.7%). The only other sub-sequence which occurred more than twice was where the flight crew missed an observation, and this was attributed to a distraction, which appeared in four possible causal sequences (18.2%). These two sub-sequences overlapped in three possible causal sequences.

The overall picture for the incidents involving a state misinterpretation of type *Position in air short of action point* is rather a mixed bag. Where a distraction was attributed to a competing task, in three cases the flight crew were performing normal standard procedures (such as carrying out pre-descent checks), whilst in the other two cases one of the members of the flight crew was busy listening to the latest ATIS message.

8.8 Radio Navigation Frequency Setting OK

The route that the flight crew follow is plotted with reference to various radio navigation beacons. When the flight crew tune the aircraft's RNAV equipment into the frequency of one of these beacons, the information it supplies is displayed on the aircraft instrumentation. The aircraft's position is therefore displayed with respect to the beacon that the flight crew have currently tuned into. So, if the flight crew fail are tuned into the wrong beacon, the aircraft's current position will be shown with respect to that (wrong) beacon.

Table 8.8 shows the frequencies of occurrence for the sequences of actions in those incidents where the flight crew believed that that the radio navigation frequency settings were correct. Only two possible causal sequences of actions occurred more than once for this type of state misinterpretation. The first is where the aircraft travelled in the wrong direction when the flight crew missed an observation because they were distracted by a competing task (EM6-OB1-TP3-

Table 8.8 Frequency of occurrence of sequences for state misinterpretation of type *RNAV Frequency setting OK.*

Sequence Length	Description (Sequence Codes)	Frequency of Occurrence	%
4	Aircraft travelled in the wrong direction when an observation was missed due to a distraction by a competing task (EM6-OB1-TP3-TP3.S4).	2	6.1%
2	Action occurred at the wrong object when an observation was missed (EM7-OB1).	2	6.1%
-	Other sequences which occurred only once	29	87.9%
	Total	33	100.0%

TP3.S4), which comprised two of the 33 possible causal sequences (6.1%). The second is where an action occurred at the wrong object when an observation was missed by the flight crew (*EM7-OB1*), which also occurred twice (6.1%).

If the error modes are ignored, then the longest most frequently occurring sub-sequence of actions is where an observation was missed by the flight crew due to a distraction which was attributed to a competing task (*OB1-TP3-TP3.S4*). This sub-sequence occurred in six of the 33 possible causal sequences (18.7%). The sub-sequence which occurred most often, is where a missed observation was attributed to a distraction (*OB1-TP3*) which appeared in nine of the possible causal sequences (27.3%).

For this type of state misinterpretation the general picture is one in which missed observations and distractions played a large part, but not always together. The most frequently attributed cause of the error modes (15 out of 33 possible causal sequences; 45.5%) was a missed observation: the flight crew did not spot that they were tuned into the wrong radio frequency, and hence to the wrong radio navigation aid. A distraction (*TP3*) appears in 14 possible causal sequences (42.4%), and, although several of these distractions were attributed to a competing task (seven out of the 13), there were also three that were attributed to extraneous conversation between members of the flight crew.

8.9 Descending to Cleared Altitude <x>

The process of descending to a particular altitude is the counterpart of climbing to a particular altitude during the initial stages of a flight. The altitude to which the aircraft is required to descend is specified by ATC, and will often be accompanied by a location (often defined by reference to a radio navigation station) by which the aircraft should have achieved the required altitude. The descent phase usually happens as the aircraft starts to approach the destination airport, and hence will be entering areas of air space where the traffic levels are higher because of aircraft being queued to land, for example.

In Table 8.9 the frequencies of occurrence for the sequences of actions which happened during incidents where the pilots mistakenly believed that the aircraft was descending to the altitude to which it had been cleared by ATC are summarised. The table shows that the most frequently occurring possible causal sequence was where the aircraft flew the wrong distance when the flight crew missed an observation due to a distraction that was attributed to a competing task (*EM4-OB1-TP3-TP3.S4*). This sequence of actions comprised 19 out of the 67 possible causal sequences (28.4%). It is also interesting to note there are two similar sequences of actions—similar in that the aircraft flew too far when an observation was missed due to a distraction (*EM4-OB1-TP3*)—which both occurred five times (7.5% each). For the first of these sequences the attributed cause of the distraction was adverse weather conditions (full sequence *EM4-OB1-TP3-AC7*), whilst in the second the attributed cause of the distractions was a radio call interruption (*EM4-OB1-TP3-TP3.S8*).

If the error modes are ignored, then the longest, most frequently occurring sub-sequence of actions was a missed observation which was attributed to a distraction by a competing task (*OB1-TP3-TP3.S4*). This sub-sequence occurred 20 times altogether (29.9%). The most

Table 8.9 Frequency of occurrence of sequences for state misinterpretation of type *Descending to cleared altitude <x>.*

Sequence Length	Description (Sequence Codes)	Frequency of Occurrence	%
4	Aircraft travelled the wrong distance because an observation was missed due to a distraction by a competing task (EM4-OB1-TP3-TP3.S4).	19	28.4%
4	Aircraft travelled the wrong distance because an observation was missed due to a distraction by adverse weather conditions (EM4-OB1-TP3-AC7).	5	7.5%
4	Aircraft travelled the wrong distance because an observation was missed due to a radio call interruption (EM4-OB1-TP3-TP3.S8).	5	7.5%
2	Aircraft travelled the wrong distance because an observation was missed (EM4-OB1).	3	4.5%
4	Aircraft travelled the wrong distance because an observation was missed due to a distraction by an extraneous conversation (EM4-OB1-TP3-TP3.S6)	2	3.0%
4	Aircraft travelled the wrong distance because an observation was missed due to a distraction through a flight attendant interruption (EM4-OB1-TP3-TP3.S7).	2	3.0%
3	Action occurred at the wrong time due to a distraction caused by a competing task (EM1-TP3-TP3.S4).	2	3.0%
-	Other sequences which occurred only once.	29	43.3%
	Total	67	100.0%

frequently occurring sub-sequence of actions, though, was an observation that was attributed to a distraction (*OB1-TP3*) which occurred in 37 of the 67 possible causal sequences (55.2%). In eight of these instances the distraction was attributed to adverse weather conditions (11.9%).

The general picture for those incidents that occurred where the flight crew mistakenly believed that they were descending to the altitude to which they had been cleared by ATC is one in which distractions play an influential part. The flight crew are distracted to such an extent that they fail to monitor the aircraft's progress towards the cleared altitude, or even fail to initiate the descent. In several instances the distraction was attributed to a competing task. The nature of the competing task varies somewhat, but generally falls into three different types: communication (getting ATIS information, talking to company, talking to ATC); navigation (retuning RNAV equipment); or vigilance monitoring (looking for traffic).

8.10 Rate of Descent OK

The descent phase may be performed as a number of steps, co-ordinated by ATC. In such cases, ATC will sometimes require that an aircraft be at a particular altitude at some specific location—a crossing restriction—in order to safely fit in with the rest of the traffic in that area. Although the final altitude and the location by which it must be achieved are specified by ATC, the rate of descent is often left up to the flight crew. The selected rate of descent will depend on the flight crew's strategy. Probably the most general strategy is to use a rate of descent that takes into account fuel considerations with the need to make the crossing restriction, although some flight crews prefer to be at the required altitude before the crossing restriction.

The frequency of occurrence for sequences of actions for those incidents in which the flight crew believed that their rate of descent was sufficient to comply with the crossing restriction are shown in Table 8.10. There were only four sequences which occurred more than once; each of them comprised two of the 20 possible causal sequences (10%). The first sequence is where an action occurred at the wrong time which was attributed to the flight crew being fatigued through exhaustion. The other three sequences all have different error modes: in the first an action occurred at the wrong time (*EM1*); in the second the aircraft travelled the wrong distance (*EM4*); and in the third the aircraft travelled at the wrong speed (*EM5*). In each of these three sequences, however, the preceding sequence of actions consisted of an observation being missed by the flight crew which was attributed to being distracted by a competing task (*OB1-TP3-TP3.S4*). Unsurprisingly, if the error modes are omitted, this sequence (*OB1-TP3-TP3.S4*) is the longest most frequently occurring sequence, occurring in seven of the possible causal sequences (35%).

Table 8.10 Frequency of occurrence of sequences for state misinterpretation type of *Rate of descent OK*.

Sequence Length	Description (Sequence Codes)	Frequency of Occurrence	%
4	Action occurred at the wrong time because of fatigue that was caused by exhaustion (EM1-OB1-TP4-TP4.S1).	2	10.0%
4	Action occurred at the wrong time because of a missed observation due to a distraction by a competing task (EM1-OB1-TP3-TP3.S4).	2	10.0%
4	Aircraft travelled the wrong distance because an observation was missed due to a distraction by a competing task (EM4-OB1-TP3-TP3.S4)	2	10.0%
4	Aircraft travelled at the wrong speed because of a missed observation due to a distraction by a competing task (EM5-OB1-TP3-TP3.S4).	2	10.0%
-	Other sequences which occurred only once.	12	60.0%
	Total	20	100.0%

The most frequently occurring sub-sequence of actions was a missed observation which was attributed to a distraction (*OB1-TP3*), which was identified in 10 possible causal sequences (50%). Only two other sub-sequences of longer than two actions occurred more than once. The first was where the flight crew missed an observation due to fatigue which was attributed to exhaustion (*OB1-TP4-TP4.S1*) which occurred twice (10%). The second was where the flight crew missed an observation due to a distraction which was attributed to an interruption by one of the flight attendants, which also occurred twice (10%).

For those incidents in which the flight crew falsely believed that their rate of descent was sufficient to make a crossing restriction the picture is dominated by distractions. In most of these cases (7 out of 10), the distraction was attributed to a competing task that was related to the descent of the aircraft, or the subsequent approach and landing. In two cases, however, the distraction was caused by an interruption by one of the flight attendants; the remaining instance was attributed to carrying on a conversation with the person travelling in the extra (jump) seat in the cockpit.

8.11 Cleared To Land

The flight crew will communicate with several different air traffic controllers during the course of the flight, which involves a change in radio communication frequency. One such change occurs when the aircraft approaches the destination airport and prepares for landing. The approach controller will normally hand off the aircraft to the tower controllers who will issue the clearance to land. In theory, if the aircraft is not cleared to land, a go-around operation should be performed, whereby the aircraft basically circles around the airport and attempts to land again. The flight crew will normally be busy when preparing for landing. These preparations include tasks such as running through predefined checklists, for example, to make sure that the landing gear has been lowered, the spoilers have been set, and the rest of the flight crew have been informed.

Table 8.11 shows the frequency of occurrence for sequences of actions in incidents where the flight crew thought that they had been cleared to land by ATC when, in fact, they had not. The sequence of actions which occurred most often was an action happening out of sequence which happened when the flight crew missed an observation due to a distraction which was attributed to a competing task (*EM2-OB1-TP3-TP3.S4*). This sequence of actions occurred in 10 of the 36 possible causal sequences (27.8%). If the error modes are ignored, however, the longest, most frequently occurring sequence of actions is a missed observation, due to a distraction by a competing task (*OB1-TP3-TP3.S4*) which arose 12 times (33.3%).

The sub-sequence *TP3-TP3.S4* (a distraction by a competing task) occurred most frequently (16 times; 43.2%), whilst *OB1-TP3* (a missed observation attributed to a distraction) came a close second (13 occurrences; 36.1%). The rest of the sub-sequences only occurred once or twice, with the exception of a distraction that was attributed to adverse weather conditions (*TP3-AC7*), which occurred three times (8.3%).

Table 8.11 Frequency of occurrence of sequences for state misinterpretation type of *Cleared to land*.

Sequence Length	Description (Sequence Codes)	Frequency of Occurrence	%
4	Action occurred out of sequence because an observation was missed due to a distraction by a competing task (EM2-OB1-TP3-TP3.S4).	10	27.8%
3	Action occurred out of sequence due to a distraction caused by a competing task (EM2-TP3-TP3.S4).	4	11.1%
3	Action occurred out of sequence due to a distraction caused by adverse weather conditions (EM2-TP3-AC7).	2	5.6%
4	Action occurred at the wrong time when an observation was missed due to a distraction by a competing task (EM1-OB1-TP3-TP3.S4).	2	5.6%
4	Action occurred out of sequence when an observation was missed due to excessive demand brought about by parallel tasks (EM2-OB1-WC1-WC1.S1).	2	5.6%
-	Other sequences which occurred only once.	16	44.4%
	Total	36	100.0%

This type of state misinterpretation is dominated by incidents in which the flight crew do not notice that they have passed the given point (often a landmark) where they were supposed to change radio frequency to contact the tower controller who will clear them to land. In a few cases this was attributed to being preoccupied by the adverse weather conditions which were affecting control of the aircraft. In several cases, however, the flight crew missed the landmark where they should change radio frequency because they were performing other duties, some of which were normal (such as running through check lists in preparation for landing), whilst others were possibly unexpected (such as having to look out for traffic in the vicinity).

8.12 General Discussion

The most significant factor across the different types of state misinterpretation covered here is the prevalence of monitoring failures (missed observations) among the possible causal sequences. The most frequently occurring sequence of actions is where a monitoring failure was caused by a distraction which was attributed to a competing task, which occurred in 106 out of a total of 513 possible causal sequences (20.7%). Although there are several individual instances in which the competing tasks are tasks that are in addition to the normal tasks that the flight crew would perform—in some cases these tasks are not directly related to the task of flying the plane—there are more instances where the competing task is valid in the context of flying the plane.

There are some types of state misinterpretation which seem to have a more typical incident profile than others. So, for example, for *Climbing to altitude <x>*, 17 of the 50 possible causal sequences ended in the aircraft travelling the wrong distance, and this was attributed to a missed observation, which resulted from the flight crew being distracted by a competing task (*EM4-OB1-TP3-TP3.S4*). Similar profiles were detected for *Descending to altitude <x>* (17 out of 67), and *Cleared to land* (10 out of 36).

The lack of a typical profile for some types of state misinterpretation could be attributed to a lack of data. In other words, more possible causal sequences are needed before a clear picture emerges. In some cases, however, there does not appear to be a typical incident profile. *Cleared to altitude <x>*, for example, has two typical profiles, both of which occurred 10 times in 76 sequences. This result is a reflection of the fact that the misinterpretation can arise in two different ways: either the flight crew believe that altitude is correct when it is wrong; or the flight crew lose track of the altitude to which they have been cleared.

One other fact which also is not clear from the sequence data is whether the predominance of one sequence is a reflection of a flaw in the underlying taxonomy of state misinterpretation. The only way that this can be determined is by comparing the trees for the different types of state misinterpretation.

8.13 Summary

In this chapter, the incident data has been considered from the level of abstraction of sequences of actions. The link between the individual actions in a sequence is a qualitative one, based on attributed causality. In the next chapter, the final part of the data analysis is described, in which an attempt is made at quantifying the relationship between actions in a sequence. The different types of state misinterpretation are laid out as trees (or networks) of causal sequences, and methods—both qualitative and quantitative—for comparing the trees across the different types of state misinterpretation are introduced.

9 Data Analysis: Trees

In Chapter 7 the data was analysed in terms of individual actions and error modes to produce a concordance listing of frequencies of occurrence for each type of state misinterpretation. Then, in Chapter 8, the data was analysed in terms of the possible causal sequences of actions for the various types of state misinterpretation. This chapter provides another level of abstraction at which the data can be viewed: the data is considered at the level of the trees of causal sequences for the types of state misinterpretation.

The data analysis in this chapter covers the fundamental ESDA operations of converting, comparing and computing. The data for each type of state misinterpretation is initially converted into a graphical representation, whereby the data is summarised as a tree of possible causal sequences of actions. These trees are used to qualitatively compare different types of state misinterpretation. After highlighting the shortcomings of such an approach, an alternative method is presented which is used to quantitatively compare pairs of types of state misinterpretation. The final analysis involves computing the relationship between the actions and error modes for the individual types of state misinterpretation, in order to identify those actions which can be considered to be most closely associated with the various types of state misinterpretation. The issues raised by the different analyses of the data are then presented and summarised in a general discussion section.

9.1 Pictorially Representing State Misinterpretation

The fundamental operations of ESDA were previously described in Chapter 6, where they were related to the CREAM-AIR method. One of the operations which is not provided by the CREAM-AIR is the conversion of data, which involves transforming the data to a different format which may allow new patterns within the data to emerge. In this section, the use of data conversion is employed to translate the sequences of data described in Chapter 8 into a graphical representation. In this instance, the data for each of the different types of state misinterpretation is transformed into a causal tree.

Each of the types of state misinterpretation can be represented as a tree in which the type of state misinterpretation can be considered as the root node of the tree. At the next level of the tree are the error modes identified for all the incidents that are classified as involving this particular type of state misinterpretation. These error modes can, in themselves, be considered as root nodes of a subtree of possible causal sequences of actions generated from the analysis of the narrative section of the incident reports using the CREAM-AIR for this type of state misinterpretation.

The tree structure for each of the different types of state misinterpretation can be generated by analysing the sequences of actions identified in Chapter 8. The simplest approach to generating the trees is to have every sequence of actions represented by a branch on that tree. This approach has the drawback, however, that the structure of the tree becomes unwieldy if there are more than a few branches (or possible causal sequences) to the tree. One alternative approach, adopted here, is to optimise the branching structure of a tree to create a generic form, in which every instance of a sub-sequences of actions is not necessarily represented in the tree. Sub-sequences of actions are only added to the tree if they are not already represented in the tree as part of a larger sub-sequence. The resultant tree therefore has several levels of branching, unlike the non-optimised version which has only one level of branching, because every sequence of actions is represented by a branch directly below the error mode nodes in the tree.

The process is best illustrated by means of an example. In Table 9.1 the frequency of occurrence for the possible causal sequences for the state misinterpretation type *Altimeter setting OK* is

Table 9.1 Frequency of occurrence for each of the possible causal sequences for state misinterpretation type of *Altimeter setting OK*

Sequence of Actions	Frequency
EM2-OB1-TP3-TP3.S4	2
EM2-PR1	1
EM2-TP3-AC7	2
EM2-TP3-TP3.S6	1
EM2-TP6	1
EM4-CO1	1
EM4-CO1-CO1.S13	1
EM4-CO1-CO1.S4	1
EM4-EM4.S1	2
EM4-OB1	4
EM4-OB1-EF1	1
EM4-OB1-TP3-AC7	2
EM4-OB1-TP3-CO1-CO1.S4	1
EM4-OB1-TP3-TP3.S4	6
EM4-OB1-TP3-TP3.S8	1
EM4-OB1-TP4-TP4.S1	1
EM4-OB1-TP4-TP4.S2	2
EM4-PL1-CO1-CO1.S5	1
EM4-PL1-CO2	1
EM4-PL1-PL1.S4	2
EM4-TI1-TI1.S7	1
EM7-CO1	1
EM7-OB3	1
EM7-TP1	1
Total	38

shown, sorted in ascending alphabetical order. The data only shows the complete sequences which begin with the error mode, and end with the last antecedent that can be found; the frequency of occurrence of sub-sequences of actions is omitted. If every one of these sequences was shown separately, the tree of sequences would have 38 branches at the highest level. The data show that there some sub-sequences which recur, however, such as *EM4-OB1* which appears on its own four times, and 14 times as part of longer sequences which begin with *EM4-OB1*.

When the optimised version of the tree is drawn, as shown in Figure 9.1, the overall structure of the tree is simpler than that for the non-optimised version in that the branching is spread across multiple levels of the tree. The total number of branches is only slightly less than that of the non-optimised tree (32 compared to 38), but the software tool which is used to draw the tree can

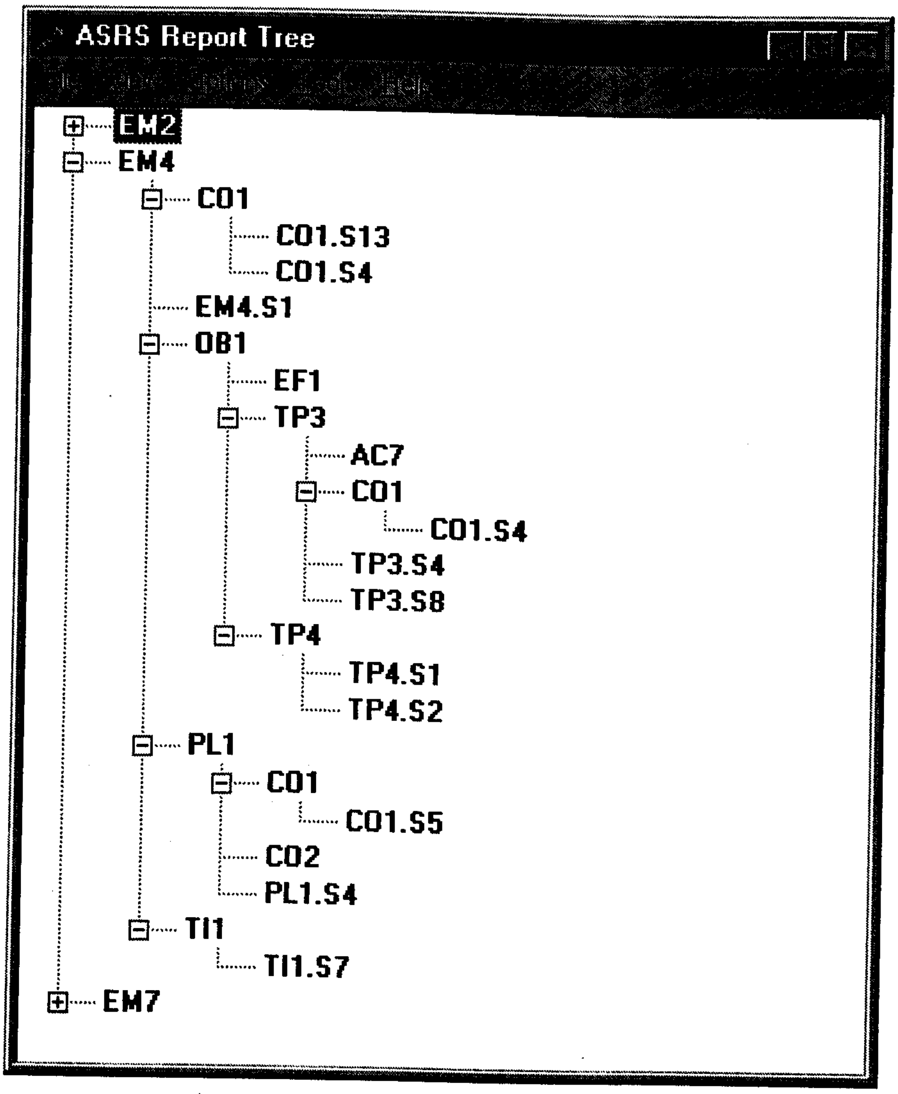

Figure 9.1 Tree Structure for Possible Causal Sequences of Actions for State Misinterpretation type of *Altimeter setting OK*.
(The figure is shown with only the branches for error mode 4 expanded in full.)

be used to collapse and expand individual branches in the tree which makes it easier to focus on those branches which are similar when comparing trees.

One advantage of the optimised format of the tree structure is that it makes it possible to carry out another of the fundamental ESDA operations: comparison of data sets. In this case, different types of state misinterpretation can be compared by simply comparing the appropriate trees. The results of the comparisons can be used to provide a measure of confidence in the correctness of the taxonomy of types of state misinterpretation. If two types of state misinterpretation really are different, then it should be expected that the trees of possible causal sequences associated with each of the types of state misinterpretation should also be different.

If the tree structure for each of the different types of state misinterpretation is generated (like that in Figure 9.1), these trees can then be visually compared to identify any sequences of actions that are common across the types of state misinterpretation. An example of this is shown in Figure 9.2 where the generic tree structures for *Altimeter Setting OK* and *Climbing to Altitude <x>* have been collapsed so that all of the common sub-sequences of possible causal branches are shown in full. The figure shows that there are indeed some similarities between these two types of state misinterpretation, with the following coded sequences appearing in the trees for both types of state misinterpretation:

- *EM4-CO1*
- *EM4-EM4.S1*
- *EM4-OB1-EF1*
- *EM4-OB1-TP3-AC7*
- *EM4-OB1-TP3-TP3.S4*
- *EM4-OB1-TP3-TP3.S8*
- *EM4-OB1-TP4-TP4.S1*
- *EM4-OB1-TP4-TP4.S2*

Care needs to be exercised when comparing the trees, however, since the number of unique sequences (which have been deliberately omitted from Figure 9.2) is generally greater than the number of sequences which occur more than once. In other words, the similarity apparent from visually comparing the trees is usually quite small.

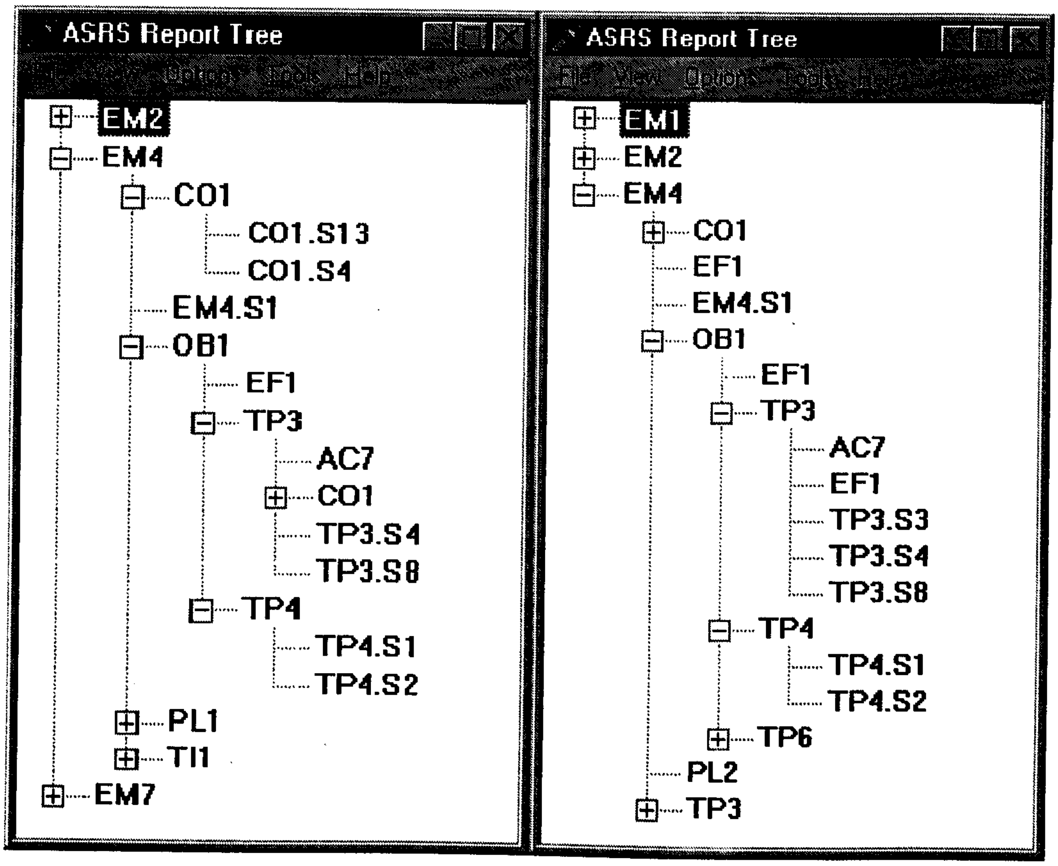

Figure 9.2 Tree of matching possible causal sequences.

(Left hand picture shows the tree for state misinterpretation type of *Altimeter setting OK*; right hand picture shows the tree for *Climbing to altitude* <x>).

Although the comparison of generic tree structures provides a relatively quick and easy way of measuring the similarity between types of state misinterpretation, there are two shortcomings of such an approach. The first is that the generic tree structures abstract away the frequency of occurrence of sub-sequences of actions. In general, the more times a sub-sequence of actions appears for a particular type of state misinterpretation, the more that sub-sequence can be considered to be characteristic of that state misinterpretation type. The generic trees reduce multiple instances of a sub-sequence of actions to a single instance which can distort the visual comparison. All that the visual comparison shows is those sub-sequences of actions which are common across different trees, rather than a precise measure of the similarity of the different types of state misinterpretation.

The second shortcoming, which also applied to the non-optimised tree structures, is that even with the generic structure it quickly becomes difficult to make visual comparisons as the number of branches in the trees increases. Each time a new incident is reported for a particular type of state misinterpretation, any possible causal sequences for that incident which are not already represented in the tree have to be added. Every new sub-sequence that gets added to the tree increases the complexity of the tree structure, which makes it more difficult to visually compare it to other tree structures.

The shortcomings suggest that what is required is some way of comparing the data which takes into account all of the possible causal sequences of actions for each of the different types of state misinterpretation. The next section shows one approach that can be used to quantify the similarity between pairs of trees.

9.2 Comparing Types of State Misinterpretation

Representing the different types of state misinterpretation as optimised generic trees makes it easier to perform a quick visual comparison, but what is required is a more robust method for determining the degree of commonality between the different types of state misinterpretation. The existence of a possible causal sequence in more than one type of state misinterpretation means that finding a means of breaking or preventing this sequence will help to reduce or eliminate multiple instances of state misinterpretation. Measuring the commonality of sequences across trees is not a simple task, however, and any measure of commonality will necessarily be a relative measure which reflects the level of refinement of the various types of state misinterpretation. The more incidents there are associated with a particular type of state misinterpretation, the more representative will be the causal tree of sequences of actions for that type of state misinterpretation.

There are three ways of comparing the causal trees (in a pairwise manner) that will be considered here, each of which revolves around the basic idea that all of the trees are built up from a set of causal sequences which has an error mode at its root. The commonality between the trees is therefore measured by working out the intersection between the two sets of causal sequences and expressing this as a ratio of the union between the two sets. The three methods, which are all related, are:

1. Simple comparison. In this method the number of sequences that are common to the set of sequences for each of the two trees are counted. The result is divided by the total number of unique sequences in the union of the two sets of sequences.

2. Standardised comparison. This method takes some account of the fact that the number of sequences in the set of sequences for each of the trees is related to the number of incidents reported for that type of state misinterpretation. The frequency of occurrence of sequences in the smaller tree is multiplied by the ratio of the

number of sequences in the larger tree and divided by the number of branches in smaller tree. This calculation effectively extends the smaller of the two trees so that it has the same number of nodes as the larger of the trees. Once this has been done, the comparison is basically the same as that for the simple comparison method, the only difference being that both of the sets of sequences will be the same size.

3. Weighted standardised comparison. The standardised comparison does not take into account the length of the sequences in the tree. The weighted standardised comparison overcomes this shortcoming by first standardising the two trees that are being compared (as in the standardised comparison method) and then multiplies the frequency of occurrence of the individual sequences (branches) by the length of the branch. The final step is then the same as the simple comparison method.

The values produced by the three types of comparison are each slightly different, and all of them require some interpretation. The results of the comparisons generally need to be discussed in context, since only values of 0—the trees are distinct—and 1—the trees are identical—can be understood independently. The reduction of similarity to a single value is a convenience, although it can be used as a way of testing whether two trees really represent different types of state misinterpretation. The closer the result of the comparison calculation is to unity, the greater the reason for suspecting that the two trees may be representations of the same type of state misinterpretation. The comparison value should not be used in isolation for this purpose, however, and other factors such as the number of branches in the trees need to be taken into account. The larger the number of branches in both trees, the greater can be the confidence in the comparison value.

Although the comparison values provide a convenient way of expressing the similarity between two trees, they should not be used to compare pairs of trees. So, for example, if two different pairs of trees both produce a simple comparison value of 1 this does not necessarily mean that all four trees are identical. The trees are only identical if all of the branches are the same for all four of the trees.

In the context of comparing state misinterpretation in pilot behaviour some general predictions can be made regarding the relative magnitude of the values yielded by the comparisons. The workload of the flight crew, measured simply as the number of tasks that have to be performed, varies according to the phase of the flight. Since flying an aircraft is a highly proceduralised skill, the tasks which need to be performed will be more or less invariant across different flights. It can therefore be predicted that pairs of trees for those types of state misinterpretation which occur at the same phase of the flight should generally be more closely related than those types of state misinterpretation that occur during different flight phases. The basic reason for this is that during the same flight phase the context in which the actions take place will be more or less the same, i.e. the same things are going on in the cockpit, so there will be the same potential for the same sequences of actions to arise.

The results of each of the different types of comparisons are described in the next three sections. The data for each is presented as a table in which the types of state misinterpretation are used to label the rows and the columns. The diagonals of the tables are blank because the comparison between one type of state misinterpretation and itself would obviously yield the value 1. The lower left hand half of each of the tables is blank too, because the comparison value for a pair of state misinterpretation types is independent of the ordering. So, for example, the comparison value for *Altimeter setting OK* and *Climbing to altitude <x>* is the same as the comparison value for *Climbing to altitude <x>* and *Altimeter setting OK*.

9.2.1 Simple Comparison of State Misinterpretation Types

Table 9.2 shows the results of performing the simple comparison calculation on all of the possible pairs of state misinterpretation types (or trees). In all 55 comparison calculations were performed, with 10 results being generated for each pair of state misinterpretation types. Out of the 55 comparisons, three of the results are zero, and in two of these cases the pair of state misinterpretation types included *Position on ground short of action point*.

The vast majority of the comparisons (45 out of 55) yielded a value that was less than 0.1. For five types of state misinterpretation (*Cleared to land, Cleared to push by ground crew, Flying cleared route, Position in air short of action point,* and *RNAV frequency setting OK*) all of the comparison values were less than 0.1.

Nine of the ten remaining comparisons yielded values in the range 0.1-0.2. In five of these cases the types of state misinterpretation were those which directly involve altitude clearances (*Altimeter setting OK, Climbing to altitude <x>, Cleared to altitude <x>, Descending to cleared altitude <x>*). In a further two comparisons *Position on ground short of action point* was involved, which ostensibly seems completely unrelated to the other types of state misinterpretation mentioned.

Table 9.2 Results of simple comparisons between pairs of types of state misinterpretation.

	Altimeter Setting OK	Climbing to altitude <x>	Cleared to altitude <x>	Cleared to land	Cleared to push by ground crew	Descending to cleared alt. <x>	Flying cleared route	Pos. in air short of action point	Pos. on ground short of action point	Rate of descent OK	RNAV frequency setting OK
Altimeter Setting OK		0.189	0.100	0.072	0	0.193	0.056	0.052	0.172	0.074	0.060
Climbing to altitude <x>			0.125	0.012	0.044	[illegible]	0.041	0.029	0.115	0.077	0.038
Cleared to altitude <x>				0.021	0.047	0.133	0.062	0.071	0.072	0.048	0.036
Cleared to land					0.018	0.051	0.028	0.055	0	0.037	0.078
Cleared to push by ground crew						0.012	0.021	0.075	0.018	0.025	0.039
Descending to cleared altitude <x>							0.052	0.059	0.106	0.061	0.042
Flying cleared route								0.076	0.018	0.055	0.058
Position in air short of action point									0	0.103	0.057
Position on ground short of action point										0.056	0.015
Rate of descent OK											0.060
RNAV frequency setting OK											

Part of the reason why such apparently radically different types of state misinterpretation could appear to be similar (albeit to a lesser extent) lies in the level of abstraction used in the CREAM-AIR. A closer inspection of the two comparisons where *Position on ground short of action point* was involved shows that the sequence of events where an aircraft travels too far when an observation is missed due to a distraction by a competing task (*EM4-OB1-TP3-TP3.S4*) is common across the three different types of state misinterpretation (*Position on ground short of action point*, *Climbing to altitude* <*x*>, and *Descending to cleared altitude* <*x*>). The difference is that when the aircraft is on the ground the observation that is usually missed is a sign outside the aircraft, whereas for the other two types of state misinterpretation the missed observation is usually related to one of the cockpit instruments.

The one remaining comparison yielded a value in the range 0.2-0.3. This comparison involved *Climbing to altitude* <*x*> and *Descending to cleared altitude* <*x*>. On one level, these types of state misinterpretation can be regarded as complementary. In the first type, the flight crew are often busy completing tasks that have to be done after take-off; in the second, the flight crew are often busy making preparations for landing. Otherwise the basic goal is to make the aircraft move from one altitude to another one.

Overall the results of performing simple comparisons between the various pairs of state misinterpretation types offers a measure of confidence in the selected taxonomy of types of state misinterpretation. If the values were closer to unity, then there would be cause for a more detailed analysis in order to determine whether the pairs of types of state misinterpretation involved could be considered to be of the same type.

The main drawback of just using simple comparisons is that the comparisons are made between two types of state misinterpretation for which the number of possible causal sequences detected is often very different. If one examined the trees for the two types of state misinterpretation, the number of nodes in the two trees would be different. The standardised comparisons presented in the next section, take appropriate allowance of the differences in the size of trees.

9.2.2 Standardised Comparison of State Misinterpretation Types

The results of performing the standardised comparison calculations on all of the possible pairs of trees are shown in Table 9.3. The overall distribution of the results values using the standardised comparison calculations is slightly different to that for the simple comparisons, in that the number of values between 0.1 and 0.2 has fallen from eight to seven, whilst the number in the range 0.2 to 0.3 has risen from one to two. There are some minor variations in the distribution of the values for some of the individual comparisons too. So, for example, six of the values for *Cleared to Altitude <x>* now lie in the range 0-0.1, compared with seven for the simple comparison.

The minor change in the distribution for the individual comparisons reflects the changes to the individual comparison values, which are slightly different to those for the simple comparisons. The direction and magnitude of the change is not always the same, however. When judged against the simple comparison values, 29 of the values were reduced, six remained the same, and 20 increased. The size of the changes was relatively small, however, with the biggest reduction being 0.026, and the biggest increase being 0.036.

The standardised comparison does have the drawback that the method used for extending the smaller of the two trees is based on extending the existing tree structure in such a way that the basic structure of the tree is unaltered, but the number of instances of each of the sequences in the tree is increased. There is an underlying assumption here that the structure of the trees emerges early on, and that ensuing instances of possible causal sequences leave that structure unaltered. There is no good evidence to support or deny the truth of this assumption, and the truth probably lies somewhere in between. In other words, there is a core of sub-sequences which are common to many, but not necessarily all, of the possible causal sequences.

The other problem with the standardised comparison method is that it ignores the length of the sequences in the trees. In general, the data suggest that longer sequences of actions are more likely to end with specific consequences, than shorter sequences: longer sequences tend to offer a better explanation of the incident. It therefore makes sense to try and take account of this when comparing the trees, which is something that the weighted standardised comparison attempts to do.

Table 9.3 Results of standardised comparisons between pairs of types of state misinterpretation.

	Altimeter Setting OK	Climbing to altitude <x>	Cleared to altitude <x>	Cleared to land	Cleared to push by ground crew	Descending to cleared altitude <x>	Flying cleared route	Pos. in air short of action point	Pos. on ground short of action point	Rate of descent OK	RNAV frequency setting OK
Altimeter Setting OK		0.189	0.109	0.070	0	[redacted]	0.048	0.060	0.173	0.083	0.060
Climbing to altitude <x>			0.099	0.010	0.042	[redacted]	0.038	0.020	0.125	0.087	0.036
Cleared to altitude <x>				0.014	0.032	0.118	0.059	0.067	0.104	0.082	0.044
Cleared to land					0.024	0.039	0.034	0.053	0	0.029	0.079
Cleared to push by ground crew						0.008	0.027	0.074	0.024	0.024	0.031
Descending to cleared altitude <x>							0.048	0.039	0.132	0.087	0.039
Flying cleared route								0.078	0.013	0.041	0.080
Position in air short of action point									0	0.095	0.048
Position on ground short of action point										0.068	0.015
Rate of descent OK											0.048
RNAV frequency setting OK											

9.2.3 Weighted Standardised Comparison of State Misinterpretation Types

In Table 9.4 the results of performing the weighted standardised comparison calculations on all of the possible pairs of trees are shown. The overall distribution of the results values using the weighted standardised comparison calculations is slightly different from that for the other two types of comparisons. Compared to the standardised comparisons (Table 9.3), the number of comparisons yielding a zero value remains the same (and is the same three comparisons) as before. The number of comparison values in the range up to 0.1 has fallen slightly, however, to 42 (from 43); whilst the number of values in the range 0.2-0.3 has increased from 2 to 4. The number in the range 0.1-0.2 remained constant at seven.

The variation in the overall distribution (compared with the distribution of values for the simple and the standard comparisons) is a reflection of the changes in the underlying values. The magnitude and direction of the changes are not always the same, however, and there is no immediately obvious pattern to the changes. Compared with the standardised similarity values, nine of the values were reduced, six remained the same, and 40 increased. The size of the changes was relatively small, however, with the biggest reduction being 0.006, and the biggest increase being 0.029.

Table 9.4 Results of standardised weighted comparisons between pairs of types of state misinterpretation.

	Altimeter Setting OK	Climbing to altitude <x>	Cleared to altitude <x>	Cleared to land	Cleared to push by ground crew	Descending to cleared altitude <x>	Flying cleared route	Pos. in air short of action point	Pos. on ground short of action point	Rate of descent OK	RNAV frequency setting OK
Altimeter Setting OK		[illegible]	0.122	0.068	0	[illegible]	0.043	0.064	[illegible]	0.096	0.056
Climbing to altitude <x>			0.168	0.011	0.048	[illegible]	0.038	0.021	0.139	0.097	0.039
Cleared to altitude <x>				0.016	0.026	0.126	0.059	0.068	0.113	0.099	0.052
Cleared to land					0.030	0.037	0.032	0.064	0	0.033	0.083
Cleared to push by ground crew						0.007	0.032	0.080	0.022	0.030	0.034
Descending to cleared altitude <x>							0.048	0.036	0.146	0.094	0.040
Flying cleared route								0.079	0.016	0.049	0.086
Position in air short of action point									0	0.116	0.053
Position on ground short of action point										0.078	0.018
Rate of descent OK											0.055
RNAV frequency setting OK											

9.3 Characterising The Types of State Misinterpretation

The previous section proposed a scheme for quantitatively comparing types of state misinterpretation. It is also a useful exercise to try and quantify the relationship between individual antecedents and the error modes for the different types of state misinterpretation. Although the frequency of occurrence of individual antecedents has been described in chapter 7, the aim here is to begin to lay out how the relationship can be quantified. Once it becomes possible to quantify this relationship, it then becomes possible to start to consider methods for determining the possibility of occurrence of particular sequences of actions.

The QUORUM (Quantitative, Objective, Representative and Unambiguous Modeler) method of analysing incident reports (e.g. McGreevy, 1997) is based around performing textual analysis of the content of the narratives. As part of the set of available tools, QUORUM provides a relative measure of the distance relationship between two terms. The method proposed below is inspired by the general idea of the QUORUM relational metric value, which expresses the relationship between two terms as a function of the number of times that the two terms appear in the same sliding contextual window—in QUORUM this essentially equates to 16 adjacent words in the narrative text—and how close they appear together in that context. This general idea is extended here to apply it to encoded data, across possible causal sequences for the same type of incident.

The relational metric calculation used here is carried out as follows:

1. Work out the maximum length of the branches—possible causal sequences—in any of the trees in the complete set of incidents, i.e. across all types of state misinterpretation, and call this L_{max}.

2. For each antecedent that occurs in a possible sequence allocate a weight, such that if the antecedent is an immediate antecedent of an error mode *EM* it is assigned the weight L_{max}, if it is the antecedent of the antecedent, it is assigned the weight L_{max} - 1, and so on. (The minimum weight that can be allocated is 1.)

3. Calculate a weighted average for each error mode by summing the weights of the antecedents and dividing by the sum of the weights ($L_{max} + L_{max} - 1 + \ldots + 2 + 1$). This weighting ensures that the occurrence of an antecedent at a particular position always receives the same weighted value.

The final result provides a useful benchmark for comparing the different relationships for a single type of error mode within a single type of state misinterpretation. In order to make the results more generally applicable, the final result can be weighted using one (or both) of the following methods:

4. Weight the result by the total number of reports for that type of error mode within that type of state misinterpretation, so that different values can be compared for the same type of state misinterpretation.

5. Weight the result by the total number of reports for that type of state misinterpretation, so that different values can be compared across different types of state misinterpretation.

Generating all of the values for all of the possible relationships is a time consuming exercise, although it can at least be partially automated. Since this chapter is generally concerned with issues related to trees, the relationships were only calculated between the error modes and all their antecedents for each of the types of state misinterpretation. Accordingly, rather than exhaustively list out all of the values, the maximum and minimum values are summarised as separate tables below. The maximum values for the relationships are shown in Table 9.5; the minimum values are shown subsequently in Table 9.6. The results are sorted in the order of expected occurrence of the types of state misinterpretation during the course of a flight.

As a simple example, consider the case of the relationship between *Error Mode 4: Action of wrong type: distance* (*EM4*), and *Communication Failure* (*CO1*) for the state misinterpretation type *Cleared to altitude*. The maximum length of the branches (L_{max}) is 5, so the sum of the weights is 15. In 20 causal sequences *CO1* occurs as the first antecedent (and has an associated weight of 5), and in two sequences it occurs as the second antecedent (and has an associated weight of 4). So the unweighted value is calculated by:

$$\text{Unweighted value } (EM4, CO1) = (20*5) + (2*4))/ 15 = 7.2$$

which is the value shown in Table 9.5. The weighted values in the table are simply calculated by dividing the unweighted value by the number of reports for the appropriate error mode (within the selected type of state misinterpretation), or by the number of reports for the relevant type of state misinterpretation.

The three different values shown in the tables—unweighted, error mode weighted, and state misinterpretation weighted—reflect the different relationships between the action and the error mode. The general interpretation that should be applied to the maximum values in Table 9.5 is the same, however: the higher the value, the more closely related the action and the error mode are. In addition, for the results of the two weighted calculations, a higher value also generally represents the fact that the relation is more characteristic of either that particular error mode (within that type of state misinterpretation) or the type of state misinterpretation itself. This reflects the fact that the result of the calculation is proportional to the number of occurrences of that relation. It should be noted, however, that the error mode weighted value may need careful interpretation, because a high value may simply be the result of the fact that there are few instances of the relation.

Of the values listed in Table 9.5 three actions account for 28 of the 36 values listed: *Observation Missed* (*OB1*; 14); *Distraction* (*TP3*; 8); and *Communications Failure* (*CO1*; 6). These actions are also generally involved in the relationships which yield the highest values for each of the calculations for the different types of state misinterpretation. In other words, it appears that these three types of action are most typical of incidents where state misinterpretation occurs.

Out of the 11 types of state misinterpretation listed, *Observation missed* (6), *Distraction* (1), and *Communications failure* (2) are involved in the relationship which yields the highest state misinterpretation weighted value for nine of them. One of the exceptions is for *Flying Cleared Route* where *Hearing Failure* (*CO1.S11*) is the dominant action, although even this is a specific antecedent of *Communication Failure*. The only other exception is for *Position In Air Short of Action Point*, where *Adverse Weather Conditions* (*AC7*) is the dominant action.

Among the state misinterpretation weighted values, only eight of the 36 calculations yield a value of 0.1 or greater. Closer inspection of the data confirms that the (*error mode, action*) pair for each of these eight relationships is down to their dominance within the various types of incidents. In other words, the *error mode* in the pair was the most common one for that type of state misinterpretation, and the *action* was the most common action for that error mode. For the other three types of state misinterpretation, the data is generally more evenly distributed across more error modes.

Table 9.5 Maximum relative values for actions within each error mode and state misinterpretation type.

State Misinterpretation	Error Mode, Antecedent	Unweighted	Error Mode Weighted	State Misinterpretation Weighted
Cleared to push by ground crew	EM1,CO1	3.40	0.179	0.170
	EM4,OB3	0.33	0.333	0.017
Position on ground short of action point	EM1,PI1	0.33	0.333	0.009
	EM4,OB1	5.67	0.157	0.153
Flying cleared route	EM1,PL1	3.07	0.133	0.040
	EM2,CO1.S11	4.00	0.500	0.053
	EM4,OB1	2.33	0.179	0.031
	EM6,TP3	1.67	0.093	0.022
	EM7,TP3	1.33	0.095	0.018
Climbing to altitude <x>	EM1,CO1	1.33	0.222	0.027
	EM2,PR2	0.33	0.333	0.007
	EM4,OB1	10.00	0.238	0.204
Altimeter Setting OK	EM2,TP3	1.53	0.219	0.046
	EM4,OB1	6.00	0.214	0.158
	EM7,CO1	0.33	0.167	0.010
Cleared to altitude <x>	EM1,CO1	7.20	0.195	0.064
	EM2,OB1	1.00	0.167	0.013
	EM4,CO1	10.27	0.168	0.092
	EM7,CO1	1.33	0.167	0.012
Position in air short of action point	EM1,AC7	4.00	0.267	0.200
	EM2,TP3	0.60	0.300	0.030
	EM4,PL1	0.33	0.110	0.017
RNAV frequency setting OK	EM1,OB1	0.67	0.133	0.020
	EM2,TP3	1.27	0.141	0.038
	EM6,TP3	1.33	0.190	0.040
	EM7,OB1	2.00	0.182	0.061
Descending to cleared altitude <x>	EM1,TP3	1.33	0.267	0.020
	EM2,TP3	1.20	0.171	0.018
	EM4,OB1	15.33	0.279	0.229
Rate of descent OK	EM1,OB1	2.00	0.286	0.100
	EM4,OB1	1.33	0.267	0.067
	EM5,OB1	1.00	0.167	0.050
	EM7,OB1	0.33	0.167	0.017
Cleared to land	EM1,OB1	0.67	0.167	0.018
	EM2,OB1	5.67	0.177	0.153
	EM7,OB3	0.33	0.333	0.009

The minimum values for the error mode weighted calculations shown in Table 9.6 also require careful interpretation. Although a lower value may reflect that the error mode and action are less closely related, the value could also be attributed to the fact that there were simply less incidents reported. In addition, it is sometimes necessary to reference the maximum values too, because in some cases the minimum value is relatively high because the action in the (*error mode, action*) relationship pair is still fairly typical—although less typical than the maximum value—of that type of error mode or state misinterpretation.

The distribution of the values across actions is much more widely spread than that for the maximum values. None of the actions dominate the results for the relationships, with the most frequently occurring antecedent—*Parallel tasks*, *WC1.S1*—only being able to account for five out of the 36 relationships.

Only non-zero values are shown in Table 9.6. In other words, there is still some relationship between the action and error mode, albeit a relatively small one in most cases. The way in which the calculations work—based on the relative distance between the action and the error mode in a possible causal sequence—means that the action should invariably be the end point of a particular sequence. This is easily confirmed by the fact that in 30 of the 36 cases, the relationship involves a specific antecedent, i.e. an action with no antecedents (treating *Equipment failure*, *EF1*, as a specific antecedent in this instance). The other six cases can be confirmed by inspecting the sequence data directly.

In 27 of the 36 cases the state misinterpretation weighted value is less than 0.01. A closer inspection of the data for the remaining nine cases shows that the results relate to single instances of the relationship, and occurred for those four types of state misinterpretation which occurred less frequently than the others. In other words, the results may be a reflection of the fact that the types of state misinterpretation do not have a prototypical underlying structure. Alternatively, it may simply be that the underlying structure has not emerged out of the data due to a lack of instances of appropriate incidents.

Table 9.6 Minimum relative values for actions within each error mode and state misinterpretation type.

State Misinterpretation	Error Mode, Antecedent	Unweighted	Error Mode Weighted	State Misinterpretation Weighted
Cleared to push by ground crew	EM1,OR2	0.13	0.007	0.007
	EM4,OB3.S4	0.27	0.267	0.013
Position on ground short of action point	EM1,WC2	0.27	0.267	0.007
	EM4,PL1.S4	0.20	0.006	0.005
Flying cleared route	EM1,TP4.S1	0.20	0.009	0.003
	EM2,WC1.S8	0.20	0.025	0.003
	EM4,WC1.S1	0.20	0.015	0.003
	EM6,OB3.S4	0.27	0.015	0.004
	EM7,CO1.S4	0.20	0.014	0.003
Climbing to altitude <x>	EM1,EF1	0.20	0.033	0.004
	EM2,CO1.S10	0.20	0.200	0.004
	EM4,TP3.S8	0.20	0.005	0.004
Altimeter Setting OK	EM2,TP3.S6	0.27	0.038	0.008
	EM4,TP3.S8	0.20	0.008	0.006
	EM7,TP1	0.33	0.017	0.010
Cleared to altitude <x>	EM1,CO1.S3	0.20	0.005	0.002
	EM2,TP4.S2	0.13	0.022	0.002
	EM4,WC1.S1	0.20	0.003	0.002
	EM7,TP3.S6	0.20	0.029	0.003
Position in air short of action point	EM1,TP3.S7	0.20	0.013	0.010
	EM2,WC1.S1	0.20	0.100	0.010
	EM4,PL1.S4	0.27	0.089	0.013
RNAV frequency setting OK	EM1,TP3.S4	0.20	0.040	0.006
	EM2,TP4.S1	0.20	0.022	0.006
	EM6,OR2	0.20	0.029	0.006
	EM7,OR1	0.13	0.012	0.004
Descending to cleared altitude <x>	EM1,WC1.S1	0.20	0.040	0.003
	EM2,TP4.S1	0.27	0.038	0.004
	EM4,WC1.S1	0.13	0.006	0.005
Rate of descent OK	EM1,TP3.S6	0.20	0.029	0.010
	EM4,TP3.S4	0.20	0.040	0.010
	EM5,TP3.S7	0.20	0.033	0.010
	EM7,TP3.S4	0.20	0.100	0.010
Cleared to land	EM1,CO2.S3	0.20	0.050	0.005
	EM2,WC1.S2	0.20	0.006	0.005
	EM7,OB3	0.33	0.333	0.009

9.4 Summary

In this chapter three different ways of comparing the trees of possible casual sequences that make up each of the different types of state misinterpretation have been described. The first involves visually comparing automatically generated pictures of the structure of the trees. Whilst this can give a useful general impression about the similarity of two types of state misinterpretation it becomes more difficult as the number of branches in the trees increases. The second way involves quantifying the similarity by counting the occurrence of sequences, and then weighting the data to take account of frequency of occurrence, and the length of the sequences. The final way is to generate a relational metric value which was used to quantify the relationship between the error modes and antecedents for each type of state misinterpretation.

The analysis of the data is now complete, and all that remains is to draw together the individual threads of the research. The final chapter pulls together these threads, generating some overall conclusions, and suggesting ways in which the research can be fruitfully extended.

10 Concluding Remarks

In this final chapter, a recapitulation of the research is presented. The work is assessed to measure how well the following original aims were met:

- General aim: To investigate the role of state misinterpretation in the operation of complex systems.
  - Specific aim 1: To investigate the general role played by the system state in the operation of complex systems, and use this, in combination with known examples of state misinterpretation to formulate a definition of the concept.
  - Specific aim 2: To use the definition of state misinterpretation as a basis for gathering data on instances of state misinterpretation. A taxonomy of types of state misinterpretation shall be developed by categorising the data appropriately.
  - Specific aim 3: To investigate the underlying causes surrounding state misinterpretation. The goal here is to try and establish if there any common causes which can be identified. Once any common features or patterns of behaviour have been identified it should be possible to start to determine how the various types of state misinterpretation can be prevented, possibly by automation, and which types give rise to situations which have to be managed by the operators.

The success in meeting the general aim can be measured by the combined level of success for the three specific aims. The results of the research have to be interpreted in the light of the strengths and weaknesses of the use of incident data as a method for performing human factors evaluations. These caveats are briefly revisited as a precursor to considering how the aims have been met, with varying success, in three parts. First, the contributions to the field of research are described. Second, the results of the data analysis are discussed in terms of what can be done to manage state misinterpretation, and how the findings relate to existing research. Third, consideration is given to ways in which the work can be improved and extended in the future.

10.1 Caveats

There are particular strengths and weaknesses to using incident data in human factors evaluations. These need to be borne in mind when interpreting the results of this research, and can be summarised as follows (Chappell, 1994). The strengths of using incident data are:

- The data is provided by the participants who have first hand experience of what happened.
- Large numbers of observations: the ASRS database receives an average of more than 2000 incident reports per month.
- Ecological validity: the incidents described in the reports happened in the real working environment.

Second, the weaknesses of using incident data are:

- The data is not validated: the level of confidentiality in the system may preclude any further investigation into reported incidents.
- Reporter biases, in terms of both who reports, and which incidents get reported. People need to be aware of the reporting system, have access to it, and be motivated to report incidents in the first place. If people perceive an incident as insignificant, they may not report it, and media interest in accidents in the domain can heighten awareness of particular issues in potential incident reporters. Furthermore, people can only report incidents if they detect that an error has occurred.

Altogether 123 out of the 476 incident reports analysed here were categorised with an error mode that involved an altitude deviation. There is a predominance of altitude deviation incident reports in the ASRS data base. This particular bias results from the fact that pilots receive an indemnity from prosecution by the FAA if they voluntarily report any such incidents, as long as there is no deliberate or malicious act involved. The net effect is that pilots are keen to report all such incidents in order to gain immunity from any possible further action, such as fines or other penalties.

The overall approach has not been fully validated by measuring the reliability of the coding scheme using an inter-rater reliability measure across different coders. The results should therefore be interpreted with some caution, since only the intra-rater reliability of the coding scheme has been tested. It should also be noted that the time and effort involved in learning to use the approach advocated here is relatively large. The main reason for this is the need to acquire all of the appropriate kinds of knowledge—knowledge about the software tools, the aviation domain, the CREAM-AIR, and human error—that are needed to successfully use this approach.

10.2 Contributions to the field

This research has provided four basic contributions to the field. The first is a theory of state misinterpretation, encapsulated as a working definition. The second is a taxonomy of the different types of state misinterpretation. The third is a method for the causal analysis of ASRS incident reports, which is an adaptation and specialisation of the CREAM (Hollnagel, 1998a). The fourth is a set of tools that support the method for analysing the ASRS incident reports. Each of these is considered below.

10.2.1 A theory of state misinterpretation

State misinterpretation is simply a failure in state interpretation. In other words, it is a failure to correctly interpret the state of the system. For the purposes of this research, the state of the system is that which is *perceived* by the operator. The perceived state of the system is determined by whether the operator knows which variables define the current state of the system, whether the operator accesses the values of all of those variables, and whether the operator interprets those values correctly. An analysis of these three factors was used to formulate the definition that: State misinterpretation occurs when the operator knows which values define the current state of the system, but does not access all of them, and may or may not interpret the values correctly.

In general the term *system state* is used in the literature to described the system state as perceived by the operator. However, systems nowadays are no longer simply considered to be the computer system which the operator controls, but as socio-technical systems which are actually distributed between machines, people and organisations. The corollary of this is that the system state must also be considered to be similarly distributed. Any part of the system state which is not directly observable from the instrumentation attached to the machine component of the system therefore has to be stored elsewhere. In the aviation domain, for example, transient instructions from ATC—clearances—form a part of the system state which cannot be directly observed on the cockpit instrumentation. The flight crew either have to transcribe these instructions, or keep them in their head.

In addition to transient components that affect the overall system state, there are other components which change more slowly but are equally influential in determining the system state. Typical examples are company operating procedures, and regulatory manuals which prescribe what the flight crew should do in particular situations. Such manuals can be read by the flight crew and the relevant parts incorporated into the system state. There are some parts of the system state which cannot readily be made observable, however, such as latent failures in the system (Reason, 1990), or cultural factors which influence flight crew behaviour (Helmreich & Merritt, 1999).

The definition of state that is used as the basis for state interpretation here (and hence state misinterpretation) does not preclude the inclusion of human, and environmental factors. There is a general shortcoming, however, in that the incident reports can only describe failures of which the flight crew are cognisant (e.g. Sarter and Woods, 1991). If the flight crew do not know about the importance or relevance of environmental or organisational factors, for example, they will not be able to include information about these factors in their reports. This general problem may be partly addressed by the introduction of human factors training for flight crews, which some airlines are already carrying out.

10.2.2 A taxonomy of state misinterpretation

Having identified and retrieved a set of reports in which state misinterpretation was implicated, a taxonomy of different types of state misinterpretation was developed. Initially each of the incident reports was categorised as belonging to a particular type of state misinterpretation. Once this had been done, the different types of state misinterpretation were analysed to see if there were any similarities between them. Those which were deemed to be very closely related, i.e. seemed to describe the same thing, were combine together into a single type.

The validity of the taxonomy was later measured by comparing all of the types of state misinterpretation in a pairwise manner. This is described in more detail below.

10.2.3 A method for the causal analysis of ASRS incident reports

The causal analysis of the ASRS incident reports was performed using the CREAM-AIR. This is a simplified version of the CREAM (Hollnagel, 1998a), which was adapted to the analysis of incident reports (rather than accident reports) and extended to take account of the particular requirements of human performance in the aviation domain. The CREAM-AIR is implemented as a software application, which makes application of the method more reliable and robust.

The data analysis goes beyond the standard CREAM, in that the results of the retrospective analysis are considered at three different levels of abstraction. At the lowest level the data were considered in terms of individual error modes and actions. At the middle level the data were considered in terms of sequences of actions and error modes. At the highest level the data were considered in terms of the different types of state misinterpretation which can be represented as trees (or networks) of possible causal sequences.

There are some general limitations to how far the CREAM can be usefully extended. The tables of categories and links that underpin the CREAM is a strength of the method, but could ultimately prove to be a weakness too. The CREAM is intended to be generic, but if it is to be applicable to all domains there is a danger that the size of the categories and the number of links between consequents and antecedents may increase to such an extent that retrospective analysis becomes unmanageable. The basic process that underpins the CREAM is a search for the attributed causes, which relies on making a match between data in the incident report and the antecedents in the CREAM categories. As the number of elements in the categories and the number of links between consequents and antecedents increases, the complexity of the search process increases because there are more possibilities to be considered at each level of the search.

10.2.4 A set of tools and techniques for analysing ASRS incident reports

An overview of the software tools appears in Appendix B. These tools fall into two basic types:

- Tools to overcome ASRS database limitations
- Tools to support the data analysis.

The CD-ROM version (97-2) of the ASRS database used here runs as a DOS application. There are some limitations imposed by features inherent in DOS (such as the DOS file naming convention), and others that are imposed by the ASRS database query form which limits the number of characters that can be used to query individual fields. In general, the problems of file naming and query field sizes have not been mentioned in ASRS studies. This is probably because most studies which utilise the ASRS database as the data source are concerned with a single specific attribute. The research here, however, is of a more general nature, and is more data driven. The queries therefore have to take account of the fact that different reporters will describe the same situation using different phraseology. This is a general problem, exacerbated by the fact that some equipment manufacturers use different names for the same piece of equipment (Billings, 1997).

The first type of tools were used to overcome the query field size limitations, by effectively combining the results of individual queries for individual dimensions of the narrative field outside of the ASRS database. The accession numbers of the reports in the combined results were then used to extract incident reports from the database.

The second type of tools are made up of a tool that implements the CREAM-AIR, and tools that can be used to process the results of using the CREAM-AIR. The CREAM-AIR implementation bounds the search for antecedents, in such a way that only valid antecedents can be selected, and the user simply has to select one from a list. The implementation also allows the CREAM-AIR to be dynamically reconfigured, such that any new antecedents or links which are discovered during the analysis of the data can be incorporated *in situ*, without having to close the tool, and restart the analysis.

The other data analysis tools provide support for the three different levels of abstraction: actions, sequences and trees. These tools are used for analysing the basic structure of the data, so that any underlying patterns can be detected.

In addition to the software tools, two techniques were developed to analyse the data. The first was a technique for comparing two causal trees to quantitatively measure how closely related the trees are. The second was a technique for measuring the relationship between two actions (or an action and an error mode) for a particular type of state misinterpretation.

10.3 Assessment of the Findings

The major findings of this research are presented below. The findings are discussed with reference to existing research in this area. Finally, some suggestions are made about how the results can be converted into positive recommendations about how to manage incidents in which different types of state misinterpretation are implicated.

10.3.1 The findings

At the level of individual actions and error modes, the most significant finding was that the different types of error modes were dominated by the CREAM-AIR error mode *Action of wrong type: Distance/magnitude (EM4)*. This finding reflects the domination of so-called altitude deviations in the ASRS database; Chappell (1994) noted that at that time more than one third of the total number of ASRS incident reports for the preceding six years related to altitude deviations. Although the proportion here is closer to a half, this could simply be the result of the fact that the population of reporters used here is more restricted than that used by Chappell (1994).

At the level of possible causal sequences, the main result is the prevalence of monitoring failures across the different types of state misinterpretation. In particular, the most frequently occurring sequence of actions was one in which a monitoring failure was caused by a distraction which was attributed to a competing task. In most cases the competing tasks were directly relevant to the general task of flying the aircraft. The new tasks generally captured the attention of one or more members of the flight crew which distracted them from the immediate task of continuing to fly the aircraft or providing backup for the other member of the flight crew.

At the level of trees, the complexity of the different types of state misinterpretation was made clear by the pictorial representation of the trees. Incidents in which state misinterpretation occurs do not fall cleanly into nice neat structures but manifest themselves in several different ways. The quantitative measure for comparing trees offered a level of support for the taxonomy of the different types of state misinterpretation developed here. The highest similarity value between two trees was 0.291 for the types of *Climbing to altitude* <*x*> and *Descending to altitude* <*x*>, which can be considered on one level to be counterparts.

10.3.2 Expertise and state misinterpretation

Although there is a school of thought that suggests that expert behaviour arises with about 1000 hours experience, the situation may not be quite so clear cut here. When the ASRS incident report includes a figure for the level of experience, this is measured by the total number of hours spent flying aircraft. It is not clear whether the time includes on the job training, and the figure does not discriminate between time spent as a first officer, and time spent as a captain. The way that the aviation industry operates also means that there is no time limited transition point which determines when a first officer becomes a captain. A first officer can normally only become a captain when a captain's seat becomes available, and the first officer has met all the other specified criteria.

The caveats outlined above may mean that the selected incident data is not be quite as homogeneous as it initially appears. In particular the 1000 hour limit may be too low, and the data may therefore include flight crew members who, whilst not complete novices, could not be considered to be experts either. Out of the total of 493 analysed incidents, the minimum reported number of hours experience is 1000 hours, and the maximum is 35000 hours. The distribution of the reported number of hours experience is skewed towards the higher number of flight hours, however, as shown in Table 10.1. So, even if the selection criteria for expertise is raised to five years (which is equivalent to 4000 hours, given that the maximum number of hours a pilot can fly in a year is limited to 800), only 24 of the 493 reported incidents (4.87%) would

Table 10.1 Number of flight hours' experience of crew member reporting the incident.

No. of Hours Experience	No. of Incidents
1000-1999	5
2000-2999	9
3000-3999	10
4000-4999	13
5000-5999	29
6000-6999	26
7000-7999	28
>8000	373
Total	493

fall below the new threshold. Indeed, in 373 out of the 493 reported incidents (75.7%) the pilots had the equivalent of 10 years or more experience (i.e. over 8000 hours).

10.3.3 Managing incidents involving state misinterpretation

For the most part, state misinterpretation appears to occur during the busiest phases of the flight. The fact that it appears to manifest itself in so many different ways means that there is no single panacea for preventing incidents in which state misinterpretation is implicated. There are, however, some general ways in which such incidents could be managed.

The first possible way of managing incidences of state misinterpretation is to introduce new checklists, or in appropriate cases add extra items to existing ones. The highly proceduralised nature of aviation means that certain aspects can be encapsulated using checklists. The flight crews already have several checklists that they are supposed to run through at particular points in the flight. A case could possibly be made for adding extra checklists to counter some of the problems of state misinterpretation. The downside of this is that checklists are often used to ensure that critical actions have been performed at particular points in the flight. Adding more checklists may simply add to flight crew workload at a time when they are already busy. From the data analysis, it is also not clear when such checklists should be performed either.

The second possible way of managing state misinterpretation relates directly to problems of determining what constitutes system state. At present most flight deck systems deal in terms of things that can be measured and displayed. They do not provide a means of augmenting the memory of the flight crew for recording verbal instructions from ATC. Although good Crew Resource Management can help, when things get busy in the cockpit the flight crew tend to focus on those tasks that they routinely perform at the same time or the same stage of the flight.

Following checklists is a good example. So there could be a case made for having slots in the checklist that can be dynamically filled in, such as the value of the frequency for contacting tower which the flight crew have to fill in when they receive it from ATC, and possibly the point at which they should make the switch. There are two potential problems here. The first, as noted above, is that the flight crew may already be heavily loaded anyway. The second is that the flight crew would have to remember to refer back to the checklist in order to determine when to contact tower, and on what frequency. In other words, it would require a change of philosophy in the use of checklists.

The third possible way of managing state misinterpretation is to provide some form of training which shows the problems of monitoring and distractions by competing tasks. The flight crew would be taught to develop strategies for handling such situations, such that when one member of the crew had to deal with a competing task, the other member of the crew gets made aware of the potential consequences. In some ways this is an extension of the CRM principle, in which the workload of flying the aircraft gets divided across the whole of the flight crew, whilst the captain remains in overall charge. Effectively what is required is a meta-level analysis of the situation in the cockpit, such that the flight crew are aware that a competing task has arisen.

The fourth possible way to manage state misinterpretation would be to allow the computer system to take over. There is a major difficulty here, however, in that the computer system could only detect the occurrence of a new competing task, and a distraction from monitoring if it was appropriately informed by the flight crew.

Although none of these approaches to managing state misinterpretation is a panacea, the need to recognise that there is a problem that has to be addressed still exists. The current state of affairs utilises ATC almost as a last line of defence. The fact that ATC have a general view of traffic in the areas, and know the route that each aircraft should be following means that they should be able to pick up on any deviations that may arise, and take appropriate action.

10.3.4 Comparison with existing classification schemes

The results of this research suggest that distractions recur as a causal factor in incidents where a state misinterpretation had occurred, and in several such cases the distraction was caused by a competing task. In Chapter 2, it was noted that there are two existing classification schemes that explicitly include misinterpretation. The results from using the CREAM-AIR are briefly compared and contrasted with the results of using the existing schemes.

Although Rasmussen's (1982) classification scheme may have uncovered the role of distractions, it would not have detected the source of the distraction. The Rasmussen classification suggests that a human malfunction (or error) is the end result of an accidental sequence of events. One of the elements in the sequence is a category that Rasmussen describes as *Causes of Human Malfunction* which explicitly includes distractions, albeit as an example of an external event. The classification scheme does not provide any method for identifying or determining the cause of the distraction, however. It also does not make clear how different instances of operator errors should be classified. The CREAM-AIR provides solutions to each of these shortcomings of Rasmussen's scheme. The cause of distractions can be determined using the CREAM-AIR (e.g. in several of the incidents that were analysed, the distraction was caused by a competing task). The CREAM-AIR tables also provide an inherent method for classifying every instance of an operator error.

Rouse and Rouse's (1983) taxonomy of errors would not directly have uncovered the influence of distractions as a causal factor. In their taxonomy there is an explicit general category of human error labelled *Observation of System State*. Associated with this general category are six specific categories of human error (*excessive*, *misinterpreted*, *incorrect*, *incomplete*, *inappropriate*, and *lack*), but there is no explicit mechanism for determining the underlying cause of any of these specific categories. Some guidelines are offered, however, for defining the potential contributing classes of factors, and the class labelled *Contributing events* does include *distractions* as one of its elements. Rouse and Rouse note the caveat that the degree to which the contributing factors are present is generally highly subjective. The potential problems of subjectivity are reduced by the CREAM-AIR classification scheme which only allows consequents and antecedents to be linked in certain specific ways, as defined by the CREAM-AIR tables.

10.3.5 Related Research

Although there appears to be no other research which has looked directly at the problems of state misinterpretation, there is some work which has also reported that distractions on the flight deck do cause problems. The importance of managing distractions have been reported elsewhere by Latorella (1996), who performed an experimental investigation into interruptions on the flight deck using a flight simulator. Latorella (1996) uses the term *interruptions* in its most general sense to include distractions.

The other important work on distractions in aviation was reported by Monan (1979; Barnes & Monan, 1990). In an analysis of the 2500 reports collected by the ASRS up to 1979, Monan

(1979) reported that 169 had a distraction as a significant cause. The distractions fell into two categories: operational, and non-operational. The former were routine flight deck tasks which, when performed at the wrong time, increased the workload of the flight crew, such as performing checklists, and looking at charts. The latter were tasks which were not directly related to the task of flying the aircraft, such as contacting the company on the radio, and making passenger announcements. It was the work of Monan (1979) which gave rise to the FAA's sterile cockpit rule, which is intended to prevent flight crew members from distracting each other from their duties during critical phases of the flight. Barnes and Monan (1990), however, found that even after the introduction of this rule, distractions are still being cited as a cause of incidents.

The research presented here is in line with the findings of both Latorella (1996), and Monan (1979; Barnes & Monan, 1990), albeit arrived at from a different angle. This research has used a more data driven approach, in which patterns of behaviour have been looked at for a particular type, rather than starting from the premise that distractions are a causal factor.

One possible explanation for distractions still being regarded as a significant problem, albeit a slightly speculative one, is that perhaps the situation in the cockpit is approaching capacity. The resources of the flight crew are close to full utilisation during busy periods, so they may not have enough resources available to appropriately deal with distractions. Some of the reports in the ASRS database allude to this problem, pointing out that things in the cockpit can be fairly hectic at times, and the addition of extra workload—an extra task, for example—almost inevitably leads to an incident.

10.4 Future work

There are several directions in which this research can be extended. Some of these are described below.

The fact that there are different identifiable types of state misinterpretation lends credence to the concept of state misinterpretation. What has not been assessed by the current research, however, is the correctness of the concept, which would be difficult to measure, partly due to the fact that it is not a directly observable phenomenon. One step towards assessing the correctness of the concept would be to look at whether the patterns of behaviour noted here are unique to incidents in which state misinterpretation is implicated. It could be the case that problems surrounding incidents in which state misinterpretation is implicated are of a more general nature, rather than specifically applicable to state misinterpretation. This could be investigated by analysing ASRS incident reports for the same period, but without using any selection criteria for the narrative field when determining which reports to export for analysis.

One very useful way of extending this work would be to derive a base level version of the CREAM. As it stands, the CREAM reflects the fact that it was developed on the basis of experience of performing research in the process industries. The CREAM-AIR has built on this, and succeeded in adding several new antecedents and links to the base level CREAM. Superficially, at least, it seems likely that there will be a core of antecedents and links between consequents and antecedents that are independent of the domain in which the CREAM is used. By having a base level CREAM which can be extended into a domain specific version, the process of retrospectively analysing the incident reports should be made slightly easier, at least, because the categories of antecedents, and the links would always be directly applicable to the domain. In other words, the search process would only ever consider relevant antecedents.

The identification, extraction, and analysis of the incident reports is a time consuming process. It would be useful if the process could be improved, without losing any of the robustness of the approach. One way of improving the process would be to integrate the software tools together, so that the analyses of the data could be conducted more quickly. It would also be useful to update the software tools, ideally to make them work directly with the ASRS database, which would speed up the process of identification and extraction of reports.

Perhaps the most useful way in which the work could be extended is to take the approach and apply it to another domain. There is some evidence that the sort of problems that arise from the use of automation in the cockpit in aviation, are also present in the maritime domain (e.g. Dekker & Woods, 1999). By applying the approach to maritime operations, it would help to identify the strengths and weaknesses of the concept of state misinterpretation.

10.5 Summary

In this thesis the role of state misinterpretation in the operation of complex systems has been investigated. Ultimately the research was restricted to the aviation domain, largely because that was the domain for which there was data readily available, and in which there is a history of such research to build upon. The research was driven initially by the concept of state misinterpretation, and then data driven as an attempt was made to try and identify patterns of behaviour that were common across instances of incidents in which state misinterpretation was implicated. The research has started to address some of the issues recently raised by Maurino (2000), in that it looks at the causal process surrounding state misinterpretation.

State misinterpretation does appear to exist as a real problem, and there do appear to be different types of state misinterpretation within the aviation domain. Several tools and techniques have been developed to support the identification, extraction and analysis of incident reports from the ASRS incident database. The data suggest that there are some sequences of behaviour which are more strongly associated with state misinterpretation than others. The fact that such patterns have been identified suggests that incidents involving state misinterpretation can be managed in order to mitigate the potential consequences.

11 References

Aeroknowledge (1994). *ASRS CD ROM User's manual*, Version 1.2. Trenton, NJ: Aeroknowledge.

Amalberti, R. (1998).Why operators' cognitive models are hard to incorporate into design? The case of human reliability models. Keynote address, *2nd European Conference on Cognitive Modelling*. Nottingham, UK.

Ashby, W.R. (1956). *An introduction to cybernetics*. London, UK: Chapman and Hall.

Bagnara, S., Di Martino, C., Lisanti, B., Mancini, G., & Rizzo, A. (1989). *A human error taxonomy based on cognitive engineering and on social and occupational psychology (EUR 12624 EN)*. Ispra, Italy: JRC.

Bainbridge, L. (1983). The ironies of automation. *Automatica, 19*, 775-780.

Bainbridge, L. & Sanderson, P.M. (1995). Verbal Protocol Analysis. In J. Wilson and E.N. Corlett (Eds.), *Evaluation of human work* (Second Edition) (pp.169-201). London, UK: Taylor & Francis.

Barnes, V.E. & Monan, W.P. (1990). Cockpit distractions - Precursors to emergencies. In *Proceedings of Human Factors Society 34th Annual Meeting*, pp.1142-1144.

Billings, C.E. (1997). *Aviation automation: The search for a human-centered approach*. Mahwah, NJ: LEA.

Boy, G.A. (1987). Operator assistant systems. *International Journal of Man-machine Systems, 27*, 541-554.

Brazendale, J. (1990). *Human errors in risk assessment*. London, UK: Health & Safety Executive.

Bruford, D. (1994). *The air pilot's glossary and reference guide*. Shrewsbury, UK: Airlife Publishing.

Cantù, M. (1997). *Mastering Delphi 3*, Second edition. Alameda, CA: Sybex.

Chappell, S. L. (1994). Using voluntary incident reports for human factors evaluations. In N. Johnston, N. McDonald, & R. Fuller (Eds.), *Aviation psychology in practice* (pp.149-169). Aldershot, UK: Avebury.

De Keyser & Woods, D.D. (1990). Fixation errors: Failures to revise situation assessment in dynamic and risky systems. In A.G. Colombo and A. Saiz de Bustamente (Eds.), *System reliability assessment* (pp.231–251). Dordrecht, The Netherlands: Kluwer Academic.

Dekker, S., & Woods, D.D. (1999). Automation and its impact on human cognition. In S. Dekker & E. Hollnagel (Eds.), *Coping with computers in the cockpit.* Aldershot, UK: Ashgate.

Dougherty, E. (1990). Human Reliability Analysis—Where shouldst thou turn? *Reliability Engineering and System Safety, 29,* 283-299.

Ericsson, K.A., & Simon, H.A. (1993). *Protocol analysis: Verbal reports as data* (Second Edition). Cambridge, MA: MIT Press.

FAA (1998). *The federal aviation facility operation and administration handbook,* 7210.3P. Updated 26 February 1998.

Faith, N. (1996). *Black box: Why air safety is no accident.* London, UK: Boxtree.

Feltovich, P.J., Spiro, R.J., & Coulson, R.L. (1993). Learning, teaching, and testing for Complex Conceptual Understanding. In N. Frederiksen, R.J. Mislevy & I.I. Bejar (Eds.), *Test theory for a new generation of tests* (pp.181-217). Hillsdale, NJ: LEA.

Frese, M., & Altmann, A. (1989). The treatment of errors in learning and training. In L. Bainbridge and S.A. Ruiz Quintanilla (Eds.), *Developing skills with information technology* (pp.65-86). Chichester, UK: Wiley.

Green, R. (1990). Human error on the flight deck. *Phil. Transactions. of the Royal Society of London, Series B,* 503-512.

Groeneweg, J. (1994). *Controlling the controllable.* Leiden, the Netherlands: DSWO Press.

Habberley, J. S., Shaddick, C. A., & Taylor, D. H. (1986). *A behavioural study of the collision avoidance task in bridge watchkeeping* (Project Report No. CON/840/35/31). Southampton, UK: The College of Maritime Studies.

Hale, A.R., Karczewski, F., Koornneef, F., & Otto, E. (1991). IDA: An Interactive Program for the Collection and Processing of Accident Data. In T.W. van der Schaaf, D.A. Lucas & A. Hale (Eds.), *Near miss reporting as a safety tool,* (pp.83-98). Oxford, UK: Butterworth-Heinemann.

Heinrich, H.W., Petersen, D., & Roos, N. (1980). *Industrial accident prevention: A safety management approach,* 5th edition. New York, NY: McGraw Hill.

Helmreich, R.L., & Merritt, A.C. (1998). *Culture at work in aviation and medicine.* Aldershot, UK: Ashgate.

Hollnagel, E. (1993a). The phenotypes of erroneous actions. *International Journal of Man-Machine Studies, 39,* 1-32.

Hollnagel, E. (1993b). *Human reliability analysis: Context and control.* London: Academic Press.

Hollnagel, E. (1998a). *The Cognitive Reliability and Error Analysis Method.* Oxford, UK: Elsevier Science Ltd.

Hollnagel, E. (1998b). Context, cognition and control. In Y. Wærn (ed.), *Co-operative process management* (pp.27-52). London: Taylor & Francis.

Hollnagel, E., Drøivoldsmo,A. & Kirwan, B. (1996). Practical insights from studies of operator diagnosis. In *Proceedings of Eighth European Conference on Cognitive Ergonomics,* pp.133-137.

HSE (2000). *Train accident at Ladbroke Grove junction, October 1999 (Paddington).* 3rd interim report. London, UK: Health and Safety Executive. (Available on-line at *www.hse.gov.uk/railway/paddrail/interim3.htm.*)

Hutchins, E. (1995). *Cognition in the wild.* Cambridge, MA: MIT Press.

Ives, G. (1991). "Near Miss" Reporting Pitfalls for Nuclear Plants. In T.W. van der Schaaf, D.A. Lucas & A. Hale (Eds.), *Near miss reporting as a safety tool* (pp.51-66). Oxford, UK: Butterworth-Heinemann.

Johnson, C.W., McCarthy, J.C. & Wright, P.C. (1995). Using a formal language to support natural language in accident reports. *Ergonomics, 38,* 1265-1283.

Johnson, P.E., Moen, J.B., & Thompson, W.B. (1988). Garden path errors in diagnostic reasoning. In L. Bolc and M.J. Coombs (Eds.), *Expert system applications*(pp.395-427). Berlin, Germany: Springer-Verlag.

Kelley, C.R. (1968). *Manual and automatic control.* New York, NY: Wiley.

Klein. G.A. (1997). The Recognition-Primed Decision (RPD) model: Looking back, looking forward. In C.E. Zsambok & G.A. Klein (eds.), *Naturalistic decision making* (pp.285-292). Mahwah, NJ: Lawrence Erlbaum Associates.

Klein, G.A., Orasanu, J., Calderwood, R., & Zsambok, C.E. (1993). *Decision making in action: Models and methods.* Norwood, NJ: Ablex.

Kletz, T. (1991). *An engineer's view of human error*, 2nd edition. Rugby, UK: Institution of Chemical Engineers.

Kletz, T. (1993). *Lessons from disaster: How organisations have no memory and accidents recur.* Rugby, UK: Institution of Chemical Engineers.

Kletz, T. (1994). *Learning from accidents,* 2nd Edition. Oxford, UK: Butterworth-Heinemann.

Kletz, T. (1999). *HAZOP and HAZAN: Identifying and assessing process industry standards* (Fourth edition). Rugby, UK: Institution of Chemical Engineers.

Landsberg, B. (1998). Landmark accidents: Cleared for the approach. *AOPA Pilot,* 41, 6, 77-80.

Latorella, K. (1999). Investigating interruptions: Implications for flightdeck performance. NASA Technical Memorandum TM-1999-209707. Langley, VA: NASA.

McGreevy, M.W. (1997). *A practical guide to interpretation of large collections of incident narratives using the QUORUM method.* NASA Technical Memorandum 112190. Moffett Field, CA: NASA.

Mach, E. (1905). *Knowledge and error.* Dordrecht, The Netherlands: Dreidel. (English translation, 1976.)

Mangold, S.J., Morrison, F.R., & Frank, S.M. (1995). The Use of ASRS Incident Reports in AQP Training. In R.S. Jensen and L. Rakovan (Eds.), *Proceedings of 8th International Symposium on Aviation Psychology.* pp.1390-1394.

Masson, M. (1991). Understanding, Reporting and Preventing Human Fixation Errors. In T.W. van der Schaaf, D.A. Lucas & A. Hale (Eds.), *Near miss reporting as a safety tool* (pp.35-50). Oxford, UK: Butterworth-Heinemann.

Maurino, D. (2000). Human factors and aviation safety: What the industry has, what the industry needs. *Ergonomics, 43*, 7, 952-959.

Monan, W.P. (1979). *Distraction - A human factor in air carrier hazard events.* NASA Technical Memorandum 78608. Moffett Field, CA: NASA.

Moray, N. (1987). Intelligent aids, mental models, and the theory of machines. *International Journal of Man-Machine Studies, 27*, 619-629.

Nagel, D.C. (1988). Human error in aviation operations. In E.L. Wiener and D.C. Nagel (Eds.), *Human factors in aviation*, (pp.263-303). San Diego, CA: Academic Press.

Nardi, B.A. (1996). *Context and consciousness: Activity theory and human-computer interaction.* Cambridge, MA: MIT Press.

Neisser, U. (1976). *Cognition and reality.* San Francisco, CA: Freeman.

Newell, A. (1990). *Unified theories of cognition.* Cambridge, MA: Harvard University Press.

Norman, D.A. (1981). Categorization of action slips. *Psychological Review, 88*, 1-15.

O'Hare, D., Wiggins, M., Batt, R., & Morrison, D. (1994). Cognitive failure analysis for aircraft accident investigation. *Ergonomics, 37*, 1855-1869.

Orlady, H.W., & Orlady, L.W. (1999). *Human factors in multi-crew flight operations.* Aldershot, UK: Ashgate.

Perrow, C. (1984). *Normal accidents.* New York, NY: Basic Books.

Petroski, H. (1985). *To engineer is human: The role of failure in successful design.* New York, NY: St. Martin's Press.

Pew, R.W., Miller, D.C., & Feeher, C.E. (1981). *Evaluation of proposed control room improvements through analysis of critical operator decisions.* Cambridge, MA: Bolt, Beranek & Newman.

Rasmussen, J. (1976). Outlines of a hybrid model of the process operator. In T.G. Sheridan and G. Johannsen (eds.), *Monitoring behavior and supervisory control.* 371-383. New York: Plenum.

Rasmussen, J. (1980). What can be learned from human error reports? In K. Duncan, M. Gruneberg, and D. Wallis (Eds.), *Changes in working life* (pp.97-113). Chichester, UK: Wiley.

Rasmussen, J. (1982). Human errors: A taxonomy for describing malfunction in industrial installations. *Journal of Occupational Accidents, 4* (2-4), 311-333.

Rasmussen, J. (1983). Skills, rules, knowledge: signals, signs and symbols and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-13*, 257-267.

Rasmussen, J. (1986). *Information processing and human-machine interaction: An approach to cognitive engineering*. London, UK: Elsevier.

Rasmussen, J. (1988). Human error mechanisms in complex work environments. *Reliability Engineering and System Safety, 22* ,155-167.

Rasmussen, J., Pejtersen, A-M., Goodstein, L.P. (1994). *Cognitive systems engineering*. Chichester, UK: Wiley.

Reason, J. (1979). Actions not as planned: The price of automatization. In G. Underwood & R. Stevens (Eds.), *Aspects of consciousness, Volume 1: Psychological issues* (pp.67-89). London: Wiley.

Reason, J. (1990). *Human error.* Cambridge, UK: Cambridge University Press.

Reason, J. (1991). Too Little and Too Late: A Commentary on Accident and Incident Reporting Systems. In T.W. van der Schaaf, D.A. Lucas & A. Hale (Eds.), *Near miss reporting as a safety tool* (pp.9-26). Oxford, UK: Butterworth-Heinemann.

Reason, J. (1997). *Managing the risks of organisational accidents.* Aldershot, UK: Ashgate.

Reason, J. & Mycielska, K. (1982). *Absent-minded? The psychology of mental lapses and everyday errors.* Englewood Cliffs, NJ: Prentice Hall.

Reynard, W.D., Billings, C.E., Cheaney, E.S., & Hardy, R. (1986). *The development of the NASA Aviation Safety Reporting System*. Reference Publication 1114. Moffett Field, CA: NASA.

Sanderson, P.M. & Fisher, C. (1994). Exploratory sequential data analysis: Foundations. *Human-Computer Interaction, 9*(3), 251-317.

Sanderson, P.M. & Fisher, C. (1997). Exploratory sequential data analysis: Qualitative and quantitative handling of continuous observational data. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics*, 2nd edition (pp.1471-1513). Wiley: New York, NY.

Sanderson, P.M., Scott, J.J.P., Maintzer, J., Johnston, T., & James, J.M. (1994). MacSHAPA: A software environment for ESDA. *International Journal of Human-Computer Studies, 41(5)*, 633-681.

Sarter, N.B., & Woods, D.D. (1991). The Flight Management System: Rumors and facts. In R.S. Jensen (Ed.), *Proceedings of 6th International Symposium on Aviation Psychology*. pp. 241-246.

Sarter, N.B., & Woods, D.D. (1995). How in the world did we ever get into that mode? Mode error and awareness in supervisory control. *Human Factors*, 37(1), 5-19.

van der Schaaf, T.W. (1991a). Introduction. In T.W. van der Schaaf, D.A. Lucas & A. Hale (Eds.), *Near miss reporting as a safety tool* (pp.1-8). Oxford, UK: Butterworth-Heinemann.

van der Schaaf, T.W. (1991b). A Framework for Designing Near Miss Management Systems. In T.W. van der Schaaf, D.A. Lucas & A. Hale (Eds.), *Near miss reporting as a safety tool* (pp.27-35). Oxford, UK: Butterworth-Heinemann.

van der Schaaf, T.W. (1992). *Near miss reporting in the chemical process industry*. Ph.D. Thesis, Eindhoven University of Technology, the Netherlands.

van der Schaaf, T.W., Lucas, D.A., & Hale, A.R. (1991). *Near miss reporting as a safety tool*. Oxford, UK: Butterworth-Heinemann.

Senders, J.W. & Moray, N.P. (1991). *Human error: Cause, prediction and reduction*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Stewart, M.G., & Melchers, R.E. (1997). *Probabilistic risk assessments of engineering systems*. London, UK: Chapman and Hall.

Taylor, R.K. & Lucas, D.A. (1991). Signals Passed at Danger: Near Miss Reporting from a Railway Perspective. In T.W. van der Schaaf, D.A. Lucas & A. Hale (Eds.), *Near miss reporting as a safety Tool,* (pp.99-114). Oxford, UK: Butterworth-Heinemann.

Thomas, R.E., & Rosenthal, L.J. (1982). *Probability distributions of altitude deviation.* NASA Contractor Report 16639. Moffett Field, CA: NASA.

Toft, B., & Reynolds, S. (1994). *Learning from disaster: A management approach.* Oxford, UK: Butterworth-Heinemann.

Tukey, J.W. (1977). *Exploratory data analysis.* Reading, MA: Addison-Wesley.

Wagenaar, W.A. & Groeneweg, J. (1987). Accidents at sea: Multiple causes and impossible consequences. *International Journal of Man-Machine Studies, 27*, 587-598.

Weir, G.R.S. (1991). Living with complex interactive systems. In G.R.S. Weir and J. L. Alty (Eds.), *Human-computer interaction and complex systems*, (pp.1-21). London, UK: Academic Press.

Wickens, C.D., & Flach, J.M. (1988). Information Processing. In E.L. Wiener and D.C. Nagel (Eds.), *Human factors in aviation* (pp.111-155). San Diego, CA: Academic Press.

Wiegmann, D.A. & Shappell, S.A. (1997). Human factors analysis of postaccident data: Applying theoretical taxonomies of human error. *International Journal of Aviation Psychology, 7*, 1, 67-81.

Wiener, E.L., Kanki, B.G., & Helmreich, R.L. (1993). *Cockpit Resource Management.* San Diego, CA: Harcourt Brace Jovanovich.

Wioland, L. & Amalberti, R. (1998). Human error management: Towards and ecological safety model. A case study in an air traffic microworld. In *Proceedings of Ninth European Conference on Cognitive Ergonomics*, pp.91-96.

Woods, D.D. (1984). Some results on operator performance in emergency events. *Institution of Chemical Engineers Symposium Series, 90*, 21-31.

Woods, D.D., Johannesen, L.J., Cook, R.I., & Sarter, N.B. (1994). *Behind human error: Cognitive systems, computers, and hindsight.* State of the Art Report CSERIAC SOAR 94-01. Dayton, OH: CSERIAC.

Zsambok, C.E., & Klein, G.A. (1997). *Naturalistic decision making*. Mahwah, NJ: Lawrence Erlbaum Associates.

12 Appendix A: Definitions of ASRS Incident Categories

The FAA Facility Operation & Administration Handbook, 7210.3P is the official source used to categorise reported incidents which are entered into the ASR database. The definition of the three categories—Pilot Deviation, Operational Error, and Operational Deviation—used during this research are as follows:

12.1 Pilot Deviation

The actions of a pilot that result in the violation of a FAR or a North American Aerospace Defense Command (NORAD)/Air Defense Identification Zone (ADIZ) tolerance.

12.2 Operational Error

An occurrence attributable to an element of the air traffic system which:

1. Results in less than the applicable separation minima between two or more aircraft, or between an aircraft and terrain or obstacles, as required by FAAO 7110.65, Air Traffic Control, and supplemental instructions. Obstacles include vehicles/equipment/personnel on runways; or

2. Aircraft lands or departs on a runway closed to aircraft operations after receiving air traffic authorization.

12.3 Operational Deviation

A controlled occurrence where applicable separation minima, as referenced in [the definition of Operational Error], but:

1. Less than the applicable separation minima existed between an aircraft and protected airspace without prior approval.

2. An aircraft penetrated airspace that was delegated to another position of operation or another facility without prior coordination and approval.

3. An aircraft penetrated airspace that was delegated to another position of operation or another facility at an altitude or route contrary to the altitude or route requested and approved in direct coordination or as specified in a LOA, pre-coordination or internal procedure.

4. An aircraft, vehicle, equipment, or personnel encroached upon a landing area that was delegated to another position of operation without prior consideration or approval.

13 Appendix B: Software Support Tools

Two types of software tools were developed to support the research described in the main part of this thesis. The first type is the applications that were required to overcome the particular limitations of the ASRS database. In particular, the tools were developed to overcome the limitations on the size of the fields that are provided on the ASRS database query form. The tools that fell into this category were:

- Perceived State Dimension Combination.
- ASRS Query Generator.

The second type is those tools that were required to generate the human behaviour data from the selected incidents, and subsequently support the analysis of that data. The tools that fell into this category were:

- CREAM coder.
- Concordance and sequence generator.
- Causal Tree generator.

In general, both the design and the implementation of the tools were a compromise between generality, ease of use and pragmatism. The overriding criterion was to produce software tools that would support the analysis of incident reports from the ASRS database.

The tools were all implemented using version 3 of the Inprise Delphi software development tool, running under Windows 95 on an IBM PC compatible machine. Each of the tools is briefly described below.

13.1 Perceived State Dimension Combination Tool

The Perceived State Dimension Combination Tool is used to overcome the limitations on the size of the query field for the narrative section in the ASRS database. The Perceived State Dimension Combination Tool consists of 331 lines of Delphi code, and the user interface is implemented using two forms: the main form, and a dialog to change the source directory for the input files. The main form is shown in Figure 13.1.

The tool, which is specific to the ASRS database, takes as input three files of ASRS report numbers—one file for each dimension of the perceived state—and creates a single output file

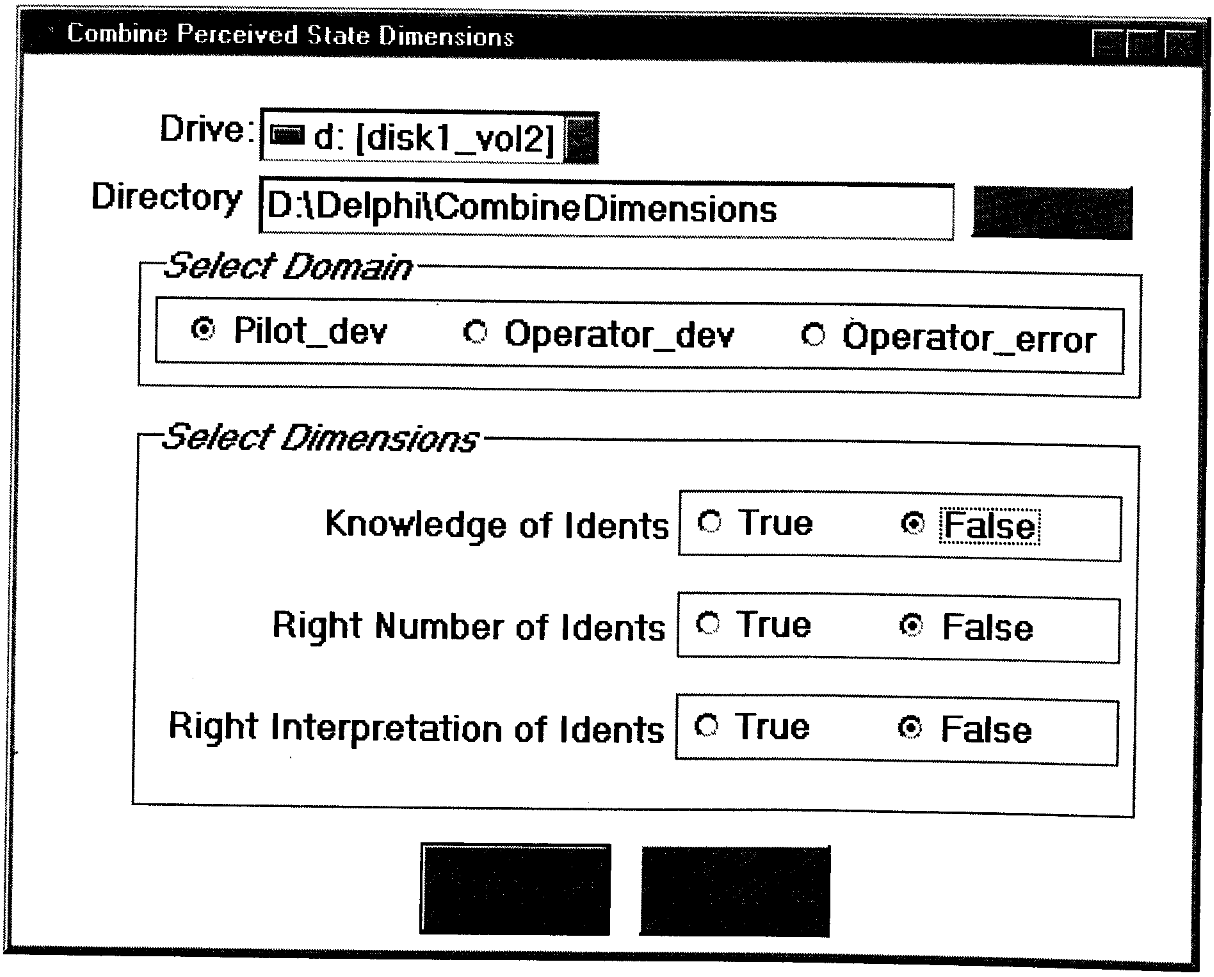

Figure 13.1. Screen shot of the main form for the user interface for the Perceived State Dimension Combination Tool.

which contains those report numbers which appear in all three of the input files. The naming convention used for the input files and the output files is hard coded into the tool. The user selects which input files to use by selecting the type of reports required (Pilot Deviation, Operator Deviation or Operator Error) and then selecting the appropriate combination of dimensions. The radio button settings are used internally to produce an eight character file name for the output file.

13.2 ASRS Query Generation Tool

The ASRS Query Generation Tool is used to automatically generate queries to the ASRS database taking into account the maximum length of the accession number field on the database query form. The consists of 107 lines of code, with a user interface that is implemented using a single form as shown in Figure 13.2.

The tool is specific to the ASRS database. The input file should contain a set of ASRS incident report numbers, one per line, saved as plain text. The *Accession nos. per line* determines how

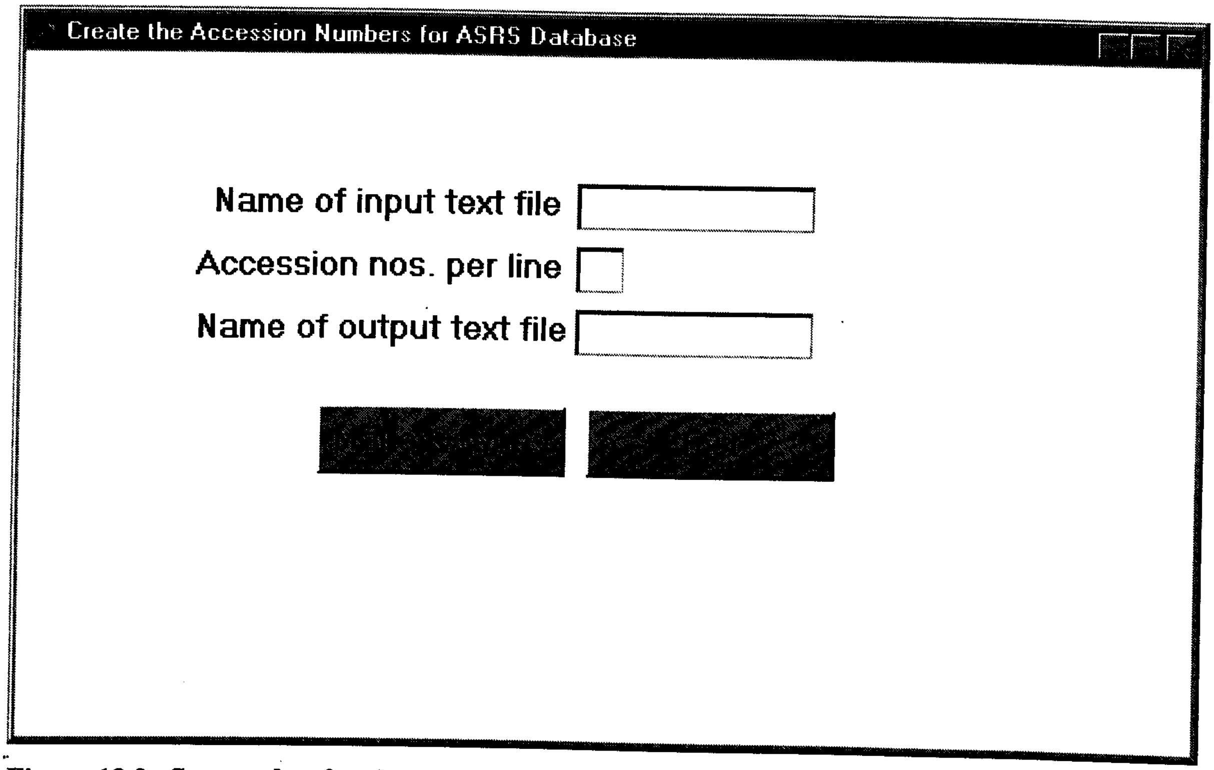

Figure 13.2. Screen shot for the user interface to the ASRS Query Generation Tool.

many incident report numbers will appear in the ASRS database query file that the tool produces. The output file name specifies the general form of output file name. Each output file is structured as a database query that can be loaded and run by the ASRS database. The tool will produce as many database queries as are required. So, for example, if there are 75 report numbers in the input file, 20 accession numbers per line are required in the output file, and the output filename is specified as "Pilots", the tool would produce four files named "Pilots1.dwq", through "Pilots4.dwq".

13.3 CREAM-AIR Coding Support Tool

The CREAM-AIR Coding Support Tool implements the CREAM as adapted and extended for the retrospective analysis of ASRS incident reports in the aviation domain. The tool consists of 1164 lines of Delphi code, with a user interface that is implemented by a single form as shown in Figure 13.3. In addition, the tables that underpin the CREAM-AIR are all implemented as text files which act as input to the tool. The links between the tables are encapsulated in the way that the text files are structured and the way that they are used by the code.

The key features that underpin the CREAM-AIR methodology are:

- The tables which define the general and specific effects of the error modes (or phenotypes), and the general and specific consequents of the different types of causes (or genotypes).
- The tables which define the links between the classification groups. In other words, the tables which relate the general consequents to general and specific antecedents.

Each of these sets of tables is encapsulated in the CREAM-AIR coding support tool. The effects/consequents definition tables (which are listed in Appendix C) provide an abstract level of definition for the possible instances. So, for example, in the definition table for the genotype *Observation*, the general consequent *Observation Missed*, has associated with it two specific consequents. The definition of the first of these is "A signal or an event that should have been the start of an action (sequence) is missed." (which defines the overlooking of a cue or signal). The coding support tool automatically displays the definitions on the screen, rather than the names of the specific consequents, because it is not always easy to discriminate between the specific consequents just by using their names. In this way the coding support tool makes it somewhat easier to discriminate between consequents that have similar names. The coding support tool also automatically generates the corresponding shorthand code for the selected consequents.

During the analysis of a particular incident, when the user initially selects a particular error mode, the coding support tool automatically presents the user with a list of definitions for the specific effects associated with that error mode, and determines the corresponding shorthand code (such as *EM1*). In addition, the software automatically searches for all the possible antecedents for that error mode, and presents them as a list from which the user can select the most appropriate antecedent. The selected antecedent then becomes the consequent for the next step of the causal analysis, and the process is repeated—selecting the definition, and the antecedent from the generated lists—until no more antecedents can be found. The selection of an antecedent for a particular consequent (or error mode) is therefore made easier by the software, because the user does not have to find all of the possible alternatives by manually searching through the CREAM-AIR tables: the search for the antecedents is performed dynamically by the coding support tool, based on the last consequent selected by the user.

The CREAM-AIR linkage mechanism is built into the software tools in such a way that the user can only select antecedents for the last selected consequent from a list of valid alternatives. An antecedent is only regarded as a valid alternative if there is a link between the currently selected consequent and that antecedent defined in the CREAM-AIR tables. It is unlikely that all of the possible consequent-antecedent relationships for a domain can be specified prior to the analysis of a set of incident reports, however. For this reason, the CREAM-AIR coding support tool allows the user to add novel consequent-antecedent relations to the CREAM-AIR tables when they are uncovered by the analysis of the incidents. The new tables can then be loaded into the coding support tool on the fly, without the need to do any reprogramming of the software, or even rebuilding it. In this way, the CREAM-AIR tables evolve over time to more fully represent those aspects of the domain that are specific to aviation. The introduction of new antecedents or definitions of specific consequents does not mean that any of the existing analyses need to be redone, however, because the new data is only introduced when it is first encountered. Allowing the CREAM-AIR tables to be extended in this way, means that it should be possible to encode any ASRS incident report using the coding support tool.

In addition to providing shorthand coded sequences for each of the possible causal sequences for an incident, the coding support tool also generates a more readable version of the possible causal sequence(s) of events. These generalised textual descriptions of the incidents are held in a separate text file from the coded sequences. The basic format of these descriptions closely mirrors what the user sees on the screen when using the CREAM-AIR coding support tool (see Figure 13.3, for example). The only difference is that the shorthand code is shown at the start of the line in the text file, rather than at the end of the line in Figure 13.3.

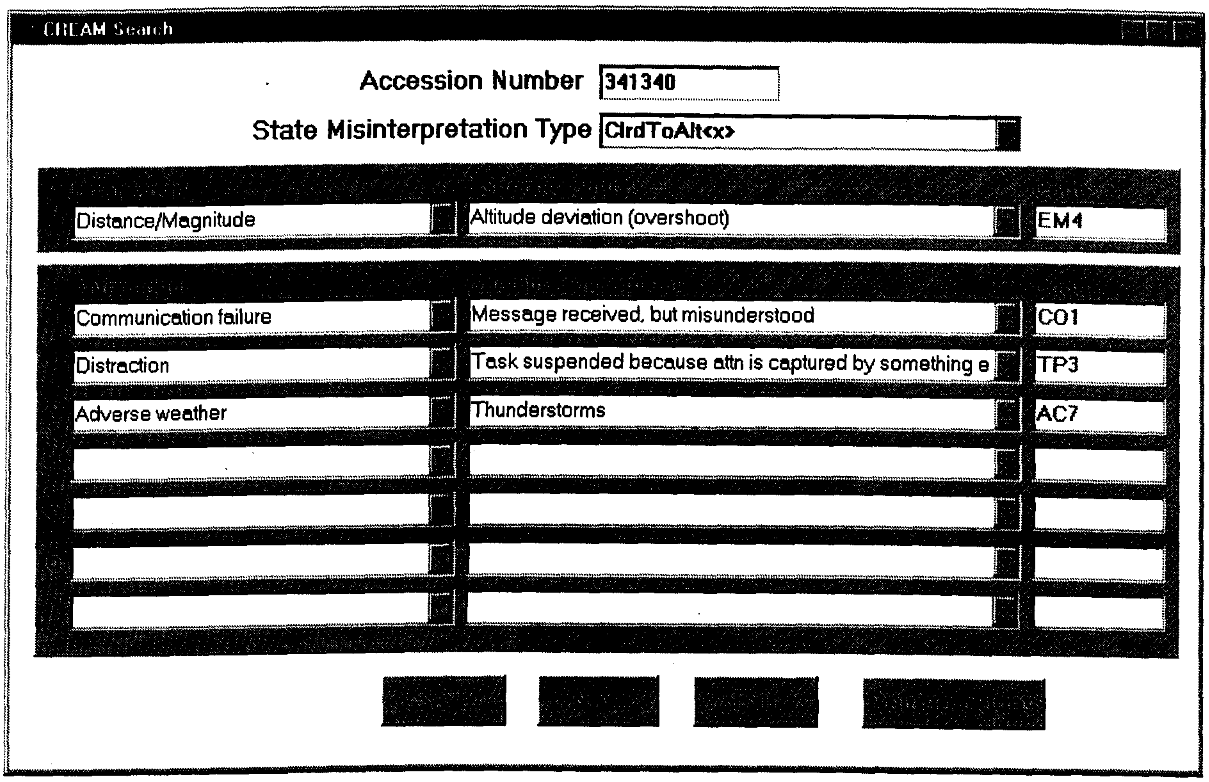

Figure 13.3. Screen shot for the user interface for the CREAM-AIR coding support tool, with the fields filled in for ASRS Incident Report Number 341340.

The user enters the report number in the text box at the top of the page, and selects the appropriate type of state misinterpretation, or adds a new type where appropriate in the box immediately below it. Filling in the form is then simply a matter of selecting the desired option from each of the drop down boxes, and clicking on the save button. The options shown on the drop down boxes are determined by the CREAM-AIR tables and links, so only valid combinations of consequents and antecedents are allowed. If a novel combination is found in an incident report, however, this can be added to the tool by simply updating the appropriate input files, and reloading them.

The tool produces two files, one of which contains encoded causal sequences of error modes and actions, and the other of which contains a normalised description of these sequences, based on the definition for each of the specific consequents and error modes. Both of the output files are text files that can be imported into Excel for further processing as appropriate.

13.4 Concordance and Sequence Generator Tool

The Concordance and Sequence Generator Tool takes the sequences that are produced by the CREAM-AIR Coding Support Tool, and can generate Concordance listings for actions and error modes, and frequency counts for sequences too. The tool consists of 671 lines of Delphi code, and has a user interface that is implemented by a single form, as shown in Figure 13.4.

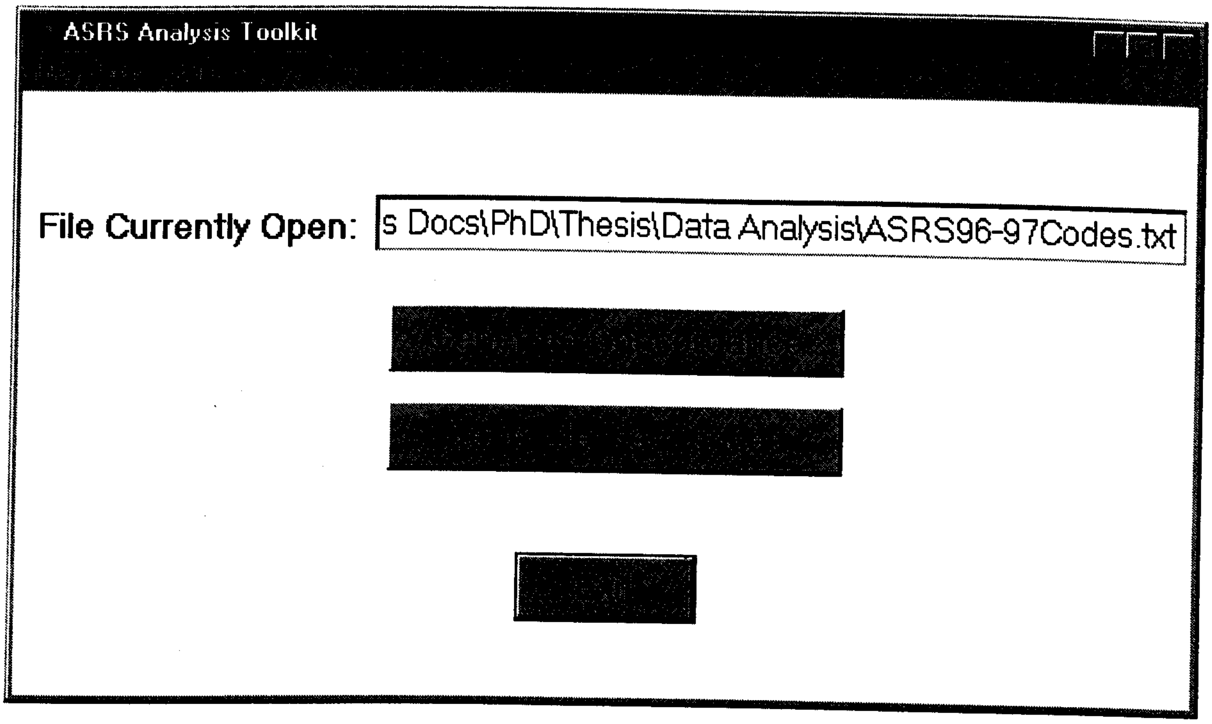

Figure 13.4. Screen shot for the user interface of the Concordance and Sequence Generation Tool.

The tool takes as input a file of coded possible causal sequences generated by using the CREAM-AIR tool described above. If the user clicks on the button labelled *Generate Concordance*, the tool analyses all of the sequences in the input file and produces an output file that contains the frequency of occurrence of each of the error modes and actions that appear in the input file. Alternatively, if the user clicks on the button labelled *Generate Sequences*, the tool analyses all of the sequences in the input file and produces an output file that contains the frequency of occurrence for each of the sequences.

13.5 Causal Tree Generator Tool

The Causal Tree Generator Tool produces a tree of the causal sequences for a particular type of state misinterpretation. The format of the tree is the same as that used in the Microsoft Explorer tool. The Causal Tree Generator Tool is an extension of the ChapTree program in Cantù (1996), and contains 385 lines of Delphi code. The user interface is implemented using a single form, as shown in Figure 13.5.

The user specifies the input file, which should be a text file produced by the CREAM-AIR Coding Support Tool described above. The tool analyses the sequences of coded data in the input file, and produces a generic tree structure in which the sub-sequences of the sequences in the input file appear as branches on the tree. Only a single instance of each sub-sequence

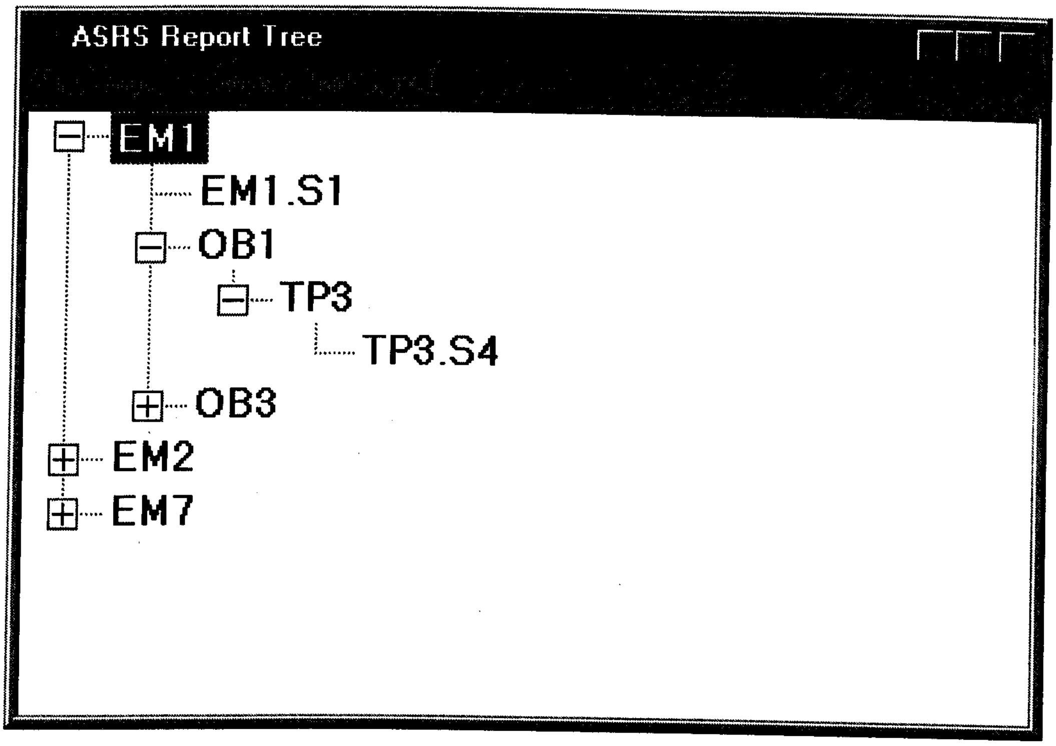

Figure 13.5. Screen shot for the user interface of the Tree Generator Tool, shown with the sub-tree for Error Mode 1 fully expanded.

appears in the tree.

Various options are provided which allow the user to manipulate various aspects of the displayed tree, such as collapsing and expanding branches and sub-branches on the tree, and changing the size of the font used on the display.

14 Appendix C: CREAM-AIR Tables

This appendix shows how the CREAM-AIR shorthand codes are related to the consequents and antecedents that are identified during the analysis of the ASRS incident reports. There are two sets of tables, one for the general effects and consequents, and one for the specific antecedents. These are listed separately below.

14.1 General Consequents

The tables in this section are used within the CREAM-AIR coding support tool (described in Appendix B) to generate the codes for particular instances of a general effect or consequent for a particular incident. Each of the tables shows the CREAM-AIR shorthand codes, the general effects (for the error modes) or consequents (for the causal actions), and the instances of the effects or consequents. The instances directly correspond to the specific effects or consequents.

Table 14.1 General effects for Error Modes

Code	General Effect	Instances
EM1	Timing/ Duration	Action started before a signal or required conditions. Action started too late. Action not performed in time allowed. Action continued past point where it should have stopped. Action stopped before it should have.
EM2	Sequence	Action was not carried out (especially the last action). One or more actions in a sequence were skipped. One or more earlier actions are carried out again. Previous action is repeated. Two adjacent actions are reversed. An extraneous/irrelevant action is carried out.
EM3	Force	Insufficient force. Too much force/effort.
EM4	Distance/ Magnitude	Movement too far. Movement not far enough. Altitude deviation (overshoot). Altitude deviation (undershoot).
EM5	Speed	Action too quick or finished too early. Action too slow or finished too late. Speed deviation (too fast/slow).
EM6	Direction	Movement in wrong direction. Wrong kind of movement (e.g. pulling instead of turning).
EM7	Wrong Object	Action on object physically close to correct object. Action on object that looks like the correct object. Action on object unrelated to the correct object.

Table 14.2 General consequents for Observation category

Code	General Consequent	Instances
OB1	Observation missed	Triggering signal/event for action sequence is missed. A measurement/information is missed. An action/event is missed (monitoring).
OB2	False observation	Response made to incorrect stimulus or event. Event/information is mis-recognised or mistaken for something else.
OB3	Wrong identification	A signal/cue is misunderstood as something else. Identification of an event/information is incomplete. Deliberate identification of an event/information is incorrect.

Table 14.3 General consequents for Interpretation category

Code	General Consequent	Instances
IN1	Faulty diagnosis	Diagnosis of situation/state is incorrect. Diagnosis of situation/state is incomplete.
IN2	Wrong reasoning	Reasoning used specific to general inferences/generalisations. Reasoning used general to specific inferences/generalisations. Selection among alternatives using wrong criteria.
IN3	Decision error	Inability to make a decision. Making wrong decision (about action alternatives). Decision does not completely specify what to do.
IN4	Delayed interpretation	Identification not made in time for action to be taken. Identification is not fast enough leading to time pressure.
IN5	Incorrect prediction	Unanticipated state change. Side effects of event overlooked. Speed of development of system is faster/slower than expected.
IN6	Misinterpretat ion	Misinterpreted the information presented.

Table 14.4 General consequents for Planning category

Code	General Consequent	Instance
PL1	Inadequate plan	Plan does not contain all details required to carry it out. Plan will not achieve its purpose. Plan based on incorrect information. Plan contains confusing/ambiguous information. Plan cannot be achieved in time available.
PL2	Priority error	Plan will not be effective because wrong goal was selected

Table 14.5 General consequents for Temporary Person Related category

Code	General Consequent	Instance
TP1	Memory failure	Item/information cannot be recalled when required. Information is incorrectly recalled. Information is only partly recalled.
TP2	Fear	Actions do not appear to follow any plan or principle. Person is unable to move or act.
TP3	Distraction	Task suspended because attention is captured by something else. Task is not completed because of a shift in attention. Person cannot remember why task/action is being done. Person cannot remember what is next or what has happened.
TP4	Fatigue	Person's response is reduced due to fatigue.
TP5	Performance variability	Reduced precision of actions. Actions increasingly fail to achieve their purpose.
TP6	Inattention	Signal/event missed (at random) due to inattention.
TP7	Physiological stress	Physiological stress.
TP8	Psychological stress	Psychological stress.

Table 14.6 General consequents for Permanent Person Related category

Code	General Consequent	Instance
PP1	Functional impairment	Deafness. Bad eyesight. Colour blindness. Dyslexia/aphasia. Other disability.

Code	General Consequent	Instance
PP2	Cognitive style	Simultaneous search for information/data. Serial search for data/information. Serial search of aspects based on an assumption.
PP3	Cognitive bias	Search for data/information changes opportunistically. New information fails to lead to an adjustment of probabilities. Interpretation of past events influenced by current outcome. Events mistakenly attributed to specific factors. Chosen actions mistakenly believed to control the system. Search restricted to data confirming current assumptions. Search restricted by strong theory about current problem.

Table 14.7 General consequents for Equipment Failure category

Code	General Consequent	Instance
EF1	Equipment failure	Autopilot failed to capture altitude. Autothrottle failure. Control cannot be moved/moves too easily. Something is obstructing the performance of an action. FMS problem. Generator malfunction. Stall warning. Altitude alerting system failed to chime or light up. Autopilot fails to engage. Landing gear malfunction. Auto pressurization controller failure. Bleed valve failure. Oil bypass light continually coming on. Flap/slat/speed brake problem. Hydraulics leak. Autopilot inoperative. Mechanical problems. OMEGA units problem. ADF inoperative. VOR/DME failure. RNAV failure. ILS out of action for selected runway. Radio communications/headset failure.

Code	General Consequent	Instance
EF2	Software fault	System performance slows down. Delays in transmission of information. System is unstable, so commands are stacked. Information unavailable due to software or other problems.

Table 14.8 General consequents for Procedure category

Code	General Consequent	Instance
PR1	Inadequate procedure	Procedure is ambiguous. Procedure is incomplete and assumes specific knowledge. Procedure is factually incorrect. Procedure does not match equipment. Procedure causes conflict of interest.
PR2	Procedure violation	Flight crew broke sterile cockpit rule. Improper use of phraseology (ATC). Improper use of phraseology (Flight crew). FO initiated Captain's manoeuvre. Flight crew failed to adhere to clearance. Action omitted from checklist. Failed to comply with altitude awareness procedure. Failed to use recommended NAV mode. Ground crew failed to follow push back procedures. Flight crew broke local procedure re visual approach clearance.

Table 14.9 General consequents for Temporary Interface category

Code	General Consequent	Instance
TI1	Access limitations	Item is temporarily out of reach. Item cannot be located or is temporarily unavailable. Item is temporarily obscured.
TI2	Ambiguous information	Mismatch between indicated and actual positions of an item. Mismatch in coding (e.g. colour or shape). Similar callsigns. Obsolete squawk code (from previous flight).
TI3	Incomplete information	Information provided by the interface is incomplete.

Table 14.10 General consequents for Permanent Interface category

Code	General Consequent	Instance
PI1	Access problems	Item is permanently out of reach. Item is permanently difficult to find. Item is located on captain's side of panel. Items located too close together.
PI2	Mislabelling	Labelling/identification of an item is incorrect. Labelling/identification of an item is ambiguous. Labelling/identification of an item is incorrectly formulated.

Table 14.11 General consequents for Communication category

Code	General Consequent	Instance
CO1	Communication failure	Message did not reach intended recipient. Message received, but misunderstood. No communication took place. Message transcribed incorrectly. Incorrect information communicated.
CO2	Missing information	Information not given when expected (e.g. missing feedback). Information is incorrect or incomplete. Misunderstanding between sender and receiver. Required information is not readily available.

Table 14.12 General consequents for Organisation category

Code	General Consequent	Instance
OR1	Maintenance failure	Equipment is unserviceable due to missing/inappropriate management. Indicators (lights etc.) failed due to lack of maintenance.
OR2	Inadequate quality control	Equipment/function inadequate due to lack of QC. Lack of resources/supplies.
OR3	Management problem	People unclear about roles and duties. No clear distribution of responsibility. Line of command is not well defined.
OR4	Design failure	Working environment inadequate due to design failure. Interface is inadequate due to design failure.
OR5	Inadequate task allocation	Organisation of work deficient due to lack of clear rules/principles. Task planning/scheduling is deficient. Procedures for carrying out work are inadequate.

Code	General Consequent	Instance
OR6	Social pressure	Individual understanding is guided/controlled by group.

Table 14.13 General consequents for Training category

Code	General Consequent	Instance
TR1	Insufficient skills	Lack of skills to perform task. Incorrect use of equipment through lack of skills.
TR2	Insufficient knowledge	Person uncertain what to do due to lack of knowledge. Person has lost general understanding due to lack of knowledge.

Table 14.14 General consequents for Ambient Conditions category

Code	General Consequent	Instance
AC1	Temperature	Too hot. Too cold.
AC2	Sound	Too loud. Too soft.
AC3	Humidity	Too dry. Too humid.
AC4	Illumination	Too bright. Too dark.
AC5	Other	Vibration.
AC6	Adverse ambient conditions	Wake turbulence. Mountainous terrain. Heavy boat traffic.
AC7	Adverse weather	Thunderstorms. Lightning. Severe icing. Turbulence. High winds. Bright sunshine. Snow. Freezing rain. Fog. Hot and humid. Drizzle and mist. Freezing fog.

Table 14.15 General consequents for Working Conditions category

Code	General Consequent	Instance
WC1	Excessive demand	Excessive task demands or lack of time/resources.
WC2	Inadequate workplace layout	Work space too cramped. Dangerous working conditions. Danger of falling down from workplace. Controls not replicated for both pilots. Inadequate signage. Inadequate gate sign lighting.
WC3	Inadequate team support	Roles within team not well defined/understood. Distribution of work/responsibility not mutually agreed. Little cohesiveness/collaboration in the team.
WC4	Irregular working hours	Psychological/physiological problems of shift work.
WC5	Inadequate staffing levels	Insufficient staff to perform task correctly.

14.2 Specific Antecedents

The tables in this section are used within the CREAM-AIR coding support tool (described in Appendix B) to generate the codes for the specific antecedents—those antecedents which have no further antecedents—in a particular incident. Each of the tables shows the CREAM-AIR shorthand codes, and the specific antecedent. Where no specific antecedents have been identified, the shorthand code for the general consequent is shown in italic text, and the specific instance is listed as *None defined.*

Table 14.16 Specific antecedents for Error Modes

Code	Specific Instance
EM1.S1	Earlier omission.
EM1.S2	Trapping error.
EM2.S1	Trapping error.
EM3.S1	Ambiguous label.
EM3.S2	Convention conflict.
EM3.S3	Incorrect label.
EM4.S1	Ambiguous label.
EM4.S2	Convention conflict.
EM4.S3	Incorrect label.
EM4.S4	Earlier omission.
EM5	*None defined.*
EM6.S1	Ambiguous label.

Code	Specific Instance
EM6.S2	Convention conflict.
EM6.S3	Incorrect label.
EM7.S1	Ambiguous label.
EM7.S2	Incorrect label.

Table 14.17 Specific antecedents for Observation category

Code	Specific Instance
OB1.S1	Information overload.
OB1.S2	Multiple signals.
OB1.S3	Noise.
OB1.S4	Parallax.
OB2	*None defined.*
OB3.S1	Ambiguous signals.
OB3.S2	Ambiguous symbol set.
OB3.S3	Erroneous information.
OB3.S4	Habit expectancy.
OB3.S5	Information overload.

Table 14.18 Specific antecedents for Interpretation category

Code	Specific Instance
IN1.S1	Confusing symptoms.
IN1.S2	Erroneous analogy.
IN1.S3	Error in mental model.
IN1.S4	Misleading symptoms.
IN1.S5	Mislearning.
IN1.S6	Multiple disturbances.
IN1.S7	New situation.
IN2.S1	False analogy.
IN2.S2	Mode error.
IN2.S3	Overgeneralisation.
IN2.S4	Too short planning horizon.
IN3.S1	Lack of knowledge.
IN3.S2	Mode error.
IN3.S3	Shock.
IN3.S4	Stimulus overload.
IN3.S5	Workload.
IN4.S1	Indicator failure.
IN4.S2	Response slow-down.
IN5	*None defined.*

Code	Specific Instance
IN6	*None defined.*

Table 14.19 Specific antecedents for Planning category

Code	Specific Instance
PL1.S1	Error in goal
PL1.S2	Inadequate training
PL1.S3	Model error
PL1.S4	Overlook precondition
PL1.S5	Overlook side consequent
PL1.S6	Too short planning horizon
PL1.S7	Violation
PL1.S8	Habit expectancy
PL2.S1	Conflicting criteria
PL2.S2	Legitimate higher priority

Table 14.20 Specific antecedents for Temporary Person Related category

Code	Specific Instance
TP1.S1	Daydreaming
TP1.S2	Long time since learning
TP1.S3	Other priority
TP1.S4	Temporary incapacitation
TP2	*None defined*
TP3.S1	Boss
TP3.S2	Comfort call
TP3.S3	Commotion
TP3.S4	Competing task
TP3.S5	Telephone
TP3.S6	Extraneous conversation
TP3.S7	Flight attendant interruption
TP3.S8	Radio call interruption
TP3.S9	External event
TP4.S1	Exhaustion
TP4.S2	Early morning start
TP5.S1	Change of system character
TP5.S2	Illness
TP5.S3	Lack of training
TP5.S4	Over-enthusiasm
TP6.S1	Temporary incapacitation
TP6.S2	Non-pertinent company radio calls

Code	Specific Instance
TP6.S3	PA announcements
TP6.S4	Flight crew sight-seeing
TP7.S1	Boredom
TP8.S1	Boredom

Table 14.21 Specific antecedents for Permanent Person Related category

Code	Specific Instance
PP1	*None defined*
PP2	*None defined*
PP3	*None defined*

Table 14.22 Specific antecedents for Equipment Failure category

Code	Specific Instance
EF1.S1	External event
EF1.S2	Fire
EF1.S3	Flooding
EF1.S4	Impact/projectile
EF1.S5	Power failure
EF1.S6	Tremor
EF2	*None defined*

Table 14.23 Specific antecedents for Procedure category

Code	Specific Instance
PR1	*None defined*
PR2.S1	Habit

Table 14.24 Specific antecedents for Temporary Interface category

Code	Specific Instance
TI1.S1	Design
TI1.S2	Distance
TI1.S3	Localisation problem
TI1.S4	Obstruction
TI1.S5	Temporary incapacitation
TI1.S6	Glare
TI1.S7	Lack of lighting
TI1.S8	Shadow
TI2.S1	Incorrect coding scheme
TI2.S2	Sensor failure

Code	Specific Instance
TI2.S3	Inadequate layout diagram
TI3.S1	Display clutter
TI3.S2	Inadequate display hardware
TI3.S3	Indicator failure
TI3.S4	Display navigation problems

Table 14.25 Specific antecedents for Permanent Interface category

Code	Specific Instance
PI1	*None defined*
PI2.S1	Same call sign, similar flight number

Table 14.26 Specific antecedents for Communication category

Code	Specific Instance
CO1.S1	Noise
CO1.S2	Presentation failure
CO1.S3	Temporary incapacitation
CO1.S4	CRM failure
CO1.S5	Lack of information
CO1.S6	Expectation
CO1.S7	Radio frequency congestion
CO1.S8	Speech rate
CO1.S9	Workload
CO1.S10	Hearback failure
CO1.S11	Habit, expectancy
CO1.S12	Readback failure
CO1.S13	ATIS information being changed
CO2.S1	Hidden information
CO2.S2	Incorrect language
CO2.S3	Noise
CO2.S4	Presentation failure
CO2.S5	Incomplete transcription
CO2.S6	Assumption about receiver's knowledge

Table 14.27 Specific antecedents for Organisation category

Code	Specific Instance
OR1	*None defined*
OR2	*None defined*
OR3	*None defined*
OR4	*None defined*

Code	Specific Instance
OR5	*None defined*
OR6	*None defined*

Table 14.28 Specific antecedents for Training category

Code	Specific Instance
TR1	*None defined*
TR2	*None defined*

Table 14.29 Specific antecedents for Ambient Conditions category

Code	Specific Instance
AC1	*None defined*
AC2	*None defined*
AC3	*None defined*
AC4	*None defined*
AC5	*None defined*
AC6	*None defined*
AC7	*None defined*

Table 14.30 Specific antecedents for Working Conditions category

Code	Specific Instance
WC1.S1	Parallel tasks
WC1.S2	Unexpected tasks
WC1.S3	High traffic volume
WC1.S4	ATC call to taxi for takeoff
WC1.S5	ATC call to meet time restriction
WC1.S6	Maintenance delays
WC1.S7	Minimum time turnaround
WC1.S8	Rushed pre-flight
WC1.S9	Short taxi
WC2	*None defined*
WC3	*None defined*
WC4.S1	Changing schedule
WC4.S2	Shift work
WC4.S3	Time zone change
WC5	*None defined*