WORKSHOP ON THE

Investigation and Reporting of Incidents and Accidents (IRIA 2002)

EDITOR: CHRIS JOHNSON

GIST TECHNICAL REPORT G2002-2,
DEPARTMENT OF COMPUTING SCIENCE, UNIVERSITY OF GLASGOW, SCOTLAND.

TABLE OF CONTENTS

Workshop Timetable

WEDNESDAY 17TH JUNE

09.30-10.00 Chris Johnson	Welcome and Introduction.
10.00-11.00	*Keynote: Normal Accidents-Yesterday and Today* B. Strauch, National Transportation Safety Board, USA.
11.00-11.30	*Coffee*
11.30-13.00 Chair: M. Ayeko, Transportation Safety Board (TSB) of Canada	**Rail Reporting** *A Human Factors Analysis of Highway-Rail Grade Crossing Accidents in Canada* J. Caird, J.I.Creaser,C.J.Edwards,and R.E.Dewar, University of Calgary, Canada. *Validating the Visualisation of Incident Statistics: A Case Study Involving SPADs* F. Spiers and C. Johnson, University of Glasgow, UK. *The Application of Why-Because Graphs to Railway Near Misses* J. Braband and B. Brehmke, Siemens AG Transportation Systems, Germany.
13.00-14.30	*Lunch*
14:30-15:30 Chair: J. Knight, Univ. of Virginia, USA.	**Forensic Engineering (1)** *Using Accident Data to Forecast Societal Cost within the Framework of an Axiomatic Safety-Critical Assessment Process (ASCAP) Simulation* D. E. Brown and J. Stile, University of Virginia, USA. *Learning from Incidents Involving Electrical/Electronic/Programmable Electronic Safety-Related Systems* M. Bowell, Health and Safety Executive, G. Cleland and L. Emmet, Adelard, UK.
15.30-16:00	*Tea*
16:00-17:00 Chair: G. Clelland, Adelard.	**Forensic Engineering (2)** *The Role of Natural Language in Accident Investigation and Reporting Guidelines* K. S. Hanks, J.C. Knight, University of Virginia, C.M. Holloway, NASA Langley.. *Error Classification for Safety Management - Finding the Right Approach* S. Shorrock, Det Norske Veritas (DNV), UK.

THURSDAY 18TH JUNE

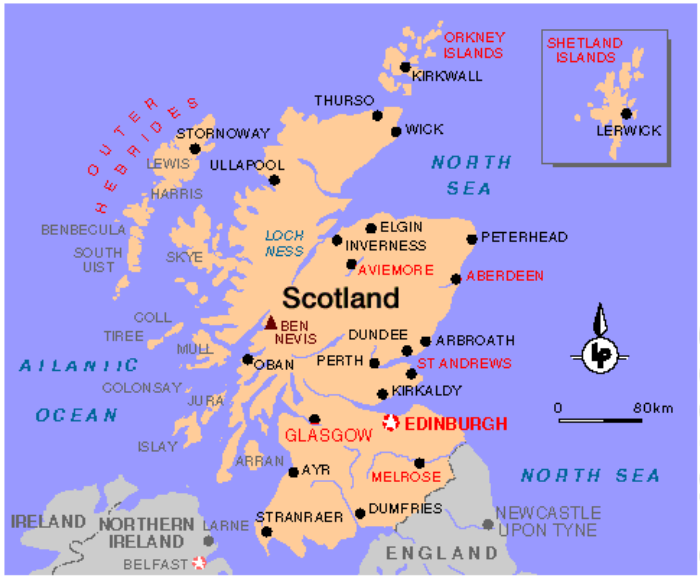
09.00-09.30 C. Johnson	**Poster presentation session**
09.30-10.30	09.30-10.30 *Keynote: Re-cycling the past for the future - issues in the use of incident data* B. Kirwan, EUROCONTROL.
10.45-11.45 Chair: H. Anderson, National Research Labs, Risoe, Denmark.	**Air Traffic Manegement and Safety Levels** *Safety Reporting and Aviation Target Levels of Safety* G. M. Graham and S. Kinnersly, AEA Technology Aviation, UK. A. Joyce, EUROCONTROL Centre Expérimental, France. *'Automatic Safety Monitoring' in Air Traffic Control - Achievements and Perspectives* A. Joyce EUROCONTROL Centre Experimental, C. Fassert, CETCOPRA (Paris-1 Sorbonne), France.
11.45-12.00	*Coffee*
12.00-13.00 Chair: M. O'Leary, British Airways.	**Theory and Practice in Air Traffic Management** *Human Error in European Air Traffic Management: From Theory to Practice* A. Isaac, EUROCONTROL, Belgium. P. Engelen & M. Polman, ATC, NL. *EPOQUES: Proposing Tools and Methods to Treat ATM Safety Occurrences* H. Gaspard-Boulinc, Centre d'Etudes de la Navigation Aérienne, France.
13.30-15.00	*Lunch*
14:30-15:30 Chair: C.M. Holloway, NASA Langley, USA	**Aviation Operations** *Safety Data Collection in British Airways Flight Operations* M. O'Leary, British Airways Safety Services, UK. C. Macrae, N. Pidgeon, UEA, UK *Assessing the Risks of Flight Safety Incident Reports* C. Macrae, N. Pidgeon, UEA, UK. M. O'Leary, British Airways, UK.
15.30-16:00	*Tea*
16:00-17:00 Chair: P. C. Cacciabue, EC, Joint Research Centre, Italy.	**Transportation reporting** *Incident investigation method for cooperative safety management* Y. Murayama, Maritime Labour Research Institute, Japan. Y. Yamazaki, Toyama National College of Maritime Technology, Japan. *Integrated Safety Investigation Methodology (ISIM) – Investigation for Risk Mitigation* M. Ayeko, Transportation Safety Board (TSB) of Canada.

FRIDAY 19TH JUNE

09.00-09.30 C. Johnson.	**Poster presentation session**
09.30-11.00 Chair: T. van der Schaaf, Eindhoven University of Technology	**Medical Systems** *Reporting Adverse Events in Hospitals: A Survey of the Views of 2000 Doctors and Nurses on Reporting Practices and Models of Reporting* H.B. Andersen, M. Dyrløv Madsen, Risø National Laboratory, N. Hermann, T. Schiøler, DSI Danish Institute for Health Services Research, D. Østergaard, Herlev University Hospital, Dept Anaesthesiology, Herlev, Denmark *Development of a Region Wide Process for the Investigation of In-Hospital Deaths* J. M. Davies, University of Calgary, B. Young, Calgary Health Region, Canada. *A Survey of Safety Culture in Hospitals Including Staff Attitudes on Incident Reporting* K. Itoh, Tokyo Institute of Technology, T. Abe, Tokyo Medical and Dental University, Japan. H. B. Andersen, Risø National Laboratory, Denmark.
11.00-11.30	*Coffee*
11.30-13.00 Chair: Shelley Jeffcott, University of Glasgow.	**Reporting Cultures and Organisational Issues** *Learning by Reporting System of Organizational Accidents in Japan* K. Tanaka, University of Electro-Communications Chofu, Japan. *A Study of Incident reporting in Air Traffic Control - Moral Dilemmas and the Prospects of a Reporting Culture Based on Professional Ethics* M. Dyrløv Madsen, Risø National Laboratory, Denmark. *Forensic Software Engineering and Stories of Failure* D. Dalcher, Middlesex University, UK.
13.00-14.30	*Lunch*
14:30-16:00 Chair: Chris Johnson, University of Glasgow.	**Analytical Tools** *Not Reporting Successful Recoveries from Self-Made Errors? An Empirical Study in the Chemical Process Industry* T. van der Schaaf and L. Kanse, Eindhoven University of Technology, Netherlands *A Framework for Re-examining Accident Reports to Support Interaction Design Processes* A. Bruseberg, I. Solodilova, R. Hourizi and P. Johnson, University of Bath, UK. *Interactive Evidence: New Ways To Present Accident Investigation Information* D. Schofield, J. Noond, L. Goodwin and J. March, University of Nottingham, UK.1
16:00-16.30	*Tea*
16:30-17:00	*Keynote: Reporting incidents: An end unto itself or a call for action? Next steps.* Sue Bogner, Institute for the Study of Medical Error, USA
17:00-17.15	*Close and hand-over.*

.

SATURDAY, 20TH JUNE



Scotland

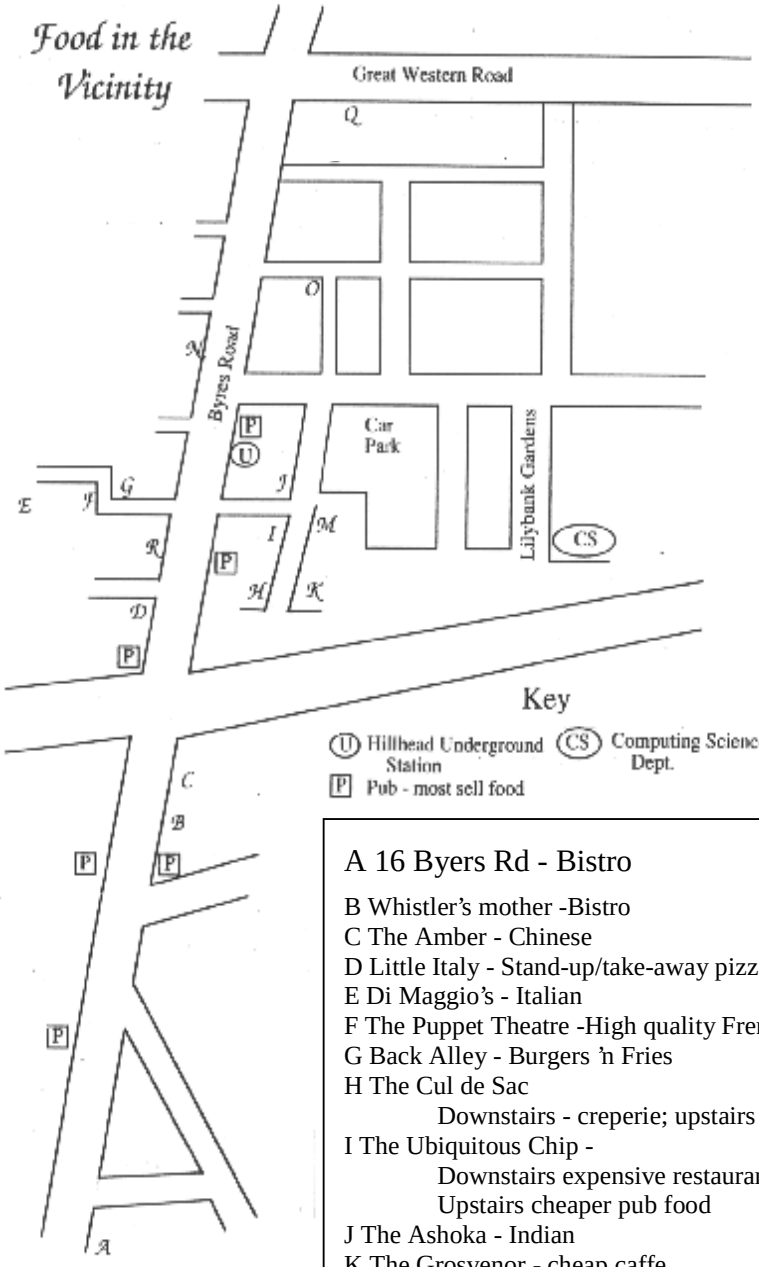
This will provide the opportunity for informal discussions about the issues raised during the workshop. The day will be spent on the Isle of Arran, off the west Coast of Scotland. The intention is to meet outside the Department of Computer Science at 07:30. We will be taking the train because this connects directly with the CalMac (http://www.calmac.co.uk) ferry onto the Island. Anyone who misses the first rendez-vous can meet us underneath the large clock at Central Station for 08:00 (Buchanan Street is the nearest Underground station). Trains depart from Glasgow Central station at 08:33, arrives at Ardrossan harbour at 09:25. The ferry leaves for Arran at 09:45. Ferry arrives at Brodick on Arran at 10:40. The ferry departs Brodick at 16:40, arrives Ardrossan 17:35. The train arrives at Glasgow Central 18:52. There is an additional service departing Brodick at 19:20, arriving at Ardrossan to connect with the 20:30 that arrives into Glasgow at 21:22.

If anyone misses this departure then they will have to spend the night on the Island (there are lots of hotels and bed & breakfast places). Arran Tourist Office can be contacted on 01770-302140 or 01292 678100 (http://www.ayrshire-arran.com/arran.htm) for hotel accommodation and other enquiries. The whiskey distillery is open for visits from 10.00-18.00 and can be contacted on 01292 678100.

Out	Monday to Saturday					Sunday			
Glasgow Central dep	**0833**	1115	1415	1650	1915	0840	1115	1405	1655
Ardrossan dep	**0945**	1230	1515	1800	2030	0945	1230	1515	1800
Brodick arr.	**1040**	1325	1610	1855	2125	1040	1325	1610	1855

Return	Monday to Saturday						Sunday			
Brodick dep	0820	1105	1350	**1640**	1920	2140	1105	1350	1640	1920
Ardrossan arr	0915	1200	1445	**1735**	2015	2235	1200	1445	1735	2015
Glasgow Central arr	1022	1322	1622	**1852**	2122	-	1328	1550	1850	2117

Restaurants in the Local Area



A 16 Byers Rd - Bistro

B Whistler's mother -Bistro
C The Amber - Chinese
D Little Italy - Stand-up/take-away pizzas
E Di Maggio's - Italian
F The Puppet Theatre -High quality French
G Back Alley - Burgers 'n Fries
H The Cul de Sac
Downstairs - creperie; upstairs - pub
I The Ubiquitous Chip -
Downstairs expensive restaurant
Upstairs cheaper pub food
J The Ashoka - Indian
K The Grosvenor - cheap caffe
M Jinty McGinty's - Irish pub & bar food
O The Metro - salads, hot dishes
Q The Parthenon - Greek food
R Burger King

Normal Accidents-Yesterday and Today

Barry Strauch

National Transportation Safety Board, Washington, DC 20594, USA[1]

Abstract: Charles Perrow's *Normal Accidents*, published in 1984, provided a unique perspective on system safety by focusing on system shortcomings and their effects on safety, rather than the operators primarily. Perrow contended that the potential high risks of many complex systems led designers to incorporate defenses that, when combined with tight coupling of system processes, exacerbated rather than enhanced safety. Several accidents are described to illustrate Perrow's points in contemporary high-risk systems. Since its publication and more recently that of a second edition, his work has influenced both students of error and accident investigators in ways that have enhanced system safety. However, with the benefit of hindsight it can be seen that his contention that accidents in high-risk systems would normally follow system shortcomings and would become normal events has not been consistently realized. Some advances in technology have exacerbated safety but others have enhanced it. Organizations that have emphasized cost reduction over safety have created system shortcomings, but others have taken steps to improve safety.

Introduction

In 1984 the first edition of Charles Perrow's seminal text, *Normal Accidents* (1999), was published. Perrow argued that accidents in high-risk systems, such as nuclear power plants, chemical refineries, and shipping, would become the normal outcomes of the systems themselves and would increase in frequency. He did, however, add that we could reduce their potential danger and hence, reduce their incidence by better understanding and addressing dangers inherent in system designs.

High-risk systems had become increasingly complex, he argued, largely because of the defenses designers added to reduce the likelihood of accidents. But these defenses have actually exacerbated safety because they increased the systems' interactive complexity. The interactive complexity, combined with the tight coupling of system processes, that is, the sequential processes that occur in strict order and within specific time frames in response to events or actions, obscures the outcomes of operator actions from the operators themselves. As he writes,

If interactive complexity and tight coupling–system characteristics–inevitably will produce an accident, I believe we are justified in calling it a normal accident, or a system accident. The odd term normal accident is meant to signal that, given the system characteristics, multiple and unexpected interactions of failures are inevitable. (p. 5)

In 1999 the second edition of Perrow's book was published. There he described several prominent accidents that had occurred since the first edition's 1984 publication, such as the 1986 accident involving the Space Shuttle Challenger, and the Chernobyl nuclear facility of the same year. These accidents, he believed, supported his earlier assertion of the inevitability of normal accidents, the consequence of the increasing complexity of high-risk systems in transportation, electrical power generation, and space exploration systems, to name a few.

Perrow's work was considerably influential and helped to change perspective of accidents in complex systems. Investigators and researchers had typically focused on the operators who committed the critical errors. When their errors led to accidents, action would be taken against the operators, rather than action taken to address the system features that had led to the errors or enabled them to occur. Many believed that operators who had been thoroughly trained to perform error free deserved to be punished, not only to bring a sense of justice to the commission of the errors, but to send a message to others as well, i.e., that errors in these systems are unacceptable. Others believed that by focusing on operator deficiencies primarily safety would be enhanced. *Normal Accidents* helped to change this. In many, but not all, parts of the industrialized world today errors are viewed as the "normal" result of shortcomings in equipment design, procedures, training, oversight, or other system feature, rather than the result of operator negligence. By addressing those shortcomings rather than punishing the operators, future operators would be less likely to encounter similar situations, and if they do encounter these circumstances despite efforts to the contrary, they would be better prepared to respond. Today many, if not most students of error have accepted the

[1] The opinions expressed herein represent those of the author and not necessarily those of the National Transportation Safety Board.

Perrow view of the contribution of system elements to accidents. This transformation of error perspective has, I believe, led to profound changes in the way we investigate, consider, and respond to accidents.

With the benefit of almost two decades of hindsight, the accuracy of Perrow's predictions can begin to be gauged. While catastrophic accidents have continued to occur, it can be seen that their frequency has decreased, not increased, and accidents are still infrequent events. In some systems, such commercial aviation, nuclear power generation, chemical refineries, and others, system safety has increased. There have been no major accidents at nuclear generating facilities since Chernobyl, and while there have been major transportation accidents, some accident types that had once occurred with some regularity now rarely take place. Although almost two decades may be insufficient to accurately gauge the accuracy of Perrow's implication, the effects of certain events on system safety can be assessed.

Nonetheless, as Perrow noted, accident types that were experienced before 1984 have continued to take place. Worse, accident types not seen before 1984 have occurred since then, and system shortcomings identified before 1984 that were not addressed have led to accidents. In hindsight, several of the accidents that have taken place support Perrow's assertion; they were the "normal" outcomes of system deficiencies. In this respect Perrow's contention has been supported; accidents have been the normal outcomes of system deficiencies.

Normal outcomes of system deficiencies since 1984

Accidents that have occurred since 1984 demonstrate how system elements can hide the results of operator actions from the operators themselves, the result of a combination of interactive complexity and tight coupling. The 1995 accident involving an American Airlines Boeing 757 in Cali, Colombia (Aeronautica Civil, 1996), for example, a regularly scheduled flight from Miami, Florida, which resulted in the death of 151 passengers and crew, supports Perrow's assertion.

The sequence of events leading to the accident began during the flight's descent into the Andean city, after the Cali air traffic controller offered the pilots the opportunity to land to the south on its one runway. Almost all of American's flights from Miami landed to the north in Cali, requiring the flights to fly over Cali, turn around, and land opposite to the direction of the initial descent path. Those landing to the south maintained the same southerly heading into the Cali area on to the runway. Pilots prefer straight in approaches because they save time and therefore, expense.

After accepting the offer to land to the south, the pilots had to quickly perform a number of additional tasks to prepare for that approach. They had to review the requirements of the approach to the south and reset equipment to its navigation, speed, descent, and communication parameters. One requirement called for pilots to inform the controller when their flights had crossed over a navigation beacon called Tulua, a routine request for information that controllers needed to locate airplanes in airspace that was not equipped with air traffic control radar. The beacon, Tulua, served as the "entry" to, or the "initial approach fix" of the approach. However, unlike the format routinely used to designate approaches, this approach was not named after Tulua, the initial approach fix, but after another beacon called Rozo, located just before the runway and which served as the "final approach fix."

The captain began the accident sequence after he misinterpreted the controller's clearance to execute the approach to the southern runway. He incorrectly commanded the airplane's flight management system, computers that, at the pilot's direction, navigate and control the airplane's flight path, to fly directly to Cali rather than to the initial approach fix, as the approach had required. This, the first of a series of errors that he and the first officer committed, caused the flight management computers to delete the Tulua beacon from the data needed for the approach. Without locating Tulua the pilots could not perform the approach to Cali as required. Neither pilot could identify the reason for their inability to locate Tulua; each repeatedly and unsuccessfully attempted to locate it through the flight management system, and each became increasingly frustrated as the flight continued to descend to Cali. As they got closer they resorted to a non-standard and unapproved means of executing the approach; they entered into the system what they believed to be abbreviator for the Rozo beacon, and commanded the flight management system to fly to it, in the belief that it was the initial approach fix. However, they did not realize that 1) Rozo was the final and not the initial approach fix, and 2) the standard abbreviation format, and the abbreviator used on the navigation chart used to identify it did not correspond to the electronic designation for Rozo. Instead, the pilots had unknowingly commanded the flight to fly away from rather than towards Cali. Thirty seconds later, after they realized their error, they returned the airplane to the initial flight path. Shortly thereafter, the airplane struck a mountain.

As Perrow had previously described, several critical system elements were hidden from the pilots who regularly relied on these elements to conduct approaches. These included features of the airplane's flight management system navigation database design, and inconsistencies in both the method used to designate instrument approach procedures and in the format used of abbreviations of navigation beacons. Being unaware of these elements had not affected their performance before, but on the night of the accident the combination of the three, with other factors,

led the pilots to command the flight management system to perform certain actions without understanding their consequences. Here, as in other accidents, the interactive complexity and tight coupling of the system, in its flight management logic, the navigation data base design, the designation of the approach, and the abbreviation of navigation beacons, obscured the consequences of their actions from the pilots themselves.
Perrow had also pointed out that designers and operators do not consistently apply the lessons of previous events to correct the deficiencies that those events had highlighted. An accident remarkably similar to the Cali accident had occurred in 1992, over three years before this accident, with pilots of a relatively comparable aircraft, an Airbus A-310, had crashed while on a descent into Katmandu, Nepal, but the lessons of that accident were not fully addressed. As in Cali, those pilots had also departed their assigned flight path and veered off course in a mountainous region of the Himalayas, after an unexpected event required that they reinitiate their approach. In both accidents the pilots had to perform certain steps in sequence, but only after they had reached specific locations, at particular altitudes, and at set airspeeds. Performing one action incorrectly, or becoming distracted, led to a serious of outcomes that reduced the options available to the pilots to safely operate the airplanes, creating an accident that was a "normal event" given the system deficiencies. In a less interactively complex or less tightly coupled system, the accidents would probably not have occurred.

Accident frequency since 1984
The incidence of major accidents since 1984 provides only partial support to Perrow's assertion that the number of normal accidents would increase. In aviation in the United States the frequency of major accidents has been about the same for over two decades, despite the fact that, at least until September 11, the number of commercial aircraft being operated has steadily increased. Since 1984 the United States has witnessed two major marine accidents, each very different from the other, the 1989 grounding of the tanker Exxon Valdez in Alaska, and the 2001 collision of the US nuclear submarine Greenville with a Japanese fishing vessel off the coast of Hawaii. On the other hand, the United Kingdom has sustained nine fatal rail accidents since 1994, most recently, this past May at Potters Bar.
Therefore, since the publication of *Normal Accidents*, the incidence and rate of system accidents can be seen to have stayed the same, increased, or decreased, depending on the country and the particular complex system. Consequently, Perrow's assertion may have been accurate, but its accuracy is affected by industry variables, cultural variables, or an interaction of the two, effects that he did not explicitly address, but which his work has helped others to outline (e.g., Helmreich and Merritt, 1999).
Perrow highlighted the adverse effects of advanced technology on system safety and the Cali accident demonstrated some of these effects. But since 1984 technology can also be seen to have helped enhance system safety, in ways that he did not predict. Further, with the benefit of hindsight his work itself has helped to enhance safety by leading to better tools to identify and respond to safety deficiencies. Together, these have served to limit the accuracy of his prediction.

Technology: As Perrow predicted and as seen in the Cali accident, since 1984 more advanced and more complex equipment, with more interactive complexity and tighter coupling, have been introduced into service and this has led to an increase in system accidents. For example, after Airbus Industrie introduced an aircraft, the A-320, which was considerably more advanced than previously designed aircraft, it experienced a relatively high rate of fatal accidents. However, after several years the frequency of accidents involving the A-320 and its derivatives has leveled off to that consistent with other aircraft.
Amalberti (1999) attributed the rise in accidents that often follows the introduction of new technology to a period of learning that operators, regulators, and managers need to adjust to and master new system features, features that likely include the interactive complexity that Perrow had identified. Because users typically have gained considerable experience with "older" technology by the time they change to the new, the period of adjustment may be especially long because the changes required to adjust to the new may be more encompassing than had been previously considered. As he writes:

> People do not yet know how to optimize complex systems and reach the maximum safety level without field experience. The major reason for this long adaptive process is the need for harmonization between the new design on one hand and the policies, procedures, and moreover the mentalities of the whole aviation system on the other hand. This harmonization goes far beyond the first months or years following the introduction of the new system, both because of the superimposition of old and modern technologies during several years, and because of the natural reluctance of people and systems to change. (p. 173)

Notwithstanding these adverse effects, advanced technology has also improved safety in high-risk systems. It has allowed operators to train in devices that replicate system operations to a remarkable degree. Realistic scenarios

enable them to learn and practice responses in a constructive environment, and respond to scenarios that would otherwise be far too dangerous to practice in the actual systems. The devices can also capture parameters of trainee performance to enable instructors to better focus on performance needs. These have also given managers the flexibility to schedule training sessions independently of system operating phases or cycles, to maximize opportunities for student learning. Systems that have employed these devices have seen improved operator skills and enhanced responses to unexpected and potentially catastrophic situations. They also provide an additional safety benefit; training-related accidents have been all but eliminated when these devices have been used as the primary training platforms.

In aviation, the domain with which I am most familiar, advances in technology have reduced the incidence of several types of accidents. For example, in the late 1970s, in response to a series of what are referred to as CFIT accidents (controlled flight into terrain), in which pilots inadvertently fly controllable aircraft into terrain after getting off course or prematurely descending, the United States government mandated the installation of devices, known as ground proximity warning systems or GPWS, to warn pilots of impending collisions with terrain. Since then, advances in technology have corrected a deficiency of early GPWS, a result of its use of radio wave propagation from the aircraft to the ground. As the Cali, Colombia, accident demonstrated, at high speeds GPWS provides little warning of impending collisions with rapidly rising terrain. But since the late 1970s, because of inexpensive data storage capabilities, enhanced software, and advanced navigation techniques, vertical terrain could be reliably predicted well in advance of a collision. With reliable and accurate navigation, devices were developed that "know" at all times the location of an airplane relative to terrain, enabling pilots to be alerted to terrain far sooner than they could with GPWS. The Cali accident helped to persuade the industry of the need for this advanced and more sophisticated GPWS, called terrain aural warning systems or TAWS, and many airlines have begun to install TAWS on their aircraft. Although it is still relatively early in the application of the technology, no TAWS equipped aircraft has been involved in a CFIT accident. This device has, and one can predict with confidence, will continue to reduce the incidence of CFIT accidents.

Technology has also addressed another type of aviation accident, the collision of two aircraft in flight. Again, as a result of a major accident, the in flight collision of a commercial transport with a smaller aircraft, the United States government required passenger carrying transport aircraft that fly within or to the United States to be equipped with devices, called TCAS for terminal collision alerting systems, that warn pilots of the presence of potentially conflicting aircraft. Even with airspace that has almost universal radar coverage, collisions involving transport aircraft continued to occur in the United States. However, since TCAS began to be installed, no two TCAS aircraft equipped have collided. By contrast, on November 12, 1996, the fifth worst aviation accident in history occurred when a Saudia Boeing 747 and a Kazakstahn Illyushin 76 collided near New Delhi, India, killing 349 persons. Neither aircraft was equipped with a TCAS. Ironically, Saudia had installed TCAS on its aircraft that operated into the United States, but not on its aircraft, such as this one, that did not.

Investigations and system safety Advances in technology have also enhanced the quality of accident investigations, allowing investigators to better understand the causes of accidents and better identify the issues related to them. System recorders, such as flight data recorders, give investigators access data that was unavailable just a few years earlier. These advances have been accompanied by new and more sophisticated software that can provide a comprehensive view of system processes and actions. System recorders have also been installed in a wide variety of systems such as railroad locomotives, marine vessels, and highway vehicles, including automobiles, giving insights into system activities that are similar to those available to aviation accident investigators.

Further, whereas accident investigators in systems as diverse as commercial aviation and college bonfires had tended to stop their search for accident causes after identifying the errors and the operators who committed them, many investigations today go far beyond that (e.g., Special Commission, 2000; Air Accidents Investigation Branch, 1992). This has improved the ability of regulators, managers, and designers to address and correct shortcomings that had led to the errors. As investigators have highlighted deficiencies in training, procedures, and oversight, to name a few, organizations and regulators have addressed the shortcomings, thereby improving safety. In general, organizations that have adopted this view of error, such as British Airways after the 1990 BAC 1-11 accident, have experienced few accidents. By contrast, countries with cultures that emphasize the punishment of error have seen little reduction in system accidents.

Organizational accidents: Perrow helped to shape the current recognition of the role of organizations in system safety. His work helped Reason and others to more fully address the affect of organizations on the safety of their operations. As Perrow wrote,

> …We have seen that the potential for a system accident can increase in a poorly-run organization. If there is poor regulation, poor quality control, or poor training, there is an increased chance of failures in the DEPOSE (System) components, and these can make the unexpected interaction of failures more likely, because there are more failures to interact. (p. 343)

Unfortunately, since the first edition of *Normal Accidents*, the history of system accidents supports his contention. Since 1984, highly visible accidents that were affected by organizational features, that demonstrate the importance of organizations in maintaining system safety, have occurred.

Organizational factors were evident in the 1996 crash of a ValuJet DC-9 that killed all 110 persons onboard the regularly scheduled revenue passenger flight (NTSB, 1997). A fire that had broken out on the airplane caused the accident. It was initiated by canisters of chemical oxygen generators that had been mistakenly loaded into the aircraft's cargo bin. Ten minutes after takeoff the fire penetrated the fire resistant cargo liners and entered the cabin, causing the crash into the Everglades.

After investigators learned of the nature of the cargo, retrieved the flight recorders, and confirmed the presence of a fire, they focused on determining how the canisters were loaded onto the airplane. Passenger-carrying jet aircraft are required to provide supplemental oxygen to passengers and crew in the event of decompression. Shipping canisters of chemical oxygen generators, used on many aircraft to supply the oxygen, without special housing or locking devices, is prohibited because the oxygen generation process heats the canisters to the point that they can ignite adjacent material. In the event of a fire, the generators would feed pure oxygen to the fire, substantially increasing its intensity.

ValuJet was a relatively new airline, having begun service three years earlier. In that time it grew from two to 52 airplanes. Its success had been due, in part, to its ability to keep its costs low. It performed few of the operational tasks that "traditional" airlines had performed, such as pilot training, instead contracting those out to other vendors. In contracted out its maintenance to multiple vendors, according to its needs, including a maintenance facility in Miami, SabreTech, retained to bring two aircraft it had recently acquired to flight operations standards. The canisters that were on the accident aircraft had been removed from these two aircraft. SabreTech, as ValuJet, had also contracted critical functions. For example, the services of many of the maintenance technicians it used had been obtained from contractors that provided licensed technicians as needed. Over half the maintenance technicians that had removed the canisters of oxygen generators from the two aircraft were contract and not SabreTech employees.

The canisters were loaded onto the airplane following a series of relatively minor errors that the maintenance technicians, and others in the maintenance facility, had committed about two months before the accident. SabreTech was then facing costly expenses because its work on the two aircraft was running behind schedule, and its contract called for substantial penalties for delays in delivering the aircraft to the airline. SabreTech management had placed considerable pressure on facility personnel to complete the work on the two aircraft, requiring all technicians to work 12-hour days, seven days a week, until the work was completed.

The canisters had been removed because they were approaching or had passed their expiration dates. Neither SabreTech nor ValuJet had anticipated the need for parts necessary to disable the canisters. Therefore, when maintenance personnel removed them, the only other approved method of disposing of each of the canisters was to expend or initiate the generation process, a time consuming and potentially hazardous procedure given the intense heat this caused. This too was not done. Further, because SabreTech had run out of the necessary color coded labels, personnel placed incorrectly coded labels, which signified components that retained value, on the removed canisters. They then placed them loosely in boxes and set the boxes aside. Days before the accident, after the initial maintenance work had been completed, facility warehousing personnel saw the boxes, and without recognizing the nature of, or the potential danger associated with their contents, decided to return them to their "owner," the airline, incorrectly ascribing value to them because of the packing labels.

Although ValuJet had contracted heavy maintenance to other facilities for financial reasons, it was still legally responsible for ensuring that maintenance performed on its behalf was carried out in accordance with its procedures. It was also contractually responsible for ensuring that SabreTech had the parts needed to perform the maintenance. SabreTech, which had contracted out costly functions of its own, was legally responsible for ensuring that the maintenance it performed was carried out in accordance with ValuJet's procedures, and contractually responsible for informing ValuJet of the parts that it would need to perform the maintenance. The Federal Aviation Administration (FAA) was responsible for ensuring that both ValuJet and SabreTech followed the maintenance procedures and requirements applicable to each.

However, the FAA was hampered because it was unprepared to oversee the type of maintenance that ValuJet employed. Its oversight had been developed in response to maintenance that "traditional" airlines, not newer airlines with newer ways of performing maintenance, performed. Traditional airlines had designed maintenance programs and carried out maintenance at their own facilities. These were generally located near their operational centers. The

FAA in turn located its inspectors near the airline's maintenance centers to enhance their ability to regularly visit the facilities, interact with personnel, and observe maintenance practices. This tended to foster a ready FAA inspector familiarity with the airline's maintenance programs and the maintenance they performed.
However, ValuJet was a "non-traditional" airline. It had developed its own maintenance program as all airlines did, but then contracted all major maintenance activities. Experience that maintenance personnel gained interpreting and applying provisions of the maintenance program did not apply to the "non-traditional" models, and experience that parts departments gained anticipating the parts that maintenance departments would need did not apply. Although ValuJet provided personnel to oversee the maintenance performed at SabreTech, two of the three employees it assigned to the Miami facility were not its employees.
In such cases, with both airline and contractor maintenance shortcomings, it is incumbent upon the regulator to properly oversee maintenance operations to ensure a minimum level of safety. This was not done. The FAA stationed inspectors near the airline's operational center in Atlanta, but the maintenance critical to this accident was carried out in Miami. Atlanta-based inspectors did not visit the Miami facility before the accident. Miami-based FAA maintenance inspectors did visit the facility, but they were unfamiliar with the airline's maintenance program and therefore, could not properly oversee the maintenance performed.
The ValuJet accident supported Perrow's contention regarding the role of organizations in creating the conditions for normal accidents, and demonstrated that such accidents have continued to occur. By emphasizing cost reductions over procedural safety, both ValuJet and SabreTech contributed to the cause of the accident. However, the accident could still have been prevented had the regulator performed its mission, one made even more critical when the major organizations involved manifest shortcomings, a mission developed largely to prevent organizations from circumventing safety practices.
Since this accident the lesson of the important role of both organizations and regulators in maintaining system safety has been further demonstrated in domains removed from those that Perrow discussed. For example the Enron Corporation collapse was a financial catastrophe that showed that organizational safety breakdowns extend to financial as well as to high-risk systems. The practices of a multi-national, publicly traded company that sought to circumvent "safe" financial procedures failed to obtain proper oversight by those responsible for carrying it out, outside auditors, financial analysts, and securities regulators. Subsequent investigations showed that the independence of the auditors and analysts was compromised by their own financial self-interests. Although no lives were lost as a result of its collapse, except for the suicide of its former vice-chairman, when correct information about the nature of Enron's finances emerged a financial accident ensued. Thousands of employee's lost their jobs and their retirement funds, investors lost billions of dollars of equity, and the company was forced into bankruptcy.

Applying the lessons of system accidents

Although accidents are often the mechanisms through which system deficiencies become manifest, the very manifestations of these deficiencies often leads to their being addressed, and therefore to a reduction in the risk of that particular type of normal accident. For example, in response to the findings of the Cali accident, members of the aviation industry have attempted to address shortcomings in the design of navigation data stored in electronic navigation databases. Similarly, after the ValuJet accident the FAA has changed the nature of its oversight of airline operations. In the United States, financial regulators, analysts, and accountants have begun to address the financial system deficiencies that led to the Enron collapse. Already, a major stock brokerage firm has altered the pay structure of its analysts to reduce their incentives to deliberately disseminate misleading analytical forecasts.
Similar enhancements to safety have been implemented in aviation after a series of accidents in which pilots inadvertently flew their aircraft close to thunderstorms, encountering among the most severe weather possible. Airlines, airframe manufacturers, and the FAA developed a training aid to assist pilots to recognize the potential presence of microbursts and avoid them, and to assist those who inadvertently encounter them to escape them. Since the training aid was disseminated to United States pilots, there has been only one accident in which a United States operated passenger jet crashed because of an encounter with a microburst.
Likewise, in the early 1990s the airlines, airframe manufacturers, and the FAA jointly developed a training aid to reduce a type of accident, known as a rejected take off, in which pilots attempt to stop their airplanes at speeds that are too high for the remaining runway distances. The program was modeled after the microburst avoidance one. Before it was implemented, rejected takeoff accidents occurred in the United States with some regularity. Since then no rejected takeoff accident has occurred in the United States.
Unfortunately, system deficiencies are not consistently addressed and deficiencies may continue to be present after being identified. For example, the Exxon Valdez accident was particularly embarrassing to the Exxon Corporation because the vessel's master, after being diagnosed as an alcoholic and sent by the company to an alcohol-rehabilitation program, was found to have been drinking on the vessel and was determined to have been under the

influence of alcohol at the time of the accident (NTSB, 1990). Since the accident, there have been groundings in United States territorial waters, but none have caused the severe consequences of the Exxon Valdez.

Looking back on the years since *Normal Accidents* was published, however, it can be said with some confidence that systems that have not responded to the deficiencies raised in normal accidents have been doomed to repeat them. For example, in 1987, a Northwest Airlines MD-80 crashed in Detroit shortly after takeoff, killing over 150 passengers and crew (NTSB, 1987) because the airplane had not been properly configured for takeoff; the crew had failed to extend the airplane's flaps and slats, devices that are necessary for takeoff. This single error led to the accident, but it had been preceded by an additional failure, an airplane component called the central aural warning system that had been installed to alert pilots to just such an event, a takeoff attempt with retracted flaps and slats, had failed to sound. The crew received no warning when they attempted to takeoff, despite the improperly configured flaps and slats.

The accident led to calls for the inspection of aural warning components and to changes in pilot training so that pilots would be less likely to commit this error. Yet, almost exactly one year later an aircraft operated by different airline crashed in a manner almost identical to the earlier accident (NTSB, 1989). The crew had failed to extend the flaps and slats for takeoff from Dallas-Ft. Worth airport and the airplane crashed shortly thereafter, killing 14 passengers and crew. Again the crew's error and the failure of the takeoff warning system to alert the crew to the unsafe airplane condition before they attempted to takeoff were cited as having caused the accident.

This accident, coming as it did so close to the previous nearly identical accident the year before, led to significant changes in the airline. It modified its procedures, training, and oversight to ensure that the accident was not repeated. In the United States, these efforts have been successful; since the 1988 accident no airplane has crashed in the United States because the flaps and slats were not extended for takeoff. By contrast, an Argentine aircraft crashed in 1999 just after takeoff from Buenos Aries after the crew had failed to extend the flaps and slats. All onboard the airplane were killed.

In July 2000 an Air France operated Concorde crashed just after takeoff from Paris. All 109 passengers and crew onboard were killed, along with four people on the ground. The accident was attributed to tire failure on takeoff after the Concorde's wheels had struck metal debris from the aircraft that had taken off just before it (BEA, 2000). The debris damaged the tire, and tire debris then punctured the fuel tanks and were ingested into the engines. Two of the four engines failed, and leaking fuel from the tanks fed fires that had erupted in the failed engines. The pilots were unable to control the airplane.

The Concorde employed a unique design. Unlike other passenger carrying aircraft, the landing gear and tires were placed forward of the engines. Tire failure posed a danger because tire debris could be ingested into the engine and harm them, and damage the fuel tanks located aft of the wheels. This possibility had been demonstrated to the aircraft's designers and operators. Investigators of the 2000 accident cited six earlier incidents, from 1979 through 1993, in which tire bursts or wheel failures had occurred. In one, pieces of the wheel rim damaged critical aircraft systems and punctured the fuel tanks. Yet, until after the 2000 accident the Concorde's fuel tanks were not hardened to the point that they could safely withstand tire/wheel debris, and tires were not reinforced to reduce the likelihood of tire burst. Although the lessons of the 2000 accident had been known for years, they were not applied until the accident in Paris.

Perrow cites a number of possible reasons for the absence of effective responses to system related shortcomings. The rarity of accidents can make managers unwilling to take on the costs and potential risks of addressing safety deficiencies, and many come to believe their own rhetoric about the importance of safety in their operations. In addition, they may learn from accidents but learn the "wrong" things and hence may not address the real system vulnerabilities.

The lack of effective response to demonstrated deficiencies remains a significant threat to system safety and an accident can be a rather difficult way to learn of them. Unfortunately, the failure to respond to demonstrated vulnerabilities has not been found exclusively with regard to system safety, but to system security as well. For example, despite experiences with foreign-based terrorists seeking to destroy the World Trade Center, such as the 1994 bombing of the facility, and despite information on terrorists seeking to destroy civilian air transport aircraft, the United States failed to effectively recognize and respond to this threat until after the horrific events of September 11, 2001. Hopefully, we have begun to address the security and intelligence deficiencies that failed in response to those particular terrorist actions, and that responses to both safety and security shortcomings will effectively address them.

Conclusions

In suggesting that features designed to enhance system safety can actually create deficiencies, Perrow introduced a concept that, at the time, was truly radical and ultimately, influential. He forced investigators and researchers to consider error not as an outcome of the person committing the error, but as a function of design features themselves. As he wrote, "our key term system accident or normal accident … focuses on the properties of systems themselves, rather than on the errors that owners, designers, and operators make in running them" (pp. 62, 63). His work helped that of Neville Moray (1994, 2000), Jens Rasmussen and his colleagues (e.g., Rasmussen et al., 1994), James Reason (1990, 1997), and David Woods (Woods et al., 1994), among others, to further explore human error in the context of high-risk systems, enhancing our understanding of the nature of system accidents.

With the hindsight of almost two decades, Perrow's largely pessimistic view of system safety has not been borne out. To some extent this has been the result of public pressure brought to complex systems to raise safety standards. The accidents at Three Mile Island, Cali, and the Everglades, as well as the Exxon Valdez accident, brought intense public pressure to improve the safety of the nuclear power industry, commercial aviation, and the marine industry respectively. This has led to changes in the nature of government oversight of the industry and in the practices of the regulated industries.

Perrow gave us a more nuanced and sophisticated view of accident causation than the one extant at the time, a view that permitted the investigation and analysis of system incidents to lead to real improvement in the ability of government and industries to enhance and maintain system safety. With advances in training technology and in the technology of investigations, government regulators and the regulated industries themselves could better identify system shortcomings that lead to accidents better than they before. Today it can be seen that Perrow's prediction about a steady increase in system accidents may not have been realized, but the safety of complex systems has been improved nevertheless.

References

Aeronautica Civil of the Government of Colombia. (1996). *Aircraft accident report, controlled flight into terrain, American Airlines flight 965, Boeing 757-223, N651AA, near Cali, Colombia, December 20, 1995.* Bogotá, Colombia.

Air Accidents Investigation Branch, (1992). *Report on the accident to BAC One-Eleven, G-BJRT over Didcot, Oxfordshire, on 10 June, 1990.* Aircraft Accident Report No. 1/92 (EW/C1165). London. Department of Transport.

Amalberti, R. R. (1998). Automation in aviation: A human factors perspective. In D. J. Garland, J. A. Wise, and V. D. Hopkin, (Eds.), *Handbook of aviation human factors* (pp. 173-192). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Bureau Enquetes Accidentes, (2002), *Accident on 25 July, 2000, at La Patte d'Oie in Gonesse, to the Concorde, registered F-BTSC, Operated by Air France,* Paris.

Helmreich, R. L. and Merritt, A. C. (1998). *Culture at work in aviation and medicine: National, organizational and professional influences.* Aldershot, England: Ashgate.

Moray, N. (1994). Error reduction as a systems problem. In M. S. Bogner, (Ed.), *Human error in medicine* (pp. 67-91). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Moray, N. (2000). Culture, politics and ergonomics. *Ergonomics, 43,* 858-868.

National Transportation Safety Board. (1988b). *Aircraft Accident Report, Northwest Airlines, Inc., McDonnell Douglas DC-9-82, N312RC, Detroit Metropolitan Wayne County Airport, Romulus, Michigan, August 16, 1987.* (Report Number: AAR-88-05). Washington, DC.

National Transportation Safety Board. (1990b). *Marine Accident Report, Grounding of U.S. tankship Exxon Valdez on Bligh Reef, Prince William Sound, Near Valdez, Alaska, March 24, 1989.* (Report Number: MAR-90-04). Washington, DC.

National Transportation Safety Board. (1997). *Aircraft Accident Report In-Flight Fire and Impact With Terrain Valujet Airlines Flight 592 DC-9-32, N904VJ Everglades, Near Miami, Florida May 11, 1996* (NTSB Report Number: AAR-97-06). Washington, DC.

Perrow, C. (1999). *Normal accidents: Living with high-risk technologies* (2nd ed.). Princeton, New Jersey: Princeton University Press.

Rasmussen, J., Pejtersen, A. M., and Goodstein, L. P. (1994). *Cognitive systems engineering.* New York: John Wiley and Sons.

Reason, J. T. (1990). *Human Error.* New York: Cambridge University Press.

Reason, J. T. (1997). *Managing the risks of organizational accidents.* Aldershot, England: Ashgate.

Special Commission on the 1999 Texas A & M Bonfire. (2000). *Final report.* College Station, Texas: Texas A & M University.

Woods, D. D., Johannesen, L. J., Cook, R. I., and Sarter, N. I. (1994). *Behind human error: Cognitive systems, computers, and hindsight.* Wright-Patterson Air Force Base, Ohio: Crew Systems Ergonomics Information Analysis Center.

Multi-disciplinary Evaluation of Information Visualisation

Fraser Speirs and Chris Johnson

Dept. of Computing Science, University of Glasgow, Glasgow, G12 9QQ, Scotland.
{speirsfr, johnson}@dcs.gla.ac.uk
http://www.dcs.gla.ac.uk/~{speirs,johnson}

Abstract: The problem of evaluating information visualisation systems for safety-critical applications is one of validating the design against real-world expert user requirements. The population of experts in a given domain is frequently very small indeed. At most there may be an organisation numbered in the tens of employees, or perhaps 100 in a large government agency. The difficulty of both gaining access to and conducting experiments with these experts is substantial. This paper describes the process which was taken to evaluate a novel information visualisation system with the experts from the UK railway safety authority. Evaluations of the system were also conducted with local users in order to be sure that the system was indeed usable at a very basic level . This process is still on going, and we hope to develop stronger ties within the expert community in the UK railway industry.

Keywords: Experts, design, evaluation, heuristic evaluation, prototyping.

Introduction

Users often find it difficult to express their needs and desires of software to a designer. The capabilities of the visualisation tools are often outwith the regular experience of many computer users. This makes traditional requirement gathering somewhat more complex than for many traditional classes of applications. When developing information visualisation systems, it is often necessary to give users a concrete example of what a system may be capable of before attempting to conduct evaluation.

This paper describes our experiences in designing and evaluating a novel information visualisation system for Signals Passed at Danger incidents on the UK railway network. We discuss briefly the process of approaching the problem as an 'industry outsider' and the need to make users understand how visualisation can be applied to their problem. We then discuss methods of evaluation, both with local non-experts and practicing experts in industry.

Design

As 'outsiders' in the industry, we felt that we had to be able to demonstrate an application of visualisation that is relevant to the domain in order to gain the participation and understanding of domain experts. To this end, we set out to develop a tool that allows visual interaction with information about SPAD incidents on the UK railway network.

Requirements Gathering Process: To develop requirements of our system, we used a variant of the spiral model presented by Kotonya and Somerville (1997). The main difference between our process and the more typical requirements gathering is that we did not have any contact with potential users of the system in our first iteration, for reasons discussed in section 1. We initially took a number of resources and information sources that are published by the organisations in charge of railway safety, and thought through some scenarios in which the SPAD data might be used (HMRI 1999, HMRL 2000, HSE 2002). These organisations were, specifically, the Railway Inspectorate of the UK Health and Safety Executive, and Railway Safety, an independent subsidiary of Railtrack Group PLC, the parent company of Railtrack PLC UK rail infrastructure controller.

Requirements: Our requirements elicitation process generated three guiding principles for our design. These were:

1 Visualisations must support interactive exploration of the data set

No a priori design of an information layout, such as the statistical graphs that the Railway Inspectorate currently use, can ever satisfy all potential uses of the information. Any visualisation must allow a user to pose arbitrary queries to the data.

- Visualisations must support useful display of location information

A primary role for incident reporting is to identify locations that may be particularly notable for the number of incidents occurring there. With SPAD incidents, a multiple-SPAD signal may be sited poorly, badly aligned, obscured by undergrowth or affected by the sun at particular times of day. A visualisation system for SPAD reports would need to include some means of indicating problem areas. It seemed a natural requirement that visualisations support correlating incidents and geographical locations.

- Visualisations must support correlation of incidents with common attributes

A major issue with the Railway Inspectorate's static presentation of the data is that it makes it difficult to correlate incidents with particular attributes. An appropriate visualisation should support the user's task in finding incidents with attributes in common. We developed an initial prototype application based on these major requirements. The design is discussed in more detail in Speirs and Johnson (2002), but is discussed briefly here.

Prototype Design: The design of the user interface of our prototype draws heavily on the Dynamic Homefinder application of Williamson and Shneiderman (1992). The interface is based around a starfield display (Ahlberg and Shneiderman, 1994, 1994a) - a map of the UK, upon which are superimposed small dots each representing a SPAD incident, at approximately the correct geographical location. The starfield display is manipulated by a number of controls aligned down the right hand edge of the window. Amongst these controls are double-ended sliders, list boxes, checkboxes, buttons and pop-up menus which are used to define the query or to set other attributes of the display, such as point colour.

SPADBrowser provides the user with three double ended sliders that control parameters of the query: the 'overshoot' distance beyond the signal, the severity of the incident and the time of day. A list control allows the user to restrict the display to incidents involving particular Train Operating Companies. Multiple selection is allowed in this list, so the user can define their own combination of Train Operating Companies' incidents to be displayed.

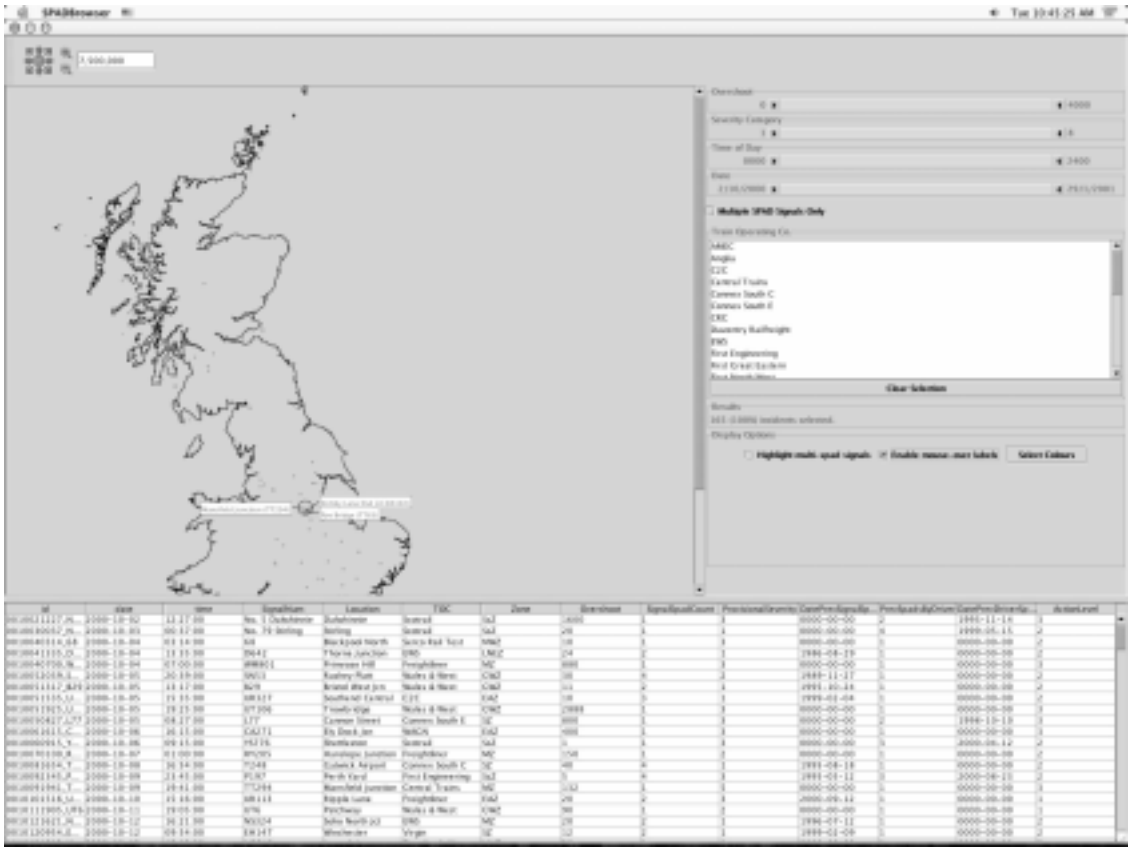


Figure 1 - Prototype screen layout

Evaluation Design

A significant difficulty in evaluating specialised applications such as a SPAD analysis tool, is that it is a domain experts' tool. The number of practicing domain experts is very small indeed. This is in contrast to evaluating a more common software tool, such as a web browser, where the number of potential users is much higher. In addition more, and more representative, test subjects are easier to recruit. In order to overcome this problem, we decided on a three-part evaluation.

Heuristic evaluation: In the early stages of development, we employed a Heuristic Evaluation technique (Neilsen 1992, Neilsen and Molich 1990). At this point, we applied some of the emerging heuristics of information visualisation tools to our system. Shneiderman gives some guidance on the expected properties of information visualisation systems (Shneiderman 1996):

- Overview: Gain an overview of the entire collection
- Zoom: Zoom in on items of interest
- Filter:Filter out uninteresting items
- Details-On-Demand:Select and item or group and get details when needed
- Relate:View relationships among items
- History:Keep a history of actions to support undo, replay and progressive refinement
- Extract:Allow extraction of sub-collections and of the query parameters

The point at which we were conducting heuristic evaluation led us to see this stage of evaluation as partly formative and partly summative evaluation. Some of the properties of visualisation and dynamic query systems that Ahlberg discusses fed into our initial design, rather than being used as heuristics to criticise a design after the fact (Ahlberg and Shneiderman 1994, 1994a). For example, Williamson reports that dynamic querying works well when coupled with a starfield display (Williamson and Shneiderman, 1992). This was a result that was formative at the design stage.

Non-expert evaluation: With an initial completed prototype, we conducted a series of evaluations with local users. These users were non-experts. All were either third, or fourth year Computer Science students. The purpose of this evaluation was to get an overview of the usability of the prototype, rather than to assess the suitability of the visualisation for the expert task in hand. We felt that it was important to consider the low-level usability issues, such as whether users could perceive items in the display, or if they could effectively use the double ended sliders, alongside the more task-oriented expert evaluation. Our use of non-experts from Computer Science allowed us to be sure that the subjects would have some degree of familiarity with computers. This meant that we could gather feedback without extensive basic computer training.

We conducted the non-expert evaluations by running experiments comparing our prototype to existing web-based techniques of presenting SPAD data. The Railway Inspectorate's SPAD report each month contains information about both the incidents for the relevant month, but also contextual information about trends in incident rates. Statistics are employed to analyse the information presented in these reports, which are distributed as web pages from the Railway Inspectorate home page. There are a number of issues relating to the design and presentation of these pages that obstruct easy navigation of the data.

The general layout of the website presenting the monthly SPAD reports is illustrated in figure 2. Each month's report is presented as a separate sub-page of the main page, but monthly reports are not linked together. This makes month-to-month comparisons difficult, except where HMRI provides a graph showing comparative monthly statistics.

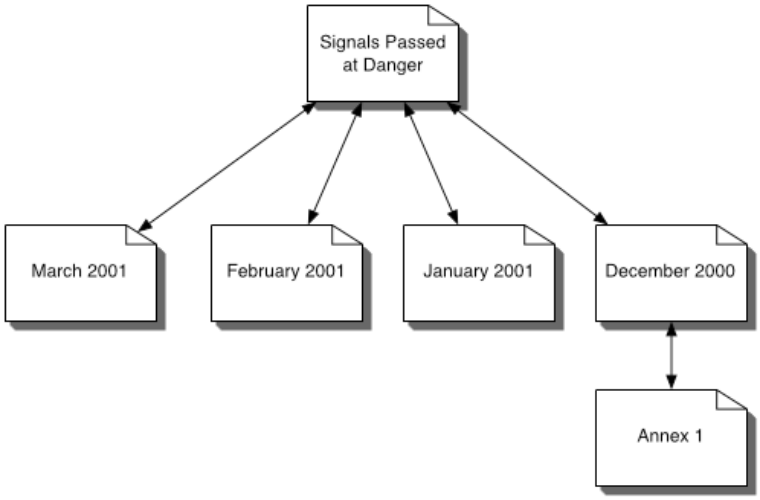


Figure 2 – HMRI SPAD Report Website

The aim of our evaluation was to compare the prototype to the existing web-based presentation, and determine if users experienced any benefit from using a visualisation tool over an HTML table. In recent months, the data has been presented in a different file format - Adobe's Portable Document Format - presenting a further difficulty in performing comparisons between monthly data sets. Users are asked to rate their agreement with five statements made about SPAD data on a scale between one and seven. There are two sets of evaluation questions and each user performs both sets of questions in separate sessions. Users are randomly assigned to a particular ordering of question sets and interfaces in order to offset any learning effects that may occur. For the purposes of evaluation of the interface, the statements were also designed to exercise as many parts of the tool as possible, from the use of double-ended sliders to the perception of points in the starfield display.

- Set 1

1. Events at Multi-SPAD signals constitute around 40 percent of all SPADs
2. Most SPADs have a severity category of three or greater
3. Most SPADs involve an overshoot of less than 200 yards
4. The number of incidents at Multiple SPAD signals is increasing from month to month
5. Incidents involving a signal that has previously been passed at danger usually also involve a driver that has been involved in a previous SPAD

- Set 2

1. Railtrack Midland Zone (MZ) is the zone with the lowest number of SPADs
2. The number of incidents is relatively stable from month to month
3. SPADs are more common in the morning (midnight-noon) than in the evening (noonmidnight)
4. The incident with the longest overshoot distance ocurred in Manchester
5. No SPAD has occurred north of the incident at Perth Yard (signal P197)

Subjects were asked to complete a questionnaire by circling items on a Likert scale similar to the following:

Railtrack Midland Zone (MZ) is the zone with the lowest number of SPADs?
Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

User Metrics: For each statement, we have an expected answer that a user would generate if they were perceiving the data correctly. User performance is measured by variation from the expected result. The greater the variation, the poorer the user's performance. Our hypothesis states that users' responses will vary less from the expected answer when using the visualisation tool than when using the webbased presentation. Response time is not measured in this experiment. We are trying to assess how accurately users can extract information through each interface, irrespective of the time they take to do it. As well as the formal questionnaire answers, users were also informally observed in their usage of the system. From this, it was hoped that some usage patterns would be observed.
Methodology: Users were given a set of four training questions designed to focus their attention on as many parts of the interfaces and data set as possible. During the training phase, users were permitted to ask questions about the use of the tool or the meaning of fields in the data set. This phase was unlimited in time. When the users had completed the training, they were given the appropriate set of evaluation questions. When completing the evaluation questions, users continued to use the tool to develop their answer. For the actual evaluation work, users were not permitted to ask questions.

Expert Evaluation: The final part of evaluating the prototype system, which is still on-going, is the evaluation of the system with respect to the tasks that experts actually perform. In essence, at this point we are using our prototype as a vehicle for evaluating the correctness of the assumptions made at the initial design phase in section 2.3. The first phase of expert evaluation took the form of informal presentations of the software to members of the UK railway safety community. In these sessions, we presented the prototype and held discussions about the possible applicability of visualisation to the domain. The initial results have been interesting and valuable.

Evaluation Results

The results of our evaluations have been encouraging. This section discusses some of the outcomes of each phase of evaluation.

Subject	Web	Visualisation	Improvement (web - visualisation)
A	9	9	0
B	9	6	3
C	4	6	2
D	10	4	6
E	1	13	-12
F	5	2	3
G	10	5	5
H	10	4	6
Average	7.25	6.125	1.125
Std. Deviation	3.25	3.54	1.13
Average (Excl. E)	7.125	4.5	2.625

Table 1 - Preliminary user evaluation results

Heuristic Evaluation: The results of the heuristic evaluation resulted in numerous small improvements in the workings of the user interface. The most immediate result was that the performance of the system needed to increase. The initial version of the system used a database as it's back-end storage and system performance was sluggish at best. This led us to find some optimisations in the existing system, and to consider using an in-memory data representation for future versions of the system.

Non-expert Evaluation: The non-expert evaluation gave us some encouragement that our system was indeed usable, and represented an improvement over current methods of presenting SPAD data. Almost all users showed reduced variance from expected results when using the Visualisation interface than when using the web interface.
A number of interesting insights into users' use of the visualisation tool have arisen during the evaluation. Firstly, many users tend to focus on the tabular view of the data rather than the map view. The indications are that the table is of importance when numerical data is in question, but also when comparisons are being made. We found that users would use the sliders to get to within a certain distance of the search target and then visually scan the table for the correct answer. A common feature request from non-expert users during evaluation was the ability to sort the table. The results obtained from our questionnaire are shown in table 1. Positive numbers in the 'Improvement' column indicate an increase in accuracy from using the visualisation interface.

Expert Evaluation: Our expert evaluation process is ongoing, but already we have gained substantial benefits from it. One of the main objectives in early evaluation was to understand whether or not the early design assumptions that we had made were valid in the domain context.
To achieve this, we used comprehensive walkthroughs of the system with multiple experts. We also spent time discussing the applications of visualisation from the point of view of domain experts' tasks and from the technological advantages of visualisation. We found that our evaluation sessions developed naturally into a form of cooperative evaluation. The domain experts we were working with took naturally to participating as if they were participating in the development of the system, which they effectively were. Many suggestions and new and surprising outcomes sprang from these sessions.
The outcome of our first evaluation session was extremely enlightening. The map-based display that we had developed was, in fact, not considered to be particularly useful for analysing SPAD data. One commentator said that "all the map shows you is where the most traffic is". This showed us that the actual visualisation of the data that we had chosen, and the assumption in section 2.3, that 'visualisations must support useful display of location information', was not correct. The most promising outcome of our initial expert evaluation was a suggestion to replace the map with a charting tool instead. It was felt that the ability to dynamically query the data was extremely valuable, but that the mode of presentation was not applicable to SPAD analysis. The existing charting techniques that are used will remain the key means of communicating information about and monitoring SPAD performance on the railway network. This confirms another of our assumptions, that 'visualisations must support interactive exploration of the data set'.
Another interesting outcome to us was that there were types of analysis, other than SPAD analysis, for which a map based display was considered to be very useful. In particular, the Suicides and Open Verdicts on the Railway Network (SOVRN) research project, and the Trespass and Vandalism (TV) project were both considered suitable

projects for the application of map-based visualisation. This is because the incidence of suicide, trespass and vandalism are not related to volume or density of rail traffic, but often to social issues in specific areas, such as low average income or high unemployment. One of the experts we met described how he had spent a week with a map of the UK on a pin-board manually plotting all the incidences of suicide, trespass and vandalism. We feel that our initial approach of developing an initial prototype in order to help experts understand the possibilities of information visualisation has been validated. Figure 3 shows an early version of our redesigned prototype.

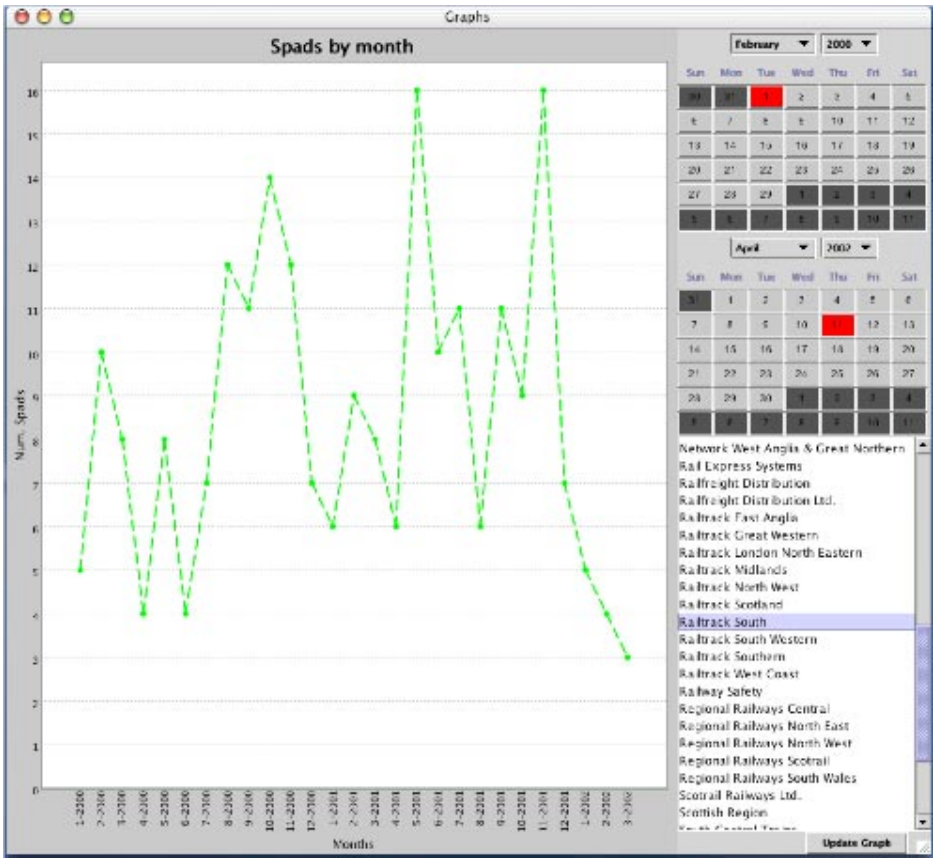


Figure 3 – Redesigned prototype

Conclusion

This paper has discussed a particular approach to designing and evaluating novel user interfaces. The difficulty in evaluating advanced user interfaces is that users may be unable to imagine what such an interface might look like, or to appreciate the way in which it might be applied to their domain of expertise. Our approach was firstly to make some deductions about the users' tasks based on the publicly available information. We then developed a prototype based on these assumptions, which was subsequently used for evaluation purposes. The results of both our non-expert and expert evaluations have been encouraging and very useful.

We believe that this approach to evaluation has shown benefits such as being able to acquire more useful feedback than would be possible with a formative-type interview where the user has no experience or example of information visualisation. We found that our sessions could become constructive very quickly if a supporting example application was shown. However, there is a certain degree of risk involved in this approach. Conducting the formative design stage without expert input leads to an increased risk of being 'surprised' when one eventually presents to expert users. This turned out to be our experience, when we discovered that the map-based display was not as appropriate as we had expected.

Going forward, we hope to redevelop our first prototype to better suit the tasks that experts require to perform. At that point, we intend to conduct further, more detailed, evaluations with domain experts.

References

Ahlberg and Shneiderman (1994). Tight coupling of dynamic query filters with starfield displays. In Proc. CHI'94, ACM Conference on Human Factors in Computing Systems, pages 313-317,1994.

Ahlberg and Shneiderman (1994a). Visual information seeking using the filmfinder. In Conference Companion of CHI'94, ACM Conference on Human Factors in Computing Systems, 1994.

HMRI (1999). Her Majesty's Railway Inspectorate. Signals Passed at Danger (SPAD)s Report For October 1999. Technical report, The Health and Safety Executive, 1999.

HMRI (2000). Her Majesty's Railway Inspectorate. Signals Passed at Danger (SPAD)s Report For March 2000. Technical report, Health and Safety Executive, London, United Kingdom, 2000. http://www.hse.gov.uk/railway/spad/march00.htm.

Kotonya and Somerville (1997). Requirements Engineering: Processes and Techniques. Wiley.

Nielsen (1992). Finding usability problems through heuristic evaluation. In Conference proceedings on Human factors in computing systems, pages 373-380. ACM Press, 1992.

Nielsen and Molich (1990). Heuristic evaluation of user interfaces. In Conference proceedings on Empowering people : Human factors in computing system: special issue of the SIGCHI Bulletin, pages 249-256. ACM Press, 1990.

Shneiderman (1996). The eyes have it: A task by data type taxonomy for information visualizations. Proc. 1996 IEEE Conference on Visual Languages, 1996.

Speirs and Johnson (2002). Designing information visualisation for incident databases. In Proceedings of the 20th International System Safety Conference, 2002.

HSE (2002). The Health and Safety Executive. Assessment of Railtrack's management of multi-SPAD signals. Technical report, The Health and Safety Executive, 2002. http://www.hse.gov.uk/railway/spad/manmss.pdf.

Williamson and Shneiderman (1992). The dynamic homefinder: evaluating dynamic queries in a real-estate information exploration system. In Proceedings of the Fifteenth Annual International ACM SIGIR conference on Research and development in information retrieval, pages 338-346, 1992.

Application of Why-Because Graphs to Railway Near-Misses

Jens Braband, Bernd Brehmke,

Siemens AG Transportation Systems, P.O. Box 3327, D-38023 Brunswick, Germany
e-mail: {jens.braband|bernd.brehmke}@siemens.com

Abstract: Observations based on experience with the root cause analysis of near-misses in rail transport are described. A causal analysis methodology using a simplified version of Ladkin's Why-Because (WB) graph is presented and the advantages and disadvantages of this approach, including the critical success factors for an industrial root cause analysis programme, are discussed.

Keywords: root cause analysis, near-misses, Why-because graph, success factors.

Introduction

Although rail transport in Europe has generally achieved a very high level of safety (European Transport Safety Council, 2002), even catastrophic accidents do happen from time to time (see Danger Ahead (2002) for historical and recent examples). A comprehensive study has not yet been carried out, but several examples indicate that the root causes of rail accidents are similar to those for aerospace accidents (Leveson et al., 2001), which have been broadly classified as follows:

- Flaws in the safety culture
- Ineffective organisational structure and communication
- Ineffective or inadequate technical activities

Unfortunately none of these categories, except perhaps the last, are considered in the risk analyses which form part of the safety cases for railway signalling systems. As a rule, safety-related products in the rail industry have to be assessed and certified. This normally implies compiling a documented safety case, in which typically analyses such as Failure Modes and Effects Analysis (FMEA) and Fault Tree Analysis (FTA) are performed. So how come things still go wrong? In our experience, one problem is common-cause failures (CCFs) and another complexity. CCFs are often related to human factors, which are generally not addressed in technical safety cases. Although it may be feasible to analyse a system with, say, 100 components and 10 failure modes on average per component by FMEA, a full analysis of all the combinations of double, triple etc. faults is impracticable. "Creative" analysis techniques such as FTA therefore concentrate on the most likely combinations, or those which have caused problems before. Often attention focuses on aspects which can be quantified. In practice, however, it may be that a problem e. g. has four root causes or that events assumed to be independent are not. Another general problem is that the scope of a product safety case is often limited to a particular system definition and a particular environment. If a system integrator buys, say, three generic products from different suppliers and integrates them into a new system, a standard interface safety analysis will not cover all the possible combinations of faults from all the subsystems.

Thus the remaining problems, due largely to apparently unlikely combinations of multiple faults in complex systems, need to be addressed by some other means. As the validity of the "iceberg model" or "safety pyramid" (see Figure 1) has also been corroborated for railway systems (Nolte, 1993), a natural way forward seems to be a rigorous analysis of near-misses as proposed by The Near-Miss Project (2002). This was discussed after the Concorde disaster, where it transpired that tyre bursts had already been reported on 57 previous occasions, of which 6 had resulted in fuel tank leakages (Kochs, 2001).

To our knowledge, root cause analysis is already applied by several approaches within the rail industry, generally driven by quality management considerations. Hitherto, however, we have seen only a few published results. Although accidents on the other hand are investigated and reported in textual form (usually at the behest of the safety authorities), we have yet to find an approach with a structured root cause analysis or graphical representation. Frequently, numerous parties (from subcontractors to the railway operator) contribute to the root causes of an accident and it would greatly enhance the investigation and reporting of such incidents if a standard approach could be adopted.

For these reasons, the Rail Automation division of Siemens Transportation Systems has undertaken research in two directions:

- How can the reporting and analysis of (internal) near-misses be enhanced and standardised?
- How can accident investigation and reporting benefit from a structured approach?

This work has been carried out in a pilot project called ”Total Safety Management”, which, following successful application in one department, was extended to the entire division. Preliminary results of this ongoing work are presented in this paper, in particular conclusions drawn from the use of a simplified WB graph for near-misses.

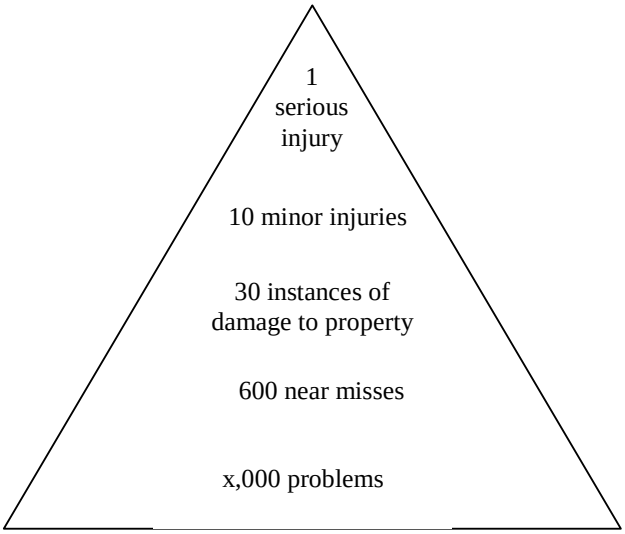


Figure 1 - The iceberg model

Analysis of near-misses

We agree with the general definition that a near-miss is *an opportunity to improve safety practice based on a condition, or an incident with the potential for more serious consequences* (The Near-Miss Project, 2002). While it is recognised that many companies undertake near-miss analysis in some form or another as part of their quality management scheme (Andersen and Fagerhaug, 2000), certain vital questions remain:

1. How do we know and ensure that near-misses are reported at all?
2. How do we know that all the major root causes are identified?
3. How do we ensure that countermeasures for all root causes are defined and implemented, and monitored until the problem has been solved?

Near-miss reporting:

It is a well-known fact (The Near-Miss Project, 2002) that companies which successfully encourage their employees to report near-misses tend to create safer products or operate safer systems than companies where only few problems or none at all are reported. The problem of near-miss reporting can only be solved by having a company-wide culture which makes problem-solving, not the penalisation of the party responsible, the top priority. To achieve this, there must be a continuous commitment on the part of senior management to the issue and the publication of success stories. Another thing which has proved to be very helpful is to have the party responsible for the problem involved in the analysis, because they frequently have the best ideas on how to prevent any recurrence. This is only possible in a positive ”no blame culture”. By comparison, the use of tools is a side issue.

In the Rail Automation division at Siemens Transportation Systems employees have been sensitised via e-mail and the Intranet by the top management, as well as at project kick-off meetings by department heads. The reporting of near-misses has also been facilitated by an easy-to-use Intranet form. Once an anticipated initial resistance had been overcome, the employees soon began reporting potential problems on a large scale.

A critical issue here is the definition of near-misses and the indenture level in the iceberg model. When a problem is reported, the number of near-misses may be very high and the effort required for analysis massive and often unjustifiable. However, adding a large number of restrictions might result in the employees becoming unsure about which problems to report and which not, and valuable information could be lost. Finally, it was decided to broaden the scope and not look at safety problems alone, but also at major availability and quality problems. Instead of restricting the general definition, minor problems were to be filtered out in a second step after the reporting. This produced a very low indenture level in the iceberg model. Figure 2 shows the general project organisation.

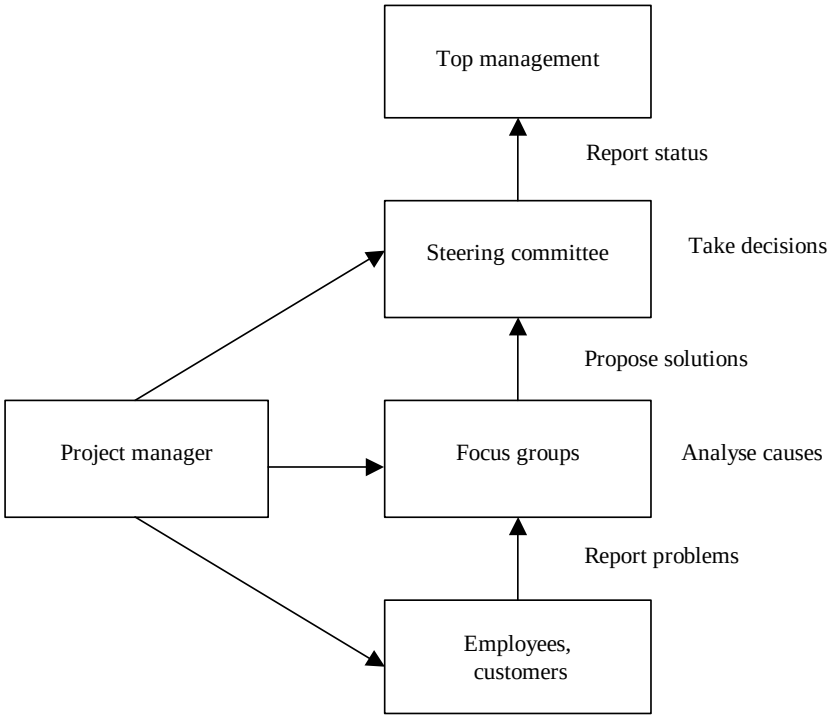


Figure 2 - Project organisation and tasks

Causal analysis:
Even where near-misses are reported, the analyses we have come across thus far often lack completeness or clarity (Andersen and Fagerhaug, 2000). Some companies have a principle whereby ”an action must result from each near-miss investigation”, but how do we know that one action is sufficient?
We have therefore examined structured root cause methodologies on the basis of following requirements:

- The method should be easy to use (with a minimum amount of training, preferably requiring no proprietary tool) by the average engineer.
- The method should provide a graphical representation (”a picture says more than a thousand words”).
- The method should allow modular approaches (different aspects analysed by different individuals).

The method which in our (subjective) evaluation came closest to these goals was Peter Ladkin’s Why-Because Analysis (WBA) (Ladkin, 2001), which is built on a formal causal model (Lewis logic) and has already been successfully applied to a number of aviation incidents (Ladkin, 2002). However, the initial field experiments with engineers were somewhat unsatisfactory because the engineers tended to be confused by the WBA terminology for events, states, non-events etc. and ended up discussing the terminology rather than the causal factors. This observation led us to simplify the procedure considerably. A precise distinction between events and states only seemed to be necessary if formal proof was being undertaken and this was not the case; we wanted the method to be applied by ordinary engineers – not formal method experts – to everyday problems in the course of their work.
The actual analysis was performed by a number of focus groups, which were created for the different product lines and technologies concerned and included the party responsible for the problem as well as other stakeholders, e.g. subcontractors, where necessary. Figure 3 shows the elements we defined for the simplified WB graph. The graphical representation was simplified in order to allow it to be drawn using standard MS Office tools.

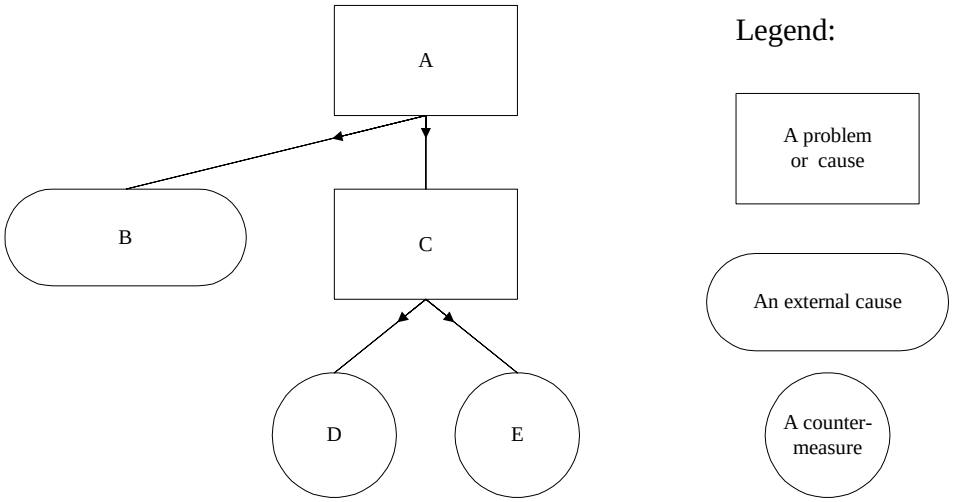


Figure 3 – Elements of a simplified Why-Because (WB) graph

The meaning of the letters in Figure 3 is as follows:

- A is the initial starting point of the analysis (problem).
- B and C are direct causes of A. If either B or C had not happened, nor would A have either.
- B is an external cause (with respect to the system under consideration).
- D and E are countermeasures with respect to the root cause C. If D or E were to be implemented, C would not be expected to recur.

In practice, the approach varied according to the complexity of the problem. While in many cases a simplified WB graph was sufficient because the problem could be easily understood (and reviewed), in more complex settings it was quite hard to review and document the problem based on the WB graph alone. For such cases a more advanced, yet still very simple, documentation scheme (see Figure 4) was devised, making extensive use of the hyperlink facility in MS Office tools. The documentation starts with a chronological chain of events, which is hyperlinked to the causal analysis, usually represented by WB graph but sometimes in simple text form (where the drawing tool was not available). The WB graph is in turn hyperlinked to the detailed text sheets with the countermeasures, as well as to source documents. As a result it is easy to trace an event from the problem report to its role in the causal analysis and to the countermeasures triggered. Depending on the implementation, the links can point either forwards or backwards or both, and they have proved to be a great help particularly when reviewing complex interrelationships.

Implementation of countermeasures and control of effectiveness:

This simple scheme has now become widely accepted among the engineers and is implemented throughout the Rail Automation division. The results are reviewed and decisions on the implementation of the countermeasures discussed and taken by a steering committee comprising the heads of the engineering departments and senior management staff.

Each countermeasure and its implementation status are described by a single sheet. If accepted, the effective implementation of the countermeasures is controlled by the steering committee using a closed-loop feedback system which tracks the implementation progress on a scale from 1 to 5:

- 1 means that the countermeasure has been identified by WB graph.
- 2 means that a detailed implementation plan and schedule have been defined.
- 3 means that all the responsible personnel have been assigned to the plan.
- 4 means that the plan has been executed.
- 5 means that the countermeasures have been applied.

The implementation of the countermeasures is also reported to senior management in regular summaries specifying the number of countermeasures, the relative percentage of countermeasures in each level and the countermeasures which are behind schedule to an excessive degree.

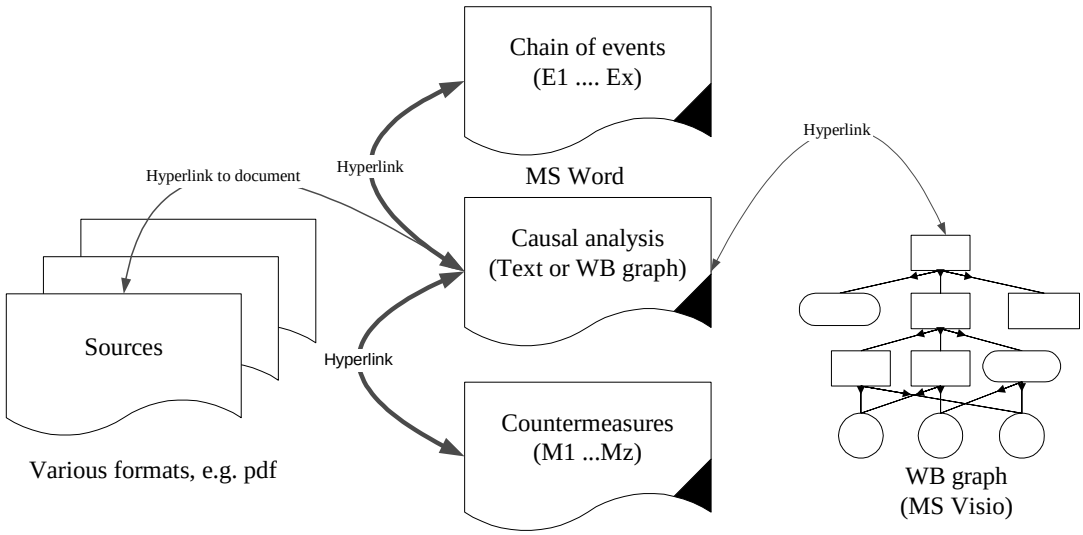


Figure 4 - Complete documentation scheme for complex problems

Extensions and further work

At Brunswick Technical University (Germany) the original WBA was applied to a well-documented case, the Ladbroke Grove accident (The Ladbroke Grove Rail Inquiry, 2001). The resulting graph (De Stefano, 2002) showed a combination of 32 root causes to be responsible. It is planned to extend the scope of the work by consistently applying the methodology to other accidents and near-misses in co-operation with rail operators, the rail industry and safety authorities.

Another interesting task would be a statistical analysis (e. g. trend analysis) and classification of the root causes and a comparison with aerospace root causes as presented in Leveson et al. (2001).

Summary and conclusions

Ladkin's original Why-Because Analysis has been both simplified and extended with respect to the following features:

- the formal rigour has been relaxed in order to increase its usability for engineers
- the symbols have been simplified and adapted to MS Office tools
- proposed countermeasures have been directly included in the graphs
- hyperlinks to the documentary evidence have been included in the WB graphs

The application of simplified WB graphs embedded in an overall problem resolution process has been very helpful in identifying potential for improvement in products and processes, even where they had already been assessed and certified. However, it should be noted that the analysis technique itself is only one link in a larger chain and that only the weakest link matters (see Figure 5). Major success factors in this approach were:

- Continuous senior management commitment, helping to implement a "no blame culture"
- A simple, but structured analysis technique
- Effective control of the implementation of countermeasures

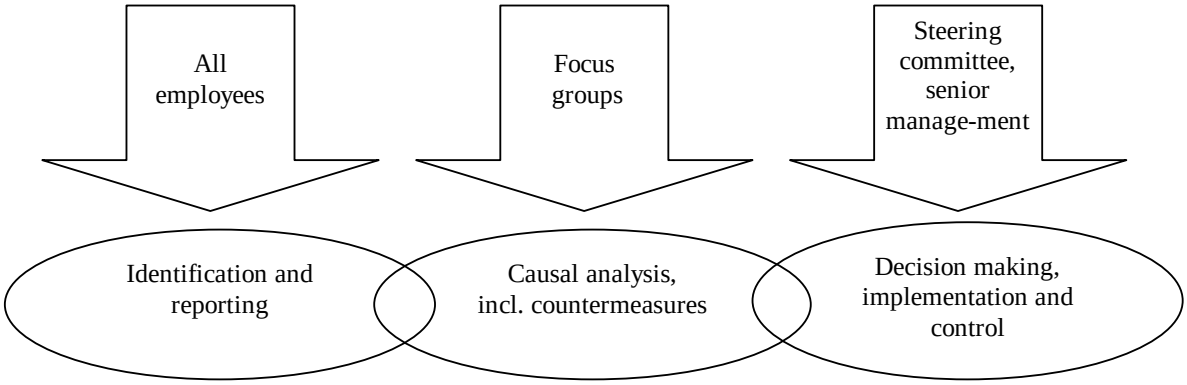


Figure 5 - The root cause analysis process model

References

Andersen, B., Fagerhaug, T. (2000). Root Cause Analysis, ASQ Press, Milwaukee.
Danger Ahead (2002). Historic Railway Disasters, http://danger-ahead.railfan.net
De Stefano, E. (2002). Application of WBA as a method for the causal analysis of railway accidents (in German), Institut für Eisenbahnwesen und Verkehrssicherung, Technische Universität Braunschweig.
European Transport Safety Council (2002). http://www.etsc.be/index.html
Kochs, H.-D. (2001). A disaster is the result of a long chain of faults (in German), VDI-Nachrichten, 26.10.01, see also http://www.concordesst.com
Ladkin, P. (2001). Causal System Analysis – Formal Reasoning About Safety and Failure, University of Bielefeld.
Ladkin, P. (2002). http://www.rvs.uni-bielefeld.de
Leveson, N. et al. (2001). Evaluating Accident Models using Recent Aerospace Accidents, Part I: Event-Based Models, NASA and MIT, http://sunnyday.mit.edu
Nolte, K. (1993). Collection and analysis of incident data (in German), Proc. Der Verkehrsingenieur, Technische Universität Dresden, 151-156
The Ladbroke Grove Rail Inquiry (2001). http://www.hse.gov.uk/railway/paddrail/lgri1.pdf and http://www.lgri.org.uk/
The Near-Miss Project (2002). University of Pennsylvania, http://grace.wharton.upenn.edu/risk/proj/nearmiss.html

Using Accident Data to Forecast Societal Cost within the Framework of an Axiomatic Safety-Critical Assessment Process (ASCAP) Simulation

D. E. Brown and J. Stile,
Department of Systems and Information Engineering
University of Virginia
Charlottesville, VA 22903

Abstract
The Axiomatic Safety Critical Assessment Process (ASCAP) (Kaufman and Giras 2000; Monfalcone, Kaufman and Giras 2001) is a simulation methodology that models complete transportation systems and analyzes the effects of system modification. This paper outlines a methodology for use within the ASCAP simulation engine that assigns a predicted severity (societal cost) to a simulated accident through the use of a regression tree. The regression tree was trained on data obtained and corrected by the Federal Railroad Association Office of Safety / Office of Railroad Development. The regression tree uses the quantitative variables train speed and tons and the qualitative variables accident type, region, and visibility to partition the data. The data is partitioned into different terminal nodes that represent societal cost in log dollars. The overall total mean square error of the tree is 1.133665 log dollars. The tree is produced as part of an evolving process in which expert opinion is used to shape both the data and the resulting tree.

Keywords: CART, Decision Trees, ASCAP, Severity, FRA, Regression Trees

Introduction
The need to effectively model transportation systems and predict their associated risk, where risk is measured as a function of accident cost, or severity, and the likelihood of occurrence, is a matter of extreme importance (Kaufman Giras 2000). Transportation systems, in particular, railway systems, serve as an integral part of society whether they are used to commute people to work or ship goods. Since the advent of these systems, industry has constantly looked to ways of improving their safety and associated risk.
The issue of system safety is especially important in the cases where the current system is to be improved or added to. This augmentation can be as simple as adding new track components, such as roadway-rail crossings and switches, or as complex as extending the current rail system by adding a whole new track structure. In order to demonstrate that the safety for the modified rail system meets or exceeds the minimum safety requirements, there is a definite need to test the changes prior to deployment. Due to the time needed to generate the rare event mishap data, it is neither cost effective nor practical to attempt to build and test a system modification prior to actual deployment. Therefore, a system simulation is needed. The Axiomatic Safety-Critical Assessment Process (ASCAP) (Kaufman and Giras 2000; Monfalcone, Kaufman and Giras 2001) is a simulation methodology that fulfills this need. Through simulation, ASCAP provides the statistical basis for predicting system risk by generating rare event mishap scenarios and data in lieu of testing prototypes in actual operational environments.
ASCAP is concerned with probabilistically determining the sequence of events that lead to various mishap scenarios as constrained by the operational environment to which a given vehicle is exposed. By identifying these mishap sequences, their risk potential and there likelihood of occurrence can be quantified (Kaufman and Giras). In an ASCAP simulation, the behavior of a given transportation system is modeled using a hybrid of time and event driven simulations. During simulation, the entire system is sufficiently tested to produce the rare event mishap data that is used to quantify system risk.
System risk is formally defined in the following manner. Given the occurrence of a simulated accident with a resulting severity, the risk for the i^{th} train and the m^{th} Mishap is calculated by equation (1):

$$Risk_i^m = function(Severity_i^m(MDV_i^m) \mid Mishap_i^m) \quad (1)$$

The ASCAP risk versus train miles traveled is the summation of all the risks estimated for a given ASCAP trial.
In This paper, we provide a methodology for forecasting accident severity, where severity is defined as the resulting societal cost of an accident, for use within the ASCAP simulation framework. Currently, ASCAP assigns accident severity by examining each simulated accident with a panel of experts subsequent to the ASCAP simulation. Severity is then assigned to each accident through a consensus of cost estimates reached by the panel. While this

method incorporates the much valued expert opinion, it proves to be tremendously time consuming, somewhat inconsistent, and infeasible in light of the multitude of simulations that will be performed. Therefore, a new methodology for forecasting ASCAP simulated accident severity that incorporates both expert opinion and historical real-world accident data is needed.

The methodology we set forth in this paper uses Regression Trees to predict the cost of an accident within the ASCAP simulation model. Regression trees have been widely used in the prediction of response variables and have had great success in medical diagnosis (Breiman et al., 1984), sensor fusion (Brown and Pittard, 1989), and power systems (Rovnyak et al., 1994). Regression trees prove to be a proficient methodology model because they incorporate both data mining and expert decision. The Trees are trained using historical data collected by the FRA from actual accidents and are applied to similar data provided by the ASCAP simulation. Because Regression Trees are easy to understand, the resulting tree is then examined by a panel of experts for final improvements. In this manner, the methodology benefits from both historical data and expert domain knowledge.

The paper is organized as follows: in the section to immediately follow, the second section, we outline the ASCAP simulation methodology; in the third section we detail the data we used for composing our analysis; the modeling techniques we used to forecast severity is described in the forth section; in section five we summarize the results of our model; and in the six section we provide overall conclusions.

ASCAP Methodology

In order to understand the methodology for assigning accident severity, it is first necessary to have a broad sense of the ASCAP methodology. Having a general understanding of how ASCAP works will better allow the reader to understand where the methodology established in this paper fits within the ASCAP simulation framework.

Prior to ASCAP the only defined method for devising railway system risk were the Corridor Risk Assessment Models (CRAM), CRAM I and CRAM II (United States Department of Transportation Federal Railroad Administration, 1998a, 1998b). These models estimate risk for rail systems by relating the occurrence and consequences of accidents as related to specific track features and traffic; that is; the models estimate the rate at which these accident occur by corridor related attributes independent of any changes in operating rules or other environmental changes (Kaufman Giras 2000). The CRAM models fail to depict the factors that contribute to accident scenarios or include expert opinion in assessing accident severity. Furthermore, the poison regression model that CRAM uses to describe the number of expected accidents and their expected cost is too complicated for railway transportation experts without a strong statistical background to understand. Having a risk assessment model that rail experts can understand is very important because this allows them to utilize their years of experience and domain knowledge to adjust the model to make it more accurate.

The ASCAP methodology, unlike the CRAM model, uses a vehicular-centric perspective to analyze a given transportation system. The ASCAP simulation reflects the simultaneous movement of *n*-vehicles concurrently from the perspective of each individual vehicle. The predicted actions of the humans operating the vehicle and the variety of physical devices encountered by the vehicle during travel predict the movement of each vehicle. Depending upon the conditions of interaction between the vehicle and these various entities, the resulting movement may generate a sequence of events that lead to a mishap.

The potential for a mishap exists when a vehicle is co-incident in both time and space with an unsafe condition. Unsafe conditions are described as results from violations of the prescribed safety-critical protocol that defines safe system operation. Such protocol violations result from inappropriate human action(s) and/or from the stimulation of hazards within the various physical devices that a given vehicle encounters (Kaufman Giras). Once the ASCAP simulation has determined that an accident has taken place, the methodology set forth in this paper is used to determine the cost of that accident.

Data

Before reviewing the regression tree model that assesses severity it is first necessary to examine the data available. The data used to create this methodology come from historical data and simulation data. Before training the model, it was important for us to only use data that are available in both the historical data and the simulation data. This section details these data.

Historical Data: Historical data was used to train the severity regression tree. The historical data comes from the accident data set used by the Accident Review Team (ART) in their report on Positive Train Control Accidents (United States Department of Transportation Federal Railroad Administration, 1998a, 1998b). These data were used because they had been thoroughly reviewed by the ART and are known to contain the most complete accurate set of accident data. The data set contains detailed information on approximately 600 actual train accidents that occurred

between 1993 and 1997. The same data set was subsequently used to compose the CRAM models discussed in the previous section. The data set contains temporal, spatial, environmental, and other information, such as "accident cause", about each accident. The following table (Table 1) lists the historical data available on a per accident basis. Each accident attribute is presented with a description and the type, whether the accident attribute is categorical (qualitative) or non-categorical (quantitative), of the data.

Table 1: Historical and ASCAP available data

Name	Description	Unit/categories	Type
Time	time of accident	Minutes	quantitative
AccType	type of accident	1-13	qualitative
Cars	# cars on consist	Cars	quantitative
CarsHzd	# cars w/ hazardous material	Cars	quantitative
Temp	Temperature	Degrees	quantitative
Visibility	Visibility	1-4	qualitative
Weather	Weather	1-6	qualitative
Hspeed	speed at accident	MPH	quantitative
TrnDir	Direction	1-4	qualitative
Tons	Gross weight of consist – power units	Tons	quantitative
TypEquip	Type of train	1-9	qualitative
TrkCls	FRA track Class	1-6	qualitative
TrkDen	Track Density	Track line/ mile^2	quantitative
TypTrk	Type of Track	(1-6)	qualitative
Loaded	Loaded cars	Cars	quantitative
Cause	Cause of accident	1-5	qualitative
AccDmg	Societal Cost Accident Damage	Dollars	quantitative
EngrMin	Minutes on duty (engineer)	Minutes	quantitative
Condctmin	Minutes on duty (Conductor)	Minutes	quantitative
REGION	Region	1-8	qualitative
Method	Method of operation		qualitative
PassTrn	Passenger train or freight	1-2	qualitative

Societal Cost: The societal cost for each accident is provided in the ART data set. The societal cost was estimated using a formula based on different accident costs. These costs cover track and rail equipment damages, accident clean up, injury/fatality compensations and evacuation compensations for each accident where applicable. A detailed analysis of how the aggregate costs form societal cost is contained in the U.S. Department of Transportation PTC Corridor Risk Assessment Study from which the accident data set was procured.

ASCAP Data: All historical accident data is collected from an accident data form (FRA Form F 6180.54) filled by the FRA for each individual accident. This form contains accident and train specific information. In order to provide a comprehensive accident analysis, the ASCAP simulation reproduces FRA Form F 6180.54 with the data produced from the simulation. This assures that the data provided by the simulation is as comprehensive as the data collected in non-simulation, real world practice. A snap shot of this data can also be seen in table 3.1

Modeling Technique

In this section of the paper we provide a general overview of the CART modeling technique, describe how the model was trained, and finally how the model was adjusted using expert domain knowledge.

CART: For predicting severity within the ASCAP simulation we consider the following problem: we have n observations D = $\{(X_i, Y_i): i = 1,\ldots,n\}$, where X_i is a vector of train accident attributes and features and Y_i is the corresponding severity or societal cost. The goal is to learn a classifier from $\{(X_i, Y_i)\}$, which can then be used to predict the severity value Y_j of any simulated accident attributes X_j. For this paper we use regression tress (Breiman et al. 1984) to solve this problem.

The goal of regression trees is to partition the data, using binary splits, into relatively homogeneous terminal nodes minimizing standard deviation or median absolute deviation within the node. The predicted value at each terminal node is the mean or median value of the data in that node. Regression trees may be larger and more complex than classification trees due to the continuum of possible values of the response variable. The method of building regression and classification trees, outlined in Breiman, et al (1984), is comprised of three stages. These stages are detailed in the sub sections to follow.

Determining Splitting Rules: The splits are determined by choosing the values in a splitting rule that maximize the decrease in the error estimate R(T) for the ensuing node. The error estimate may be computed by assessing the least squares or least absolute deviation of all values in that node. The general form of the estimate decrease in error used to evaluate and select split points is shown in Equation (2), while Equation (3) shows the optimization used in selecting the spilt.

$$\Delta R(s,t) = R(t) - R(t_L) - R(t_R) \tag{2}$$

where:

$\Delta R(s,t)$ = change in error estimate at split *s* and node *t*,

$R(t)$ = error at current node *t*,

$R(t_L)$ = error at the new left sub node t_L, and

$R(t_R)$ = error at the new right sub node t_R.

$$\Delta R(s^*,t) = \max_{s \in S} \Delta R(s,t) \tag{3}$$

where:

$\Delta R(s^*,t)$ = change in error estimate for the best split at node *t*,

s^* = best split for node *t*,

S = the entire set of possible splits, and

$\Delta R(s,t)$ = change in error estimate at split s and node *t*.

The regression tree splitting rules tend to be more robust when compared to the classification tree splitting rules. Using this criterion, the best split at a particular node is the one that partitions the variables in such a way that the high responses and the low responses are segregated. Typically, there is a choice of two different criteria to determine the best splits at each node: least squares and least absolute deviation. The Least Squares (LS) splitting method is shown in equation 4.

$$R(t) = Min \sum_{x_n \in t} (y_n - \bar{y}(t))^2 \tag{4}$$

where:

$R(t)$ = error at node *t*,

$x_n \in t$ = predictor variables in the current node *t*,

y_n = current response for input variable x_n, and

$\bar{y}(t)$ =sample mean of *y* values at node *t*.

The Least Absolute Deviation (LAD) methodology is similar to the LS method, but uses a different approach to calculating error. In LS the mean of the data points at a single node is compared to the current predicted response

and this difference is squared. When using LAD, the calculation is based on distance or deviations from the center of the data taken as the median and an absolute difference is used. This is defined by equation 5:

$$R(t) = Min \sum_{x_n \in t} \left| y_n - \upsilon(t) \right| \tag{5}$$

where:

$R(t)$ = error at node t,
$x_n \in t$ = predictor variables in the current node t,
y_n = current response for input variable x_n, and
$\upsilon(t)$ = sample median of y values at node t

Although both LAD and LS should be compared to determine the best regression tree, LAD generally proves to be a better choice for model robustness in the face of outliers and heteroscedasticity.

Pruning the tree: The error estimate measure has the property that $R(t) \geq R(t_L) + R(t_R)$, meaning that the current error estimate is greater than or equal to the sum of the error estimates of the nodes created by the current split. At some point in growing the tree, a state is reached where no further improvements are possible. This state is brought on due to the lack of data to further split, or the data are very close is in value relative to the rest of the tree.
When any of these conditions occur, the tree is grown to its maximum value (T_{max}). The tree is then pruned back to simplify it and, ideally, minimize the error on the test data. Pruning the tree reduces the number of terminal nodes of the original tree. After reduction, the error of the tree is calculated (using either LS as in Equation 4.3 or LAD as in Equation 4.4). A cost-complexity term that penalizes larger trees is also added to the error value of the tree. The formula for the resulting cost function is given in Equation 6.

$$R_\alpha(T) = R(T) + \alpha \left| \tilde{T} \right| \tag{6}$$

where:

$R_\alpha(T)$ = error-cost complexity measure at α,
$R(T)$ = error measure for tree,
α = cost complexity measure that penalizes the larger trees, and
$\left| \tilde{T} \right|$ = cardinality of the set of terminal nodes in T

The decision on where to prune is made by determining the minimum sub-tree for α. This minimization problem is formulated in Equation 7, and the resulting minimization is a sequence of increasingly smaller trees, $T_{\max} \succ T_2 \succ ... \succ \{t\}$, where *{t}* signifies a single node tree.

$$T_\alpha = \arg\min_{T \leq T_{\max}} R_\alpha(T) \tag{7}$$

where:

T_α = complexity parameter of the tree at α,
arg min = argument that minimizes the function over the range shown,
$R_\alpha(T)$ = error-cost complexity measure, and
T = sub-tree of $T_{\max}$.

Using this criterion, ties are broken by choosing the smaller tree. In regression trees, the pruning process generally deletes two nodes at a time, making the pruning process a bit more tedious and time consuming in comparison to that of classification trees, in which large branches are cut at once.

Determining the Terminal Node Predicted Value: The prediction at the terminal nodes is the mean of the data in the node for least squares models and the median of the data in the node for least absolute deviation models. The error at each node is shown as the standard deviation for least squares and median absolute deviation for least absolute deviation.

Application to Accident Data: We apply the regression tree methodology to the accident data described in the third section of the paper. The performance analysis of the models is contained in the Results section of the paper. We use the Salford Systems CART package to construct the trees. The trees are trained using a ten fold cross validation process. In a ten fold cross validation, the data are divided into ten approximately equal and random subsets, and the tree growing process is repeated from the beginning ten times. In each cross-validation replication, nine data sections are used to train the data, while the tenth section is used to test the tree. This process is repeated ten times using a different test section, each time the error counts are computed. When the ten replications have been completed, the error counts from each of the ten test sets are summed to obtain the overall error of the final tree. Before the binary recursion algorithm created by the regression tree is integrated in the ASCAP framework, a panel of experts reviews it and makes any final adjustments.

Adjusting the Data and the Rule Set: The role expert opinion plays in the model building process can best be described by figure 1. Expert opinion serves a dual purpose in constructing the severity model. The first and probably most important role is to validate the historical accident data set. The data set, like most data sets, contains some incomplete and possibly incorrect data. Having an expert panel review the data improves the overall validity of the data set and allows the model to be trained to properly portray actual accident data.
The second role that expert opinion has is in adjusting the rule set produced by the regression tree. It is especially for this reason that decision trees are used to forecast societal cost. Because the tree is easy to understand, it allows experts with domain knowledge to adjust the tree to suit what they feel are accurate decision tree splits. The expert panel can adjust the tree in several ways. They can add an extra node, adjust the split value of one node, or prune the tree where they see fit. These adjustments to the original data set and the regression tree rule set continue until the expert panel feels that the tree accurately reflects there knowledge of accident severity. Using a readable model is invaluable to determining accident severity because it effectively places emphasis on both the expert opinion and historical data without favoring one over the other.

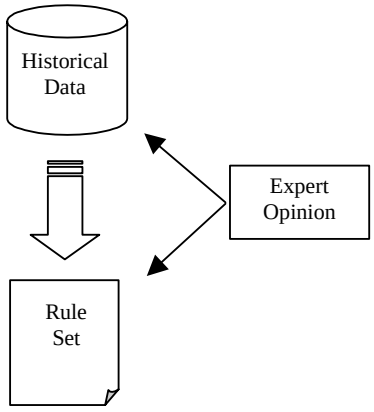


Figure 1: The model adjustment process

Adjusting Model Variance: In the preliminary severity regression model developed the variance at the terminal nodes of the tree were at too large of a scale for the tree to predict severity with any accuracy. Through analysis it was determined that data outliers cause the noted inaccuracy of Accident Damage model prediction. These extreme outliers can be seen in the histogram in figure 2 displaying accident frequency and societal cost.

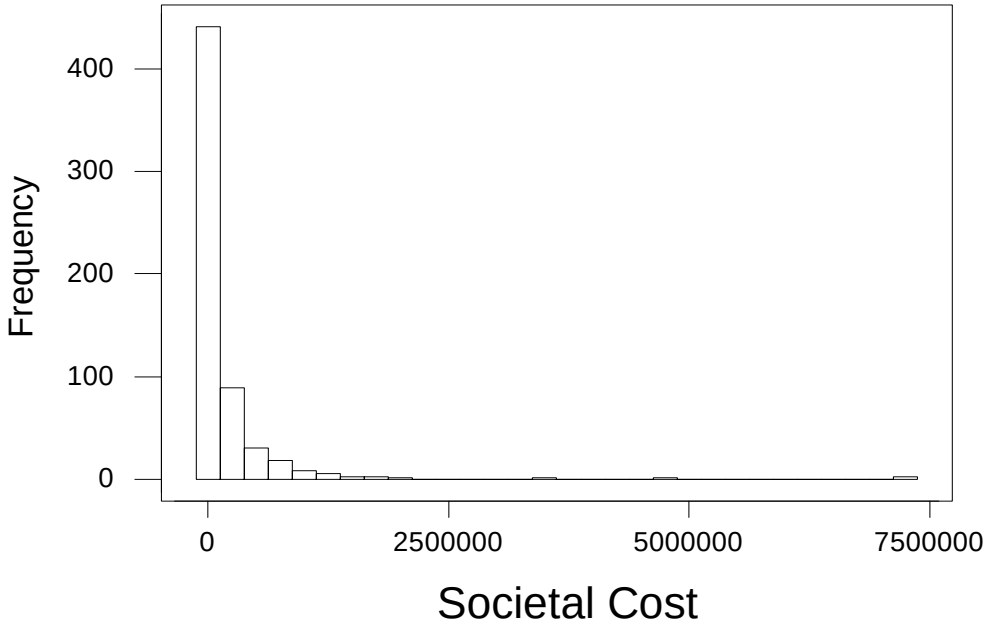


Figure 2: Accident Frequency and Societal Cost

The non-linearity of the accident cost variable can be seen in figure 3. Figure 3 matches the accident cost, along the y axis, to the nth accident arranged lowest to highest along the x axis.

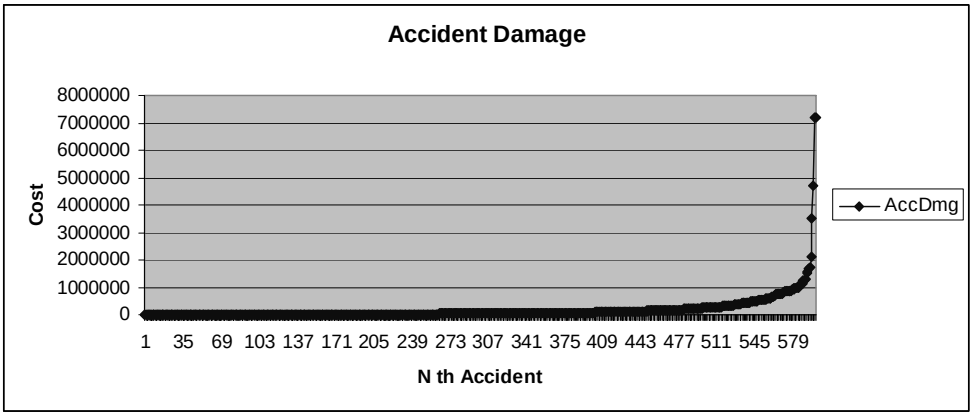


Figure 3: nth accidents arranged according to cost in dollars

To reduce the effects of outlier data, two options are considered. The first option is to remove all outlier data from the data set completely. The second is to transform the Accident Damage variable in such a way to lessen the effect of extreme damage costs. The first option is unacceptable because it will alter the original data set, and in doing so; produce a model that does not accurately portray all the data. Therefore, the second option, transforming the data, is selected.

The Accident damage variable is transformed using the natural log function. Using the natural log of the societal cost lessens the effects of extreme accident damage costs on the prediction model. This effect can be seen in figure 4, which displays the natural log of the Accident damage of the Nth damage (where the Nth damages are arranged from lowest to highest).

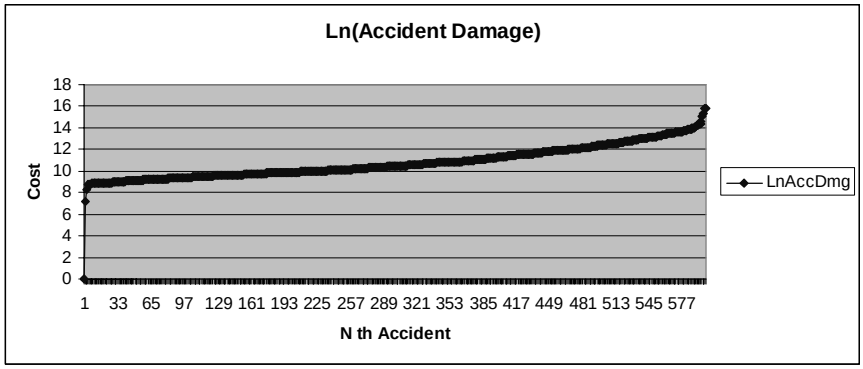


Figure 4: Ln adjustment of figure 3.

As one can see, the natural log of accident damage causes a linear effect in ordered accident damage. This linearity reduces the effect of outlier data while still allowing the data set to contain these extreme data. The benefit of the logarithmic transformation can be seen in the final severity, which uses the natural log of the accident damage cost as its target variable to be predicted.

Results

After using a logarithmic transformation to adjust the societal cost the regression tree depicted in figure 5 was produced. The tree has nine terminal nodes and seven splitting nodes with five different splitting variables. The tree uses the splitting rule Hspeed <= 10.5 Mph to split the data set practically in half. In addition to Speed (HSPEED), accident type, region, tons and visibility was also used to further partition the data. The tree classifies each accident by categorizing the accident's attribute data according to the data condition for each node. If the accident data meets the data condition of the node it goes down the tree a node to the right, otherwise, it goes down a node to the left. This continues at each node until the accident is categorized into a terminal node. To Get a better idea of how the tree assigns accident cost to a given accident we will look at a specific accident from the data. For this example we will choose accident 070696101of the data set. This accident has the following attributes: Hspeed = 20 mph, accident type = 'side collision', FRA region ='6', tons = '7304', and visibility = 'day'. Using the tree in figure 5 this accident gets assigned to terminal node 3 with a normal cost distribution of (mean =10.446 STD =1.233) ln dollars. The actual cost of this accident was 10.463 ln dollars, thus falling relatively in the statistical middle of this cost distribution.

Table 2 displays the error count of the tree and the initial mean and variance of the data. The overall total mean square error of the tree is the Resubstitution relative error * the initial variance or .535 * 2.119 = 1.133665 log dollars.

Table 2: Severity Regression Tree Summary

Model			Data	
Terminal nodes	Cross-validated Relative Error	Resubstitution Relative Error	Initial mean	Initial variance
9	0.669 +/- 0.043	0.535	10.806	2.119

To get a better idea of how successful this model performed it is better to examine the variance at the terminal node. This can be seen in table 3 and illustrated in figure 6. The regression tree presented produces nine terminal nodes with standard deviation varying from .198 to 1.453 log dollars. The standard deviation at each terminal node is relatively low and average to be about 9.3% of the mean at each terminal node.

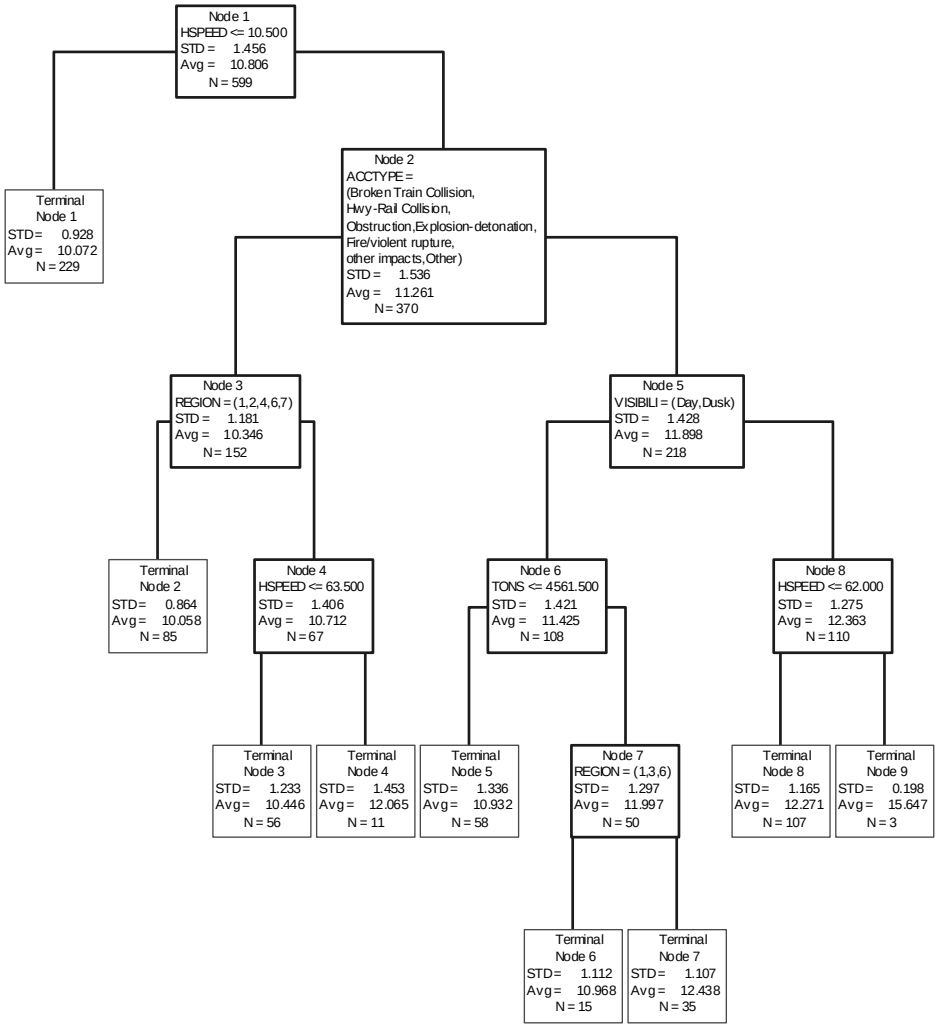


Figure5: Severity Regression Tree

Node	Count	Mean	Standard Deviation	Parent Complexity
1	229	10.072	0.928	207.713
2	85	10.058	0.864	20.067
3	56	10.446	1.233	24.09
4	11	12.065	1.453	24.09
5	58	10.932	1.336	30.48
6	15	10.968	1.112	22.692
7	35	12.438	1.107	22.692
8	107	12.271	1.165	33.258
9	3	15.647	0.198	33.258

Table 3: Regression Tree Node Summary

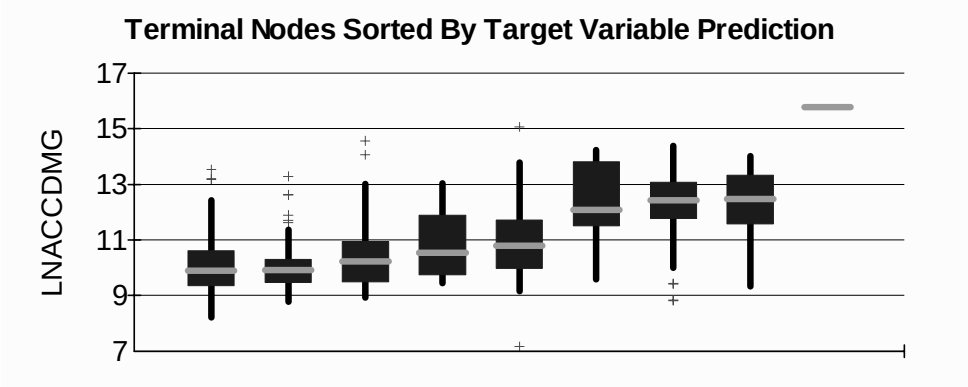


Figure 6: Terminal node box plot

Conclusions and Future Work

The ASCAP methodology is a simulation-based approach that replicates the actual behavior of a transportation system from a vehicle-centric perspective. The risk of a railway transportation system is measured as the product of severity and the likelihood of occurrence. The ASCAP sets forth the statistical basis for determining likelihood of occurrence a particular accident and this paper sets forth a methodology for predicting accident severity with in the ASCAP simulation. The methodology calculates severity using regression tress. The benefit of using the regression tree methodology over other methods is that regression trees produce results that are easily understandable to most people. Having such an easy to understand decision tree model is invaluable for predicting accident severity because it allows expert opinion to logically adjust and reshape the model according to their domain knowledge.

The data used in this paper is thought to be complete and accurate by the FRA. This data set however is just a small subset of the available FRA accident data. The remaining data for the most part is thought to be incomplete and inaccurate by various senior system safety members of the FRA. This poses a problem when we wish to consider using this modeling technique on greater sets of data. We hoped to address this problem with the inclusion of domain experts to the model development process. Allowing domain experts to see how the data can be used model accident cost gives them an understanding of how important accurate accident data are. This understanding can then, in turn, prompt them to instantiate better standards for acquiring and maintaining accurate accident data.

In conclusion, although this paper sets forth a particular rule set for finding the severity of an accident, it is important to note the methodology for calculating severity is ever evolving. Before a final rule set can be embedded in the ASCAP simulation the tree must be tested on a much larger data set than the one used for this paper. Using more data will allow for greater model test and training data sets. Also, more data would allow for separate trees to be constructed for predicting severity of certain accident scenarios. This would allow for greater specialization and possibly improve prediction results for applying the methodology to certain accident types.

References

Breiman, L., J. Friedman, R. Olshen and C.J. Stone (1984). *Classification and Regression Trees.* Wadsworth, Belmont, CA.

Brown, D.E. and C.L. Pitard (1989). "A Parallel Approach for Improved Tree Structured Classifier Construction," *Proc. Annual Cpnf. Internat. Association of Knowledge Engineers,* College Park, MD, June 1989.

Kaufman, L.M. and T.C. Giras (2000), "The Axiomatic Safety-Critical Assessment Process (ASCAP) Simulation Methodology," Proceedings of the 9th IFAC Symposium on Control in Transportation Systems 2000, Volume 2, June 2000, pp. 534-539.

Kaufman, L. M., T Giras (2001). "Simulation of Rare Events in Transportation Systems." *Proceedings of the 2001 Winter Simulation Conference*. 2001

Rovnyak, S., S. Kretsinger, J Thorp and D.E. Brown (1994). "Decision Trees for Real-time Stability Prediction. *IEEE Trans. Syst. Man Cybernet.* 21,, 660-674

United States Department of Transportation Federal Railroad Administration (October, 1998a). *Case Studies in Collision Safety,* **Volume I.** Report Number DOT-VNTSC-FRA-99.

United States Department of Transportation Federal Railroad Administration (October, 1998b). *Case Studies in Collision Safety,* **Volume I.** Report Number DOT-VNTSC-FRA-99.

Learning from Incidents Involving Electrical/Electronic/ Programmable Electronic Safety-Related Systems

Mark Bowell (1), George Cleland (2), Luke Emmet (2),

(1) Technology Division, Health and Safety Executive
(2) Adelard, UK.

Introduction

As technology develops, economic activity varies and cultural attitudes change, the factors influencing accident situations in industry also change. One area of substantial technological development has been the way in which the massive increase in computational power has allowed sweeping changes in the control of safety-related systems applied to plant and equipment. The UK Health and Safety Executive (HSE) needs to stay abreast of these changes and of their influence on accident situations in order to provide industry with best advice on how to achieve safe working environments. As part of this process, HSE has initiated a programme of work that will eventually provide:

- guidance for those responsible on how to learn from their own incident data;
- a means for HSE to ensure that it has the best information attainable on incidents involving electrical/electronic/programmable electronic (E/E/PE) safety-related systems;
- a stimulus to industry through a successor to "Out of Control" [1].

The Electrical and Control Systems Unit within HSE's Technology Division strongly contributed to the international standard IEC 61508 "Functional safety of electrical/ electronic/programmable electronic safety-related systems" [2]. This sets out specific requirements for systems involving computer control, within a high level framework that defines the safety lifecycle and safety management activities that should be followed.

One of these requirements is the need to learn from experience. Subclause 6.2.1 of IEC 61508-1 states that responsible organisations or individuals should consider specifying, implementing and monitoring the progress of:

> "procedures which ensure that hazardous incidents (or incidents with potential to create hazards) are analysed, and that recommendations are made to minimise the probability of a repeat occurrence."

The above requirement presents a goal to be achieved and, as is often the case with goal based objectives, does not say how this should be done. The implementation details will depend on the organisation that is trying to learn, its maturity in terms of data collection and analysis, and the criticality of the systems that it is responsible for. But the terminology, concepts and approach of IEC 61508 will provide grounding for HSE's programme.

In addition to the requirements of standards, organisations are increasingly realising the importance of their knowledge assets. They are keen to share throughout the company knowledge that is currently tied to individuals. The recent interest in concepts such as organisational learning, corporate memory, and knowledge management reflect this concern. Those in the safety-critical arena have an even greater need to adopt these approaches, as expertise is in short supply and is often tied to key individuals or high-performance teams.

The HSE guidance document "Out of Control" [1] analysed the causes of 34 incidents involving control systems, and presented the results as a pie chart showing the relative occurrence of primary causes by lifecycle phase. The intention is that its successor will contain results for a greater number of incidents, work within the framework of IEC 61508, and demonstrate the usefulness of incident information.

As a first step, HSE has contracted Adelard, who are also involving the Glasgow (University) Accident Analysis Group (GAAG) [3] and Blacksafe Consulting, to carry out a 7-month interactive project that will:

- identify and evaluate existing schemes for classifying causes from incident data and generating lessons to avoid recurrence of similar incidents;
- consult users of existing schemes and potential users of HSE guidance, to capture requirements for any new method and differences in approach between industrial sectors;
- select and modify an existing scheme or schemes, or derive a new one, in order to create a method for analysing and classifying incident data to match the principles and activities of IEC 61508;
- test the new method using data from a small number of real incidents; and
- identify and present the significant strengths and weaknesses of the proposed method and how it fits in with wider issues such as incident reporting, incident investigation and process improvement.

This project is part of HSE's longer-term programme to provide best advice in this field.

Technical approach

Phase 1 – Evaluation of existing schemes and stakeholder consultation

In this phase the project will evaluate existing schemes and undertake a requirements capture and analysis exercise with the main stakeholders. Although the scheme itself will address the requirements emerging from IEC 61508, we will consult on the most relevant level of support needed. For example, some industries or organisations may be looking for an overall process that they can customise, whereas others may be looking for more concrete artefacts in the form of tool support and an explicit method.

First, the project will identity and classify a selection of existing schemes from a literature review, and prior knowledge – Glasgow Accident Analysis Group has already undertaken an extensive review of these techniques. Many of these are themselves a synthesis of established approaches for particular applications. Other recent work [4]-[11] is applicable here.

In order to appropriately evaluate alternative schemes it is important to identify desirable characteristics for the new scheme. These will also be used to inform evaluation of the scheme itself. Characteristics for consideration include consistency, coverage, generality, configurability, usability, simplicity, extensibility, and understandability.

These will be developed and defined to best meet the following criteria:

- satisfaction of HSE's overall project objectives;
- effectiveness in learning all possible lessons and preventing further incidents;
- usefulness to relevant parties (taking current practice into account);
- simplicity, usability and understanding;
- ease of training;
- consistency of results;
- compatibility with IEC 61508 concepts and activities;
- flexibility when applied to data of varying detail and quality; and
- visibility of assumptions and of level of confidence in analysis results.

The characteristics above are not all mutually compatible. For example, an engineer and a usability expert might assess causes of an incident differently. Both views might be equally valid, but there may not be a consistent assessment. On the other hand coverage is improved. Other trade-offs should become clear in the analysis, and balancing decisions will be made.

Based on the results of the initial evaluation, the project will identify a number of candidate schemes that have desirable characteristics and will consult the stakeholders in these schemes. This will establish both user requirements and current best practice and experience in actually deploying incident reporting and analysis schemes.

Adelard's consultation process involves the development of briefing material followed by interview or desktop analysis. Briefing notes are used to guide discussions, but are not rigidly followed. Instead the interview process is allowed to develop freely. This usually results in a richer discussion, sometimes uncovering unexpected threads. The briefing notes are reviewed again towards the end of the meeting to ensure all planned areas have been covered.

The briefing material and subsequent interviews will address the key concerns of:

- how companies collect and use their own incident data, impediments to data collection,
- what characterises reliable incident data and how this is acted upon when available,
- the perceived applicability of data when taken from other companies, applications or sectors.

The following table is indicative of the type of coverage that will be sought.

Consultation activities will be varied, and optimised for the organisation approached and expected benefit.

Activities will include face-to-face meetings, telephone interviews, and in some cases observation of schemes in use. In addition the task will include, where appropriate, pre- or post-consultation desk reviews. Previous consultation exercises have shown that interviewees will often have more than one role and that there are diminishing returns on the amount of new information as coverage increases. The later stages often act as validation of earlier findings.

Stakeholder **Domain**	Procurers	System Suppliers	Users	Maintainers	Assessors/ licensors/ regulators	Standards/ guidance developers	Academics and consultants
Process	I	I	I	I	I	I	E
Offshore	I		I				
Machinery		I	I				
Nuclear	I	I			I	E	
Railways					E, I		
Marine	E		E				
Medical		E	E		E		E
Aviation		E			E, I		E
Defence	E					E	E

I implies interview, *E* experience capture from a review of material

Phase 2 – Develop new scheme
Phase 1 will reveal the best existing schemes to form the basis for the scheme to be developed. Phase 2 will develop a candidate new scheme and a software prototype, will identify candidate data to use in its evaluation, and will run an open consultation workshop.
In developing the new scheme, the project will specify the changes required to existing schemes to satisfy the new scheme's requirements, especially those relating to compatibility with IEC 61508. Compatibility issues include:

- interface to the main safety lifecycle phases for the E/E/PE safety-related systems,
- interface to activities in IEC 61508, e.g. looking at the phase of the product development where a fault was inserted,
- identification of criticality of systems and components and relating these to the safety integrity level (keeping in mind that the safety integrity level applies to safety functions),
- the relationship to the risk apportionment model in the standard,
- relationship to the software and hardware techniques,
- correct use of terminology from IEC 61508.

The work will also take into account sector specific derivatives such as IEC 61511. A new scheme will be developed based on these findings together and a rationale presented. The project will build an example web-based implementation of the proposed scheme that other organisations can use as a basis for their own system. This will be designed using rapid application technology, and will support enough of the scheme for the evaluation in phase 3.
During the consultation a small number of candidate organisations will be identified to assist with running of the evaluation phase. Working with them the project will identify incident data to be used on the trials, taking into account the range of criticalities and the different types of user likely to be involved in such schemes. The scheme characteristics developed in Phase 1 will be key factors in identifying candidate data and organisations.
A position paper will describe the work to date, including an outline of the scheme, and its rationale. The project will run a consultation workshop to which stakeholders will be invited. Feedback from this workshop will be valuable in positioning evaluation of the scheme and refining the final version.

Phase 3 – Evaluation of the new scheme
Simple usability testing will be conducted in-house as a sanity check for the scheme before conducting the external evaluation. Testers will not have been involved in the project up to this point, but they will be familiar with similar reporting schemes. Following this, and based upon the data gathered and organisations identified in Phases 1 and 2, the scheme will be evaluated using field trials. This will involve two organisations from different domains, with at least three users from each organisation. Stakeholders with differing viewpoints will be included in the study, and the number and scale of incidents should be representative. Concentrating on high impact, large-scale events is not appropriate as these will be visible and handled specially. Instead the project will look at typical event sets and evaluate how effective the scheme is at drawing out appropriate conclusions and patterns. The results of these tests will be evaluated against the test criteria. This information will be used to produce a set of recommendations for upgrading the scheme.

Phase 4 – Consolidation of and deployment support for scheme
This phase will provide a detailed rationale for the new scheme (based on output of Phase 2), and will update the scheme according to the recommendations from Phase 3. The prototype software developed during Phase 2 will also be updated to reflect the new scheme. An appraisal of the scheme will include its strengths, weaknesses and trade-offs that have been made in its development.
The scheme developed will be generic, implementing the requirements of IEC 61508. Practical use of the scheme will require guidance on how it should be customised for specific use. The project will draft guidance in this area, and also draft limited domain guidance for at least one domain. The audience will be base-level industry users, i.e. those with low levels of current capability or sophistication in such schemes.
More sophisticated users will be able to use the scheme in more flexible and powerful ways, building it into a safety and process improvement model, including potential use of the software tools developed on this project. Draft customisation guidelines will be produced which could support such users in deploying the scheme most appropriately for their level of sophistication.
The classification and analysis scheme developed is only part of the whole process of incident reporting. The project will summarise concerns and strategies on the wider issues of process improvement, incident reporting and incident investigation. Most of these will necessarily have been considered early in the project. The issues include:

- how this scheme will integrate with existing practice and the investigation of other types of incidents;
- how to decide whether the scheme applies in the case of any particular incident;
- ensuring that learning is applied as widely as possible, for example on other similar systems or systems developed using the same processes;
- fragmentation and cohesion of industry sectors;
- change in style of standards from prescriptive to goal based;
- maintenance of competencies in older technologies with maturing workforces;
- professional trajectories - dilution of culture; and
- organisational and individual resistance to reporting.

Way forward
HSE will use the results of this project to draft and publish guidelines on how companies can learn from their own incidents that involve E/E/PE safety-related systems.
After these guidelines have been published, our aim is to work in conjunction with several companies to gather and make available anonymised categorised incident data and analysis results from a large number of incidents. This will help HSE to:

- publish new guidance material to supplement “Out of Control”;
- determine technical priorities for HSE inspections where applicable and appropriate;
- justify HSE's policy, technical priorities and resource usage;
- develop HSE's input to standards and guidance;
- demonstrate that learning from incident data is both feasible and beneficial; and
- increase the awareness of industry and other organisations to common problems.

Acknowledgements
The authors would like to acknowledge input to this work from other project members, in particular Robin Bloomfield and Peter Bishop of Adelard, and Chris Johnson of Glasgow University.

References

[1] HSE, “Out of Control”. Health and Safety Executive, ISBN 0 7176 0847 6, 1995.

[2] IEC 61508, Functional safety of electrical/electronic/programmable electronic safety-related systems, International Electrotechnical Commission, 2000. See http://www.iec.ch/61508 for further details.

[3] Glasgow Accident Analysis Group (web site) http://www.dcs.gla.ac.uk/research/gaag/.

[4] J Henderson, C Whittington & K Wright. Accident investigation -The drivers, methods and outcomes, http://www.hse.gov .uk/research/crr_pdf/2001/crr01344.pdf.

[5] A D Livingston, G Jackson & K Priestley. Root causes analysis: Literature review, http://www.hse.gov .uk/research/crr_pdf/2001/crr01325.pdf.

[6] F Koornneef, “Organised Learning from Small Scale Incidents”, Delft University Press, ISBN 90-407-2092-4, 2000.

[7] Eurocontrol, “Reporting and Assessment of Safety Occurrences in ATM”, Eurocontrol Safety Regulatory Requirement, ESARR2, Edition 2, 03-11-2000.

[8] NASA, “NASA Procedures and Guidelines for Mishap Reporting”, Safety and Risk Management Division, Washington DC, USA, NASA PG 8621.1, 2001. See http://www.hq.nasa.gov/office/codeq/doctree/safeheal.htm.

[9] "Final Report of a Study to Evaluate the Feasibility and Effectiveness of a Sentinel Reporting System for Adverse Event Reporting of Medical Device use in User Facilities", Food and Drug Administration, Office of Surveillance and Biometrics, Center for Devices and Radiological Health". See http://www.fda.gov/cdrh/postsurv/medsunappendixa.html.

[10] T S Ferry, Modern accident investigation and analysis, 1988, Wylie, ISBN 047 16248.

[11] E D Rademaeker and J P Pineau (Eds). Accident databases as a management tool, Proceedings of the 15th ESReDA Seminar, Antwerp, 16-17 November 1998, ESReDA.

The Role of Natural Language in Accident Investigation and Reporting Guidelines

Kimberly S. Hanks (1), John C. Knight (1) and C. Michael Holloway (2),

(1) Dept. of Computer Science, University of Virginia
151 Engineer's Way, Charlottesville, VA 22904-4740, USA
{ksh4q | knight}@cs.virginia.edu

(2) NASA Langley Research Center
MS 130 / 100 NASA Road, Hampton, VA 23681-2199, USA
c.m.holloway@larc.nasa.gov

Abstract: The need to learn from incidents and accidents resulting from software failure to improve the development process and reduce the incidence of such events mandates a rigorous discipline of forensic software engineering. The proliferation of assumption in the notions and representations of critical concepts during a software process is a barrier to developing this discipline. This is true not only of documents such as requirements statements and investigation reports, but also of the guidelines that dictate how investigation of failures should take place. The goal of investigation guidelines is the production of a report with certain properties, and proliferation of assumptions in the statement of such guidelines impairs the attainment of this goal. Drawing on linguistics and cognitive psychology, we earlier motivated an approach to improving the natural language of requirements statements. In this paper, we examine the issues surrounding the natural language in which investigation and reporting guidelines are written, and suggest ways that they can be demonstrably and systematically improved with our approach. The issues and approach are demonstrated using the NASA Procedures and Guidelines for Mishap Reporting, Investigating, and Recordkeeping. Deficiencies are explored, potential consequences discussed, and a strategy for systematic improvement of the document is outlined.

Keywords: natural language, investigation guidelines, forensic software engineering.

Introduction

Incidents and accidents that can be attributed to software failure often result in tragedies and other losses. The need to learn from these events grows more critical as software systems become more complex and the ways they can fail become less intuitive. This need mandates the development of a more rigorous and systemic approach to forensic software engineering.

In exploring the issues surrounding the development of a systemic approach to forensic software engineering, Johnson, 2000 recognized that the proliferation of assumption and other communicative problems throughout the software process were recurring themes in investigations of accidents. Importantly, he also demonstrated that these same deficiencies that plagued the process also plagued the reports and recommendations resulting from these investigations. These deficiencies raise two issues. First, the potential for assumption and other miscommunication during the development process, especially in the early stages of a software project, can allow invalid conceptions of elements of the system to enter and persist, possibly leading to failures. Hayhurst and Holloway, 2001, among others, argued that requirements is a communication problem, and thus that poor requirements result from poor communication, and Lutz, 1993 showed that the majority of safety-critical errors in the systems she examined were introduced at the requirements stage. Further, unstated or unclear motivations for requirements decisions impair the ability to analyze causes. Second, the potential for assumption and other miscommunication in the reports and recommendations resulting from investigations of such failures can render such documents of little use. For example, if the analyses contained in a report are based on misconception, then a valid analysis has escaped recognition, and if the recommendations suggest ideals that are assumed to be achievable but are in reality impossible (as documented in Johnson, 2000), then time and energy that could be applied to exploring new avenues for progress is likely to be wasted in the service of unattainable perfection.

To these we add a third issue: there exists the same potential for assumption and miscommunication in the statements of guidelines that prescribe the activities and artifacts associated with incident and accident investigation and reporting. In contrast with the requirements for a software system or the reports resulting from investigations, guidelines represent meta-statements; they define the form and content of a class of instances, whereas requirements

and reports are instances of classes. The purpose of these meta-statements is in large part to standardize the results of investigations, such that, as a data set, the results can be compared with one another and analyzed for trends. In other words, the intended value of guidelines is that they predictably generate artifacts with properties that are useful to forensic software engineering. The potential for assumption and miscommunication in such guidelines impairs the likelihood that, for example, two different investigation teams will come to substantially the same conclusions while following the prescribed process, that is, this potential impairs predictability and the value of resulting documents as a data set. This creates an additional area of focus within a systemic view of forensic software engineering, in which reduction of the potential for assumption and miscommunication in investigation and reporting guidelines is a necessary task if the big picture and the role of communication throughout it are to be improved.

In previous work, we examined how the ways that humans innately use natural language render statements of requirements incomplete, inconsistent, and open to misinterpretation (Hanks, Knight, and Strunk, 2001). This analysis exploited results from cognitive linguistics that detail the ways in which humans organize and communicate conceptual information. We extended this model to account for the breakdown that occurs in communication of information across boundaries of domain expertise, breakdown that is implicated as a major limiting factor of the quality of large and complex software systems (Curtis, Krasner and Iscoe, 1988).
Miscommunicated requirements, as noted above, are themselves a detriment to forensic software engineering (Johnson, 2000). However, the model by which we analyze and characterize communicative breakdown in requirements can also be applied to investigation and reporting guidelines, as well as to the reports and other documents that are generated. Implications of the model suggest ways to improve the use of natural language in all of these areas. In this paper, we treat particularly the problem of incomplete, inconsistent, and ambiguous guidelines for the investigation and reporting of incidents and accidents. We begin with a more detailed discussion of the issues particular to guidelines and their communication. Next, we review the analysis model, which is followed by an overview of the approach that the model motivates. We then provide a case study of the language used in the NASA Procedures and Guidelines for Mishap Reporting, Investigating, and Recordkeeping. Finally, we describe a plan for improvement of the document through systematic reduction of incompleteness, inconsistency, and ambiguity.

The Role of Natural Language in Investigation and Reporting Guidelines
It is important that the investigation of any incident or accident be effective and efficient. A variety of techniques and procedures has been developed to assist with these goals, and many organizations have developed guidelines for investigation and reporting. An important objective of guidelines is to facilitate the creation of results that have *predictable* properties, in particular, ones that facilitate comparison among results of multiple investigations in order to observe patterns. Guidelines help to ensure that results are comparable by prescribing processes, procedures, and formats. Figure 1 illustrates the relationship of guidelines to a number of activities surrounding and directed by them. The structure that emerges when one considers the role of guidelines is that they are meta-documents—they are used to instantiate particular investigations and reports. Any deficiency in the guidelines, even one that seems unimportant, could have an extensive negative effect if it leads to significant imperfections in many investigative or reporting instances.
Deficiencies in guidelines do not need to take the form of factual errors to have a substantial effect. Ambiguous statements in guidelines can be extremely serious because the multiple meanings lead to results that differ from one investigation to another, thereby precluding the goal of predictability. Further, statements that are incomplete affect predictability because the incompleteness leads to instantiations that are either themselves incomplete or completed in an ad hoc manner. Finally, inconsistency in guidelines can result in instantiations that differ because of different interpretations arising from the inconsistency during instantiation.
As we show in the next section, natural language and complex cognitive structures serve our everyday needs as humans in a way that is not consistent with the goals of precise guideline statements. Unless this issue is addressed, the many opportunities for misunderstanding that are inherent in our unrestricted use of natural language can have disastrous effects in situations where completeness, consistency and precision are essential.

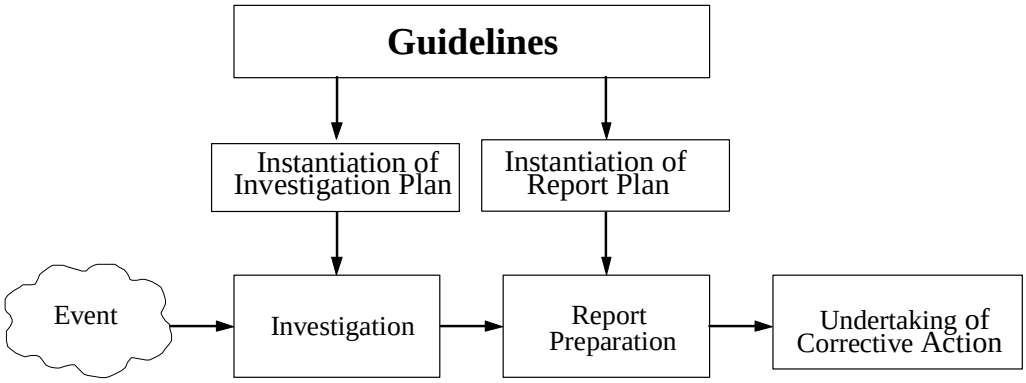


Figure 1: The role of guidelines.

The Category Model and Its Implications

Research in linguistics and cognitive psychology has demonstrated that the universe of semantics understood by any person is organized into a collection of structured entities called *cognitive categories* that possess various properties (Rosch and Lloyd, 1978, Mervis and Rosch, 1981, Ungerer and Schmid, 1996, and Langacker, 1990). For our purposes, these cognitive categories can be defined as follows:

> *Cognitive categories are collections of mental representations of entities encountered or imagined by an individual that are judged by that individual to be sufficiently similar to each other to count in some partitioning of reality as being the same.*

An individual's categories are formed as a result of his or her accumulated experience. Since there are many possible partitionings of reality that are useful to us in our interaction with the world, any entity can be a member of more than one category—which category depends on the factors considered to be significant for the task or experience at hand.

Categories are collections with internal structure based on a notion of resemblance or similarity to a prototype characterizing the category (Figure 2).

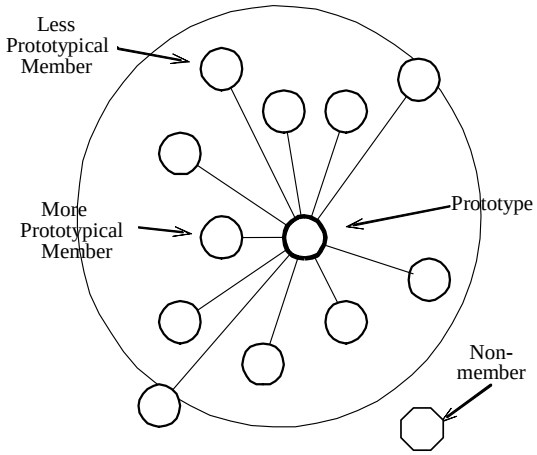


Figure 2: Structure of a cognitive category.

Members of a category that are closely clustered around the prototype bear stronger resemblances to it and instances further away have less resemblance. For example, a helicopter is a member of the *aircraft* category but is less

aircraft-like than, for example, a Boeing 747. However, by being a member of a given category, regardless of how prototypical, an instance is associated with a collection of attributes common to members of that category. By communicating a single term, such as helicopter, a speaker conveys many attributes to a listener. The listener does not need additional communication to know that the entity to which reference is made has a pilot, an engine, and flies. This aspect of human communication is known as *cognitive economy*.

Human communication in routine circumstances would be almost impossible without cognitive economy. The bandwidth of human communication is quite low, but because of cognitive economy a great deal can be communicated using that low bandwidth. Without it, we would be forced to describe all of the details about every entity we ever wished to mention. Clearly this would be impractical.

The efficacy of cognitive economy rests on shared experience. Using the word helicopter when speaking to someone who has no experience with helicopters is useless. Provided those engaged in communication (whether written or oral) have similar experience, all is well. If their experience of the domain at hand differs, even if the difference is slight, then the attributes associated with a term will not be the same. In routine communication this is not a major issue, but when using natural language in a context that requires extreme accuracy and precision, cognitive economy is the source of many invalid assumptions.

The issues that arise are of two types. In the first, the listener recognizes that, although a term used is familiar, some aspect of the communication is not clear. In this case, the listener can investigate the meaning of the problematic term by asking questions and thereby negotiate a closer match between his own conception and the speaker's. The second type of flaw is far more serious, and it occurs when the listener, through lack of or inattention to cues, does not recognize that there is a problem. In this case, the listener *assumes* that his understanding of a term and all the associated attributes implied by cognitive economy are the same as those of the speaker when they are not. All subsequent activities and behaviors will therefore be in the context of this misunderstanding.

In addition to having internal structure, the categories we possess are collectively organized into a hierarchy of specificity, with more general categories at higher levels and very constrained categories at lower levels. This hierarchy is one of inclusion, meaning that many low-level, highly-constrained categories are collected under the umbrella of a more general category, several of these more general categories are collected into still more general categories, and so on (Figure 3).

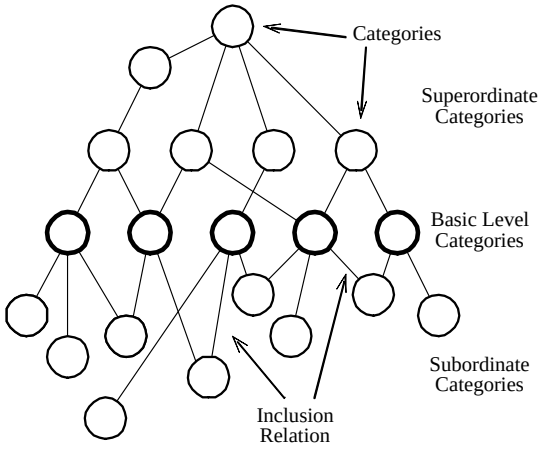


Figure 3: Hierarchical structure.

There is evidence to suggest that a particular level of this hierarchy has a special importance in our perception of the world (Rosch, Mervis and Gray, et al., 1976, Rosch and Lloyd, 1978, and Ungerer and Schmid, 1996). This level is intermediate; it is neither a very general way to describe a thing nor very specific (for example, *airplane* rather than *vehicle* or *Boeing passenger transport*), and its special role is evidenced by features particular to it in our acquisition and use of categories. It has been termed the *basic level* and represents the categories first acquired by children, the categories first evoked when classifying a newly encountered entity, and the categories used when introducing a new category into communication.

Analysis of empirical data has demonstrated a property of basic-level categories that supports their importance in communication. The basic level is that level of the hierarchy at which elements of any given category share the most

features with each other and the fewest with members of other categories (Rosch, Mervis and Gray, et. al, 1976, Rosch and Lloyd, 1978, and Ungerer and Schmid, 1996). Unfortunately, this property further complicates the reliable communication of domain knowledge. This added complication derives from the nature of the domain-specific categories used by experts. Domain experts have accumulated experience that results in their associating more attributes with domain-specific categories than a non-expert would. This is an obvious result of the very fact that an expert is an expert. The additional attributes provide more dimensions along which to collect and differentiate entities resulting in certain of these categories being basic in the expert's category hierarchy.

The implication is that experts tend to see what are commonly lower-level, more constrained categories as basic in their own hierarchies, and to use them in ways that basic-level categories are used. On being presented with a new entity in his domain, an expert is likely to associate it with a more constrained category than would a non-expert. Similarly, on using a domain-related entity in communication, the expert is likely also to invoke a more constrained category. This means that in addition to experts and non-experts possessing more and less constrained versions of certain categories, the denser expert versions are more likely to come up in discussions in a specific context because, to the expert, they are at the basic level. This results in more misalignment between the categories used by experts and non-experts than would occur because of the backfiring of cognitive economy alone.

To review, the mechanics of linguistic breakdown in the communication of domain knowledge can be characterized as follows. First, the benefits of cognitive economy that allow us to communicate adequately though somewhat imperfectly in our routine activities lead not only to the potential, but the *likelihood* that erroneous assumptions will be made in high precision, technical communications. Second, added complications arise from the specific categories that domain experts regard as basic because they are at a different (lower) level than the categories regarded as basic by the non-expert. Communication across a domain boundary, communication that is essential if investigation guidelines are to be sufficiently comprehended by investigators with diverse expertise, embodies exactly the properties that cause our natural machinery to fail. It is not a part of human nature to get this kind of communication right without serious and explicit intervention.

Approach

Using insights gained from a linguistic analysis of breakdown in domain knowledge communication, we developed an artifact designed to manage and contain the potential for such breakdown. Consider the case in which any two people with differing levels of expertise with regard to a topic are communicating regarding an entity relevant to that topic. Further assume that one or the other has in fact recognized that a breakdown is occurring. This is not representative of the more dangerous situation of no breakdown being signaled, but motivates a strategy for preventing the breakdown from occurring in the first place. In the case where breakdown is recognized as it is happening, the usual course of action taken by the interlocutors is to execute clarification activities. These activities generally take the form of paraphrasing the offending term with another term or phrase for which the sender believes the receiver is likely to possess a more compatible category topology or topologies. If this paraphrase contains terms that also invoke misaligned categories, these terms can further be paraphrased, and so on. Comprehension is recursive; we comprehend a new idea when we can put it in terms of other ideas that we already comprehend. This insight provides a direction for dealing with the problem produced by reliance on assumption in communicating domain knowledge.

Our approach is to introduce a highly structured mechanism, called the *domain map*, into communication activity that requires accuracy and precision. The map stores definitions of domain-specific terms, and documents their recursive dependence on definitions of other terms for their comprehension. It is intended to provide a systematic and complete repository of relevant domain semantics built according to the principle of making the implicit explicit. Further, once constructed, it is to serve as the exclusive point of reference for such semantics where the content of the document is concerned, providing a consistent picture to its users, for example, members of an investigation board.

Specifically, the domain map is to be constructed using a starting point of some recorded natural language, for example, a written document such as an early draft of guidelines. This body provides a corpus representing an instance of the language used to talk about the domain in question and the constraints to be placed on it. In an iterative process, experts in the various areas that contribute to the guidelines, and representatives of non-experts who are likely to be users, cooperate to partition this corpus into terms identified respectively as *domain*, those that have domain-specific meaning, and *common*, those that are unlikely to invoke relevant differing assumptions between experts and non-experts. One focus of our parallel work is refining this partitioning activity to be based rigorously on specific membership criteria for these sets. However, early experiments have been quite successful even with partitioning accomplished in an ad hoc, intuitive manner (Hanks and Knight 2002).

Once the initial *domain* set is constructed, each of its elements can be defined precisely, again in an iterative process executed by cooperating experts and non-experts. For example, a non-expert might make a first attempt, which the expert would then examine and revise. A stipulation placed on the process is that, for each term, the parties must agree that they have converged on the same understanding, as interpreted from the definition, before moving forward. This implies that both parties must have the same understanding of *each term* in the definition. Thus, terms upon which the initial term depends must themselves be classified as *domain* or *common* and defined as necessary. This realizes the recursive nature of comprehension, and forces the parties to trace these dependencies.

The bottom of the recursion is defined by design, and thus the trees representing term definitions bottom out with terms that are accepted without definition. This is the purpose of the *common* set; by virtue of its construction, it consists of those terms deemed to represent knowledge common to those both inside and outside the domain, and its use is to provide the source lexicon on which all domain definitions must eventually depend. Thus another stipulation is necessary: no cycles are allowed in the chain of dependencies associated with any term. A cycle would indicate that the recursion would never terminate, i.e., that a common understanding would not be reachable. Parties must therefore negotiate their removal, thus addressing circularities that might not otherwise have been recognized.

Our linguistic model and approach have been shown to have value in analyzing linguistic deficiencies and improving the quality of software requirements statements (Hanks and Knight 2002). Insofar as investigation and reporting guidelines can be seen as another kind of requirements statement, we extended the application of our model and approach to this area. We next discuss this extension.

Case Study: NASA Procedures and Guidelines

Linguistic Deficiencies: To illustrate the ways in which the use of natural language encourages the proliferation of assumption in investigation and reporting guidelines, we have conducted an analysis of the glossary section of the current version of the NASA Procedures and Guidelines for Mishap Reporting, Investigating, and Recordkeeping (hereafter referred to as "NPG"). We chose a subsection of the document for practical purposes of illustration; the entire document is 176 pages and a complete analysis is beyond the scope of this work. However, we found the glossary particularly compelling, since this is the section in which the document authors were *specifically tasked with* communicating explicitly the meanings of critical terms. Our analysis indicated a number of deficiencies that impair the value of the document to those attempting to realize its intended purpose.

First, a partitioning pass over the glossary text indicated incompleteness in the set of terms chosen to be defined. Since we were concerned for the moment with the 6 pages of glossary alone, this incompleteness does not yet consider problematic terms in the 170 pages of additional text; it refers rather to the necessity of access, either through prior knowledge or definition, to the meaning of terms upon which the understanding of glossary entries relies. Since this section is a glossary, we should reasonably expect that upon reading a given definition, the meaning of the defined term should be clear. If the definition includes other terms that might be problematic, we should expect to find definitions for these as well. However, *every* definition included in the original glossary contains terms that potentially have several meanings, but which are not themselves included in the glossary. For example, a term that is invoked in many of the definitions is *investigation*, indeed, it represents a concept central to the purpose of the document, yet it is not defined in the glossary. While it is true that speakers of English might collectively have a general idea of what an investigation might entail in the abstract, in this context it refers to a specific collection of activities, and a brief enumeration and description of these activities would render understanding of the terms that rely on this use of *investigation* that much more meaningful to users of the document. Further, such an enumeration would discourage users from assuming the inclusion or exclusion of activities that characterize other investigations with which they are familiar, an assumption motivated by cognitive economy, but one that might be invalid.

Second, further incompleteness was represented by occurrence in passages throughout the remainder of the document of additional terms in need of definition. We can thus say that the glossary is both locally incomplete, i.e., that the definitions included are not themselves completely grounded, as well as globally incomplete, i.e., the set of terms chosen from the main text to be represented in the glossary is not comprehensive. For example, *proximate cause* occurs in the main text, and though it may have a standard definition among those with expertise in causality analysis, the individuals tasked with using the guidelines do not all possess this expertise. Further, the authors *did* explicitly include terms like *dominant root cause* and *significant observation*, among others, indicating that they saw a need to distinguish such concepts from one another. *Proximate cause*, however, was overlooked.

Third, in addition to incompleteness, there are several forms of inconsistency in the document. An example within the glossary surrounds the definition of a NASA Mishap, which includes an enumeration of possible types, corresponding to severity estimations: "Type A Mishaps, Type B Mishaps, Type C Mishaps, Mission Failures, or Incidents". A reasonable interpretation of this list might be that the set of all NASA Mishaps can be partitioned into

five mutually exclusive subsets. However, upon reading the definition for *Mission Failure*, we find that it refers to "[a] mishap of whatever intrinsic severity" that also possesses certain other properties. This directly contradicts the understood mutually exclusive partitioning; a *Mission Failure* can apparently also constitute, for example, a *Type B Mishap*. It is not clear how to resolve the contradiction, and if a user happens to refer to only one of these definitions, he would likely not even recognize that there *is* a contradiction.

Inconsistencies relating the glossary to the remainder of the document are present as well. A number of concepts invoked in the glossary appear to be represented by different terms, or sets of terms, in different locations. For example, *NASA Mishap* has an explicit entry in the glossary, but throughout the text of definitions as well as in the text of the complete document, simply *mishap* is invoked. Since mishap has a common lay usage, a user might reasonably read it as such. A more vigilant user might suspect a domain-, i.e., NASA-specific definition, but upon looking up *mishap*, would find no glossary entry and thus also reasonably assume common usage. Only had he looked up *NASA Mishap* would he have located the presumably intended meaning, but how is he to know to look there? A similar example involves the occurrence in both the glossary text and main text of *injury/illness*, coupled with the explicit glossary entry *Lost-time Injury/Illness*. The lookup problem is the same; a user wondering what constitutes an injury or illness will not find an entry for *injury/illness* in the glossary and might assume criteria based on other experience. However in this case, it is even less clear from surrounding context whether these representations do in fact refer to the same concept, that is, are there injuries/illnesses that do not cause lost time? Such inconsistencies of representation (which in the latter case may be masking incompleteness) hinder the value of the document for directing the analysis of events by making it more difficult to classify and relate objects in the world that are of interest in an investigation.

Fourth, the glossary text includes numerous terms that have abstract common meanings for which most speakers of English have similar notions. However, the abstract meanings have little value when placed in a specific context unless criteria are provided that parameterize these meanings within that context. For example, the terms *appropriate*, *authority*, *generally*, *ordinarily*, *significant*, *similarly*, *major*, *minor*, *basic*, and several others all occur one or more times within the glossary text alone. Some of them occur many times, and a number of instances as well as additional such terms were found in the remainder of the document with only a cursory search. A section addressing the composition of NASA Mishap Investigation Boards states that "[m]embers shall have sufficient experience and technical expertise" to uphold their responsibilities, but there is no indication of how *sufficient* or even *expertise* are to be qualified or quantified. These terms are all quite transparent in the abstract sense, but since they are relative descriptions or measures, they require reference points to have any useful meaning in a given environment or domain. This renders these terms in fact domain-specific once they are actually invoked in a context, and they thus require definitions and encourage assumption without them.

So far, we have concerned ourselves primarily with individual terms, however the magnitude of the problem becomes obvious when we try to deal with several terms at once. Presented here is a passage from the document addressing the intended form and content of reports that result from investigations (and recall that we are not quite sure what exactly investigations entail). The passage is followed by a selection of indications of its insufficiency for directing the construction of such a report.

> 3.7.5 The mishap investigation report will contain a description of the structured analysis technique used by the mishap investigation board or investigator for assuring all causative possibilities are explored. The mishap investigation board or investigator will document the what, when, where, and why of the mishap investigation report. The focus and priority of the investigation report is the determination and discussion of the root cause(s) of the mishap. The report will also include significant observations, findings, and recommendations. The report will include proposed corrective actions if requested in the appointment letter, and proposed lessons learned topics for future development. The report should be technically accurate, properly documented, well defined, easily understood, and consistent with the format in Appendix H or as specified by the Appointing Official.

First, neither *mishap investigation report* nor *structured analysis technique* are defined. *Structured analysis technique*, in particular, has a definition specific to software engineering, but which is almost certainly not the meaning intended here. Since an investigation board is likely to include both software engineers and non-software engineers (and these groups understand these terms differently), a board is not likely to begin with a coherent notion of what they are to describe. Further, before developing such a coherent notion, they would first have to *recognize* this inconsistency in their experience of the terms, which they might not do until much effort has been invested under faulty assumptions. An explicit definition could reduce or avert such misappropriations and inefficiencies. In addition, since any definition of *mishap investigation report* is likely to include *structured analysis technique* among

the elements such a report must describe, a valid definition for *structured analysis technique* is necessary in order to ground the definition of *mishap investigation report*.

An explicit definition of *structured analysis technique* might also encourage critical reflection on the value and limits of such techniques; note the assumption in the above passage that the use of a structured analysis technique can "assur[e] all causative possibilities are explored." While attaining this assurance might be a useful ideal for motivating and directing analysis activities, it is quite impossible to do so perfectly and demonstrably in the complex environments with which we are concerned. Johnson, 2000 recognizes such statements in the recommendations made in existing accident reports. For example, a report on the deficiencies of the London Ambulance Computer Aided Dispatch system contains the recommendation: "A critical system such as this...must have totally reliable software." Johnson states "It is impossible by any objective measures to achieve total software reliability, contrary to what is suggested..., [and] to suggest that this is possible is to completely misrepresent the state of the art in safety-critical software engineering." We must be wary of similar such assumptions and suggestions in the guidelines we provide to investigators. It is counterproductive to assign exercises in futility and to forego a number of forms of progress in the quest for an unattainable perfection.

In addition to the incompleteness represented by the unavailability of certain definitions, an instance of the "mishap" inconsistency described earlier is also found here. Further, though definitions for *root cause*, *corrective actions*, and *lessons learned* are provided, a prescription for the form and extent of the required description of these is not. This demonstrates further incompleteness.

Also represented are additional instances of abstract common terms in need of contextual parameterization. For example, what constitutes *technical accuracy* in this domain? Similarly for *proper documentation*, *well-definedness*, and *easily understood*. *Easily understood*, in particular, begs the question of audience, i.e., the diversity of readers of these reports and their expertise.

Finally, it is not clear whether Appendix H or the Appointing Official is the final arbiter of format; if both are consulted and they disagree, which is to be attended?

With this much potential for misunderstanding and therefore unpredictability of the result contained within a single paragraph of the NPG, it is clear that there is a linguistic problem to be addressed. Next, we outline a strategy for improving the document by systematically reducing the amount of incompleteness, inconsistency, and ambiguity contained therein.

Strategy for Improving the NPG: In this paper, we have claimed that the NPG in its current form has deficiencies. The deficiencies have a basis in the way that humans innately use natural language, and derive from the fact that our cognitive heuristics are optimized for situations in which communicators share experience. Communication across a domain boundary is the pathological case that breaks these heuristics; they work in the common case by exploiting assumption, but they are the source of pervasive and often dangerous miscommunication in cases where shared experience is lacking.

We believe the NPG can be improved through the application of methods originally developed for raising the quality of requirements statements. In particular, a domain map, such as was described earlier, can systematically reduce the amount of incompleteness and inconsistency present in the NPG. The following activities, intended to be undertaken by cooperating domain experts (in this case authors or those capable of authoring the document) and analysts tasked with implementing the improvement project, represent a strategy for this systematic improvement.

We would begin with the existing glossary and its local incompleteness and inconsistency. As recognized above, not only are there many terms from the main text not explicitly defined, but the definitions that *are* provided are not themselves completely grounded. We would first complete the glossary in this down-dependency direction, that is, add to the glossary those problematic terms that are present within the definitions of already-defined terms, for example, *investigation*. It is further necessary in this step to examine added definitions for their own problematic terms. The point is to produce a domain map for the glossary that is as close as possible to being internally complete.

To address, next, the inconsistencies in the glossary, requires that all uses of defined terms be checked for usage consistent with the provided definitions. For example, the conflict in the definitions of *Mission Failure* and *NASA Mishap* above must be resolved. In addition, the cases of multiple representations of single concepts must be addressed; if *mishap* and *NASA Mishap* refer to the same concept, their representation is to be standardized, likewise for *injury/illness* and *Lost-time Injury/Illness*. These changes allow a glossary that is much closer to being internally, or locally consistent. Increasing local completeness and consistency approaches the goal of making all entries transparent to any user likely to require use of the glossary. These local steps alone improve on the original by systematically addressing the terms that the authors themselves, using even intuitive methods, believed required definitions.

Once local completeness and consistency have been addressed, a more extensive project is to address global incompleteness and inconsistency, that is, the necessity for the intended meanings of any terms occurring in the main text to be transparent to any user likely to be reading the document. Given the size of the document, this effort is likely to require a non-trivial investment of effort, but since the document has a lifetime of approximately five years (NASA QS/Safety and Risk Management Division, 2000), and further, since much of its content persists through version updates, this investment can be amortized. Further, the return is greater confidence in the value of the document to effectively direct, and standardize the artifacts resulting from, investigation and reporting of incidents and accidents.

To address global completeness, we would partition the entire text into two sets: (1) those terms that have domain-specific meaning not likely to be transparent to all users and thus requiring explicit definition; and (2) those terms for which the common lay usage is what is to be understood. All unique terms in the first set are added to the glossary as independent entries. Definitions are constructed for these terms, and those definitions are then processed as before to maintain local completeness and consistency. To address global consistency, all uses in the main text of defined terms must be checked against the definitions provided. In addition, as before, multiple representations must be standardized.

Much of the process just described can be simplified by the use of support tools. In parallel to formulating theory and conducting analysis, development of such tools is underway at the University of Virginia.

The recognition of note is that this process, undertaken manually or otherwise, is systematic. Every term is processed, the form of processing is motivated by linguistic insight regarding the ubiquity of assumption, and the team tasked with improving the document is directed by the process to give specific attention to all terms that might cause problems through the potential for invalid assumptions about their meanings. Specific constraints direct decisions during processing, for example, that there be no cycles in the dependencies among terms; this forces not just the constructing of necessary definitions, but evaluation of their usefulness. The result is a more comprehensive and considered representation of meaning essential to the effective undertaking of an investigative task.

As with all Agency directives, guidebooks, and handbooks, NASA has procedures in place that must be followed to suggest changes to official documents. If we choose to follow through on the approach to changing the NPG just suggested, we plan to formulate our recommendations for changes in the style required by these procedures.

Summary

The proliferation of assumption in the notions and representations of critical concepts during a software process is a barrier to effective forensic software engineering. This is true not only of the content of reports generated during accident investigations, but also of the requirements and designs describing software to be built, and, as we have examined, the guidelines that dictate at a meta-level how analysis and investigation of failures should proceed. The goal of investigation guidelines is the production of a report with certain properties, and proliferation of assumption in the statement of such guidelines impairs the effective realization of this goal. We argued that the way humans innately use natural language encourages the proliferation of assumption in environments where individuals with differing experience and domains of expertise must communicate. We further argued that our cognitive heuristics are so ingrained that progress in overcoming their challenges will not be made without well-founded and structured intervention. The linguistic model we provide motivates much of the structure that this intervention must take in order to be effective. The result is a systematic approach to reducing the incompleteness and inconsistency of investigation and reporting guidelines, leaving demonstrably less room for assumption, and allowing such guidelines to better serve their purpose. The issues and approach were demonstrated using the NASA Procedures and Guidelines for Mishap Reporting, Investigating, and Recordkeeping. Deficiencies were explored, their potential consequences were discussed, and an approach to systematic improvement of the document was outlined.

References

B. Curtis, H. Krasner and N. Iscoe (1988). A field study of the software design process for large systems. Communications of the ACM (31)11:1268-1287.

K. Hanks and J. Knight (2002). An experiment in applying linguistic insight to improve requirements. University of Virginia Department of Computer Science Technical Report CS-2002-18.

K. Hanks, J. Knight and E. Strunk (2001). Erroneous requirements: a linguistic basis for their occurrence and an approach to their reduction. Proceedings of the 26th Annual IEEE NASA Software Engineering Workshop, 115-119.

K. Hayhurst and C.M. Holloway (2001). Challenges in software aspects of aerospace systems. Proceedings of the 26th Annual IEEE NASA Software Engineering Workshop, 7-13.

C. Johnson (2000). Forensic software engineering. Proceedings of 19th International Conference SAFECOMP 2000, 420-430.
R. Langacker (1990). Concept, Image, and Symbol: The Cognitive Basis of Grammar. Mouton de Gruyter, Berlin.
R. Lutz (1993). Analyzing software requirements errors in safety-critical, embedded systems. Proceedings of the First IEEE International Symposium on Requirements Engineering, 126-133.
C. Mervis and E. Rosch (1981). Categorization of natural objects. Annual Review of Psychology 32:89-115.
NASA QS/Safety and Risk Management Division (2000). NASA Procedures and Guidelines for Mishap Reporting, Investigating, and Recordkeeping (NPG 8621.1).
E. Rosch and B. Lloyd, eds. (1978). Cognition and Categorization. Lawrence Erlbaum Associates, Hillsdale, NJ.
E. Rosch, C. Mervis, W. Gray, D. Johnson and P. Boyes-Braem (1976). Basic objects in natural categories. Cognitive Psychology, 8:382-439.
F. Ungerer and H. Schmid (1996). An Introduction to Cognitive Linguistics. Longman, London.

Error Classification for Safety Management: Finding the Right Approach

Steven T. Shorrock

Det Norske Veritas (DNV), Highbank House, Exchange Street, Stockport, Cheshire, SK3 0TE, UK.
steven.shorrock@dnv.com

Abstract: Human error identification systems have been criticised for failing to consider the problems of operational incident investigators and system developers. Increasingly esoteric human error modelling and classification approaches have often been met with resistance from the potential user groups that could apply them with the greatest impact. In order to improve the transfer of this technology, error classification techniques must balance a range of criteria - some more practical than have previously been considered. This paper provides an example of one such technique in air traffic management (ATM) - 'TRACEr *lite*' - that has been developed with practical aims in mind, while retaining its conceptual roots.

Keywords: incident analysis, human error identification, air traffic management, TRACEr *lite*.

The Means and Ends of Error Analysis

There is an established need in safety-critical industries to implement systems to manage the human contributions to safety. In response to this need, many well-known techniques for the identification and classification of human errors have emerged. These include SHERPA (Embrey, 1986), GEMS (Reason, 1990), CREAM (Hollnagel, 1993), as well as others integrated into Human Reliability Assessment (HRA) methodologies (e.g. HRMS and JHEDI, Kirwan, 1997; THERP, Swain and Guttman, 1983). Many techniques have been influenced heavily by Rasmussen et al.'s (1981) Skill-, Rule-, and Knowledge-based behaviour framework and Reason's (1990) classification of slips, lapses, mistakes and violations. While these techniques have been primarily associated with the nuclear and process industries, HEI has also been applied in other new sectors, such as manufacturing (Paz Barroso and Wilson, 2000), rail (Vanderhaegen, 1999, 2001), consumer products (Baber and Stanton, 1994), public technology (Baber and Stanton, 1996), and medicine (Nyssen, 2000; Taylor-Adams and Vincent, 2000).

Most methods of analysing human error have focused on particular stages of the system development lifecycle, including prospective methods (e.g. predictive human error identification, or HEI) at the design stages, and retrospective approaches (e.g. incident analysis) during operation. Furthermore, in the development stages, prototyping and real-time simulation may be used to provide evidence of safety. All approaches share a need to analyse human error, and yet accident analysis and performance prediction have been pursued as two largely separate activities, by separate communities (Hollnagel, 2000). This is exemplified in the gulf between the terms used to describe human performance issue before, during and after a design is fielded (Dekker, et al., 1997). Dekker, et al., remark that an "error of omission" may be identified in the design phase. During development, the practitioner may state that "the operator couldn't keep up with what the system was doing". And an analyst may use the phrase "lack of situation awareness" *post hoc*. While incident investigators, designers, trainers and so on are using incompatible language and methods, organisational learning opportunities and efficiencies are being lost.

Some of the predictive techniques have been used as part of the HRA process, which feeds human failure probabilities into Probabilistic Risk Assessment (PRA). Incident and accident investigation and analysis, however, rarely employ the same techniques, or even frameworks as error prediction (even though some of the quantitative data that populate human reliability databases is derived from accident and incident data). Within the same company, those responsible for performance prediction and those responsible for incident investigation and analysis may, in fact, be using entirely different approaches, "despite the obvious fact that they refer to the same reality - namely the occurrence of unexpected events leading to unwanted outcomes" (Hollnagel, 2000, pp. KN1). Safety management requires both active and reactive monitoring approaches to provide feedback on organisational performance both before and after such events (Health and Safety Executive, 1997), and harmonisation of approaches should improve efficiency and quality of research, analysis and communication.

Whither, Error Analysis?

Despite the emergence of many HEI techniques, and following considerable research and development, HEI methods are still under-utilised in industry. Lucas (2001) has reported that the UK Health and Safety Executive

(HSE) is seeing "very little evidence of use of human error prediction methods in COMAH[2] reports". And yet the HSE "expect that the part that foreseeable human failings play in initiating major accidents and the human reliability of safeguards to be understood and addressed with proportionally the same degree of rigour as for process and engineering issues" (Lucas, 2001). Although formal methods of hazard identification (HAZID) and risk and reliability assessment are being used for technical issues, human factors are relatively neglected. This is despite the fact that proportionally greater improvements in safety may often be achieved by a human focus. Cacciabue (2000) has noted that there has been a dramatic increase in the human contribution to accident development due to the very high reliability and refinement of mechanical and electronic components, and the complexity of the system and the role assigned to the human operator.

Johnson (1999) goes further to assert that human reliability approaches have had little impact upon system development in many industries, largely due to the failure of human factors research seriously to consider the problems of systems development. The problems include poor methodological support, analyst subjectivity, poor support for error prediction, focus on accidents and not incidents, individual operator/system focus, and difficulty in reaching consensus on the contextual sources of latent failures. Experience of HEI in industry suggests that additional causes include complexity, lack of demonstrable added value, and forms of denial (i.e. that there is a problem or solution). According to Johnson, until practical problems are addressed, esoteric models of cognitive and organisational failure will be of little practical benefit. Many studies (most unpublished) have shown practical benefits of such approaches applied by the HF community. But the fact remains that the transfer of this technology to the design and operation of safety-critical, interactive systems has encountered serious problems, and many non-HF specialists involved in safety management may feel 'out in the cold'.

Requirements and Trade-offs

This 'impact problem' may be partly because HF specialists and non-HF specialists emphasise different criteria when evaluating such techniques. In this paper, it is proposed that error classification techniques can be evaluated on the following criteria:

1. Comprehensiveness - the ability to discriminate and classify a comprehensive range of errors and influencing factors.
2. Consistency - the degree to which the leads consistent analyses between different users and with the same user over time.
3. Life cycle applicability - the degree to which the technique can be used throughout the formative and summative phases of system design lifecycle.
4. Predictive accuracy - the degree to which the technique accurately predicts potential errors.
5. Theoretical validity - whether the technique is based on a model of human performance, with a theoretically plausible internal structure.
6. Contextual validity - the degree to which the technique adequately captures contextual information.
7. Flexibility - whether the technique enables different levels of analysis according to the project needs, known information or expertise of the user.
8. Usefulness - whether the technique suggests, or can generate, error reduction or mitigation measures.
9. Resource efficiency (training) - the time taken to become proficient in the use of the technique.
10. Resource efficiency (usage) - the amount of time required to collect supporting information and conduct the analysis.
11. Usability - the ease of use of the technique.
12. Auditability - the degree to which the degree lends itself to auditable documentation.

These criteria (with the exception of #4) can be applied to both prospective and retrospective techniques, and are an adaptation of those proposed by Kirwan (1992). Such criteria are generally accepted, though they have been criticised for being partly based on user opinion, face validity and utilisation of the technique (Stanton and Baber, 2002). Stanton and Baber propose a quantitative appraisal approach based on signal detection theory, whereby predicted errors are classified as 'hits', 'misses', 'false alarms' and 'correct rejections'. This approach depends on prior observation of the total pool of possible error types that could occur - a possibility with simple consumer products but not so with large-scale, safety-critical, complex systems. So while this is a useful and interesting approach, it may only be useful for evaluating fairly simple products, is only applicable to predictive HEI, and would seem to focus primarily on predictive accuracy, consistency, and to a slightly lesser extent, comprehensiveness. The remaining criteria are not properly addressed by this approach.

[2] Control Of Major Accident Hazards

Some existing published techniques are biased heavily toward comprehensiveness, consistency, predictive accuracy and theoretical validity. Indeed, Baber and Stanton's study similarly valued the first three of these criteria in the appraisal approach described. When these criteria are valued to the expense of contextual validity, resource usage, usability and flexibility, for example, error classification techniques can be off-putting to those in industry, who feel that their practical needs are not being addressed. Shorrock and Kirwan (2002) note that the main problems of many techniques are:

- low usability (e.g. lack of structure, excessive requirements for supporting analyses, excessive jargon or excessive 'resolution', i.e. distinctions which were not possible to make reliably);
- low contextual validity (particularly important for Performance Shaping Factors - PSFs); and
- limited applicability (e.g. to skill- and rule-based performance only, to small-scale systems or applications only; to retrospective or predictive use only).

Methods that are used by incident investigators and designers generally emphasise practical criteria. These methods provide description and context, and are often quick and easy to use. However, the methods are often limited to a specific domain, highly dependent on experience, and fall short of providing the kind of information that is required to explain (or predict) errors. It stands to reason that if error classification techniques are to be useful, they must satisfy a more diverse range of criteria in a more balanced fashion than has previously been achieved, to meet the needs of the user group.

Technique for the Retrospective and Predictive Analysis of Cognitive Errors

One approach that has been developed within National Air Traffic Services (NATS) to gain a better understanding of air traffic controller error is called TRACEr ('Technique for the Retrospective and Predictive Analysis of Cognitive Errors') (Shorrock and Kirwan, 1999, 2002). TRACEr is a model-based approach to HEI, permitting both retrospective and predictive analysis. Shorrock and Kirwan (2002) called this 'the Janus perspective' after the Roman god who gave his name to the month of January. Janus presided over openings, beginnings and doorways, and was often depicted with two faces because he could look into the past and the future at the same time. The original version of TRACEr comprises a modular structure of eight taxonomies describing the context, error and error recovery, employing a series of colour-coded decision-flow diagrams and tables.

The process of developing TRACEr was iterative. The main inputs included:

- a literature review (covering over 70 sources);
- a controlled study of error classification;
- analysis of approximately 30 controller interviews regarding unreported human errors; and
- controller reviews of TRACEr taxonomies.

The principal applications have been:

- Analysis of UK Aircraft Proximity (Airprox) incidents (a mandatory reporting system) occurring within both controlled and unregulated (Shorrock et al., 2000) airspace between 1996 and 1999.
- Analysis of confidential incident/error reports (voluntary reporting system) from the Confidential Human Factors Incident Reporting Programme (CHIRP).
- Prediction and analysis of errors occurring in large-scale real-time simulations as part of the New Scottish Centre (NSC) programme (Shorrock, et al., 2001a).
- Prediction and analysis of errors occurring in small-scale military simulations of reduced separation standards outside controlled airspace (Shorrock et al., 2000).
- Human error prediction for the Final Approach Spacing Tool (FAST) (Evans et al., 1999; Shorrock et al., 2001b).

The original TRACEr is described more fully in Shorrock and Kirwan (2002). In a EUROCONTROL project, TRACEr was also developed for European use in the 'HERA' project - human error in ATM (see Isaac, et al., 2000, 2002a, 2002b), and, in collaboration with the Federal Aviation Administration (FAA), further developed in a joint project resulting in the 'Janus' technique.

Development of TRACEr *lite*

Initially, TRACEr was designed to be used primarily by HF specialists. However, it became clear that TRACEr could be beneficial to other ATM specialists, such as incident investigators and designers. Operational feedback revealed that TRACEr appeared too complex or time-consuming to apply in an operational environment by non-HF specialists (as with other error classification systems). If such a technique was to be used in practice, a reduced-scope version was needed. This idea was called 'TRACEr *lite*' - an error analysis and classification tool for operational ATC personnel.

Following consultation with potential users of TRACEr *lite*, some practical requirements were determined in conjunction with the intended users. These emphasised resource usage, usability, flexibility, training requirement, and contextual validity. However, it was recognised that the technique should share the core of TRACEr - the theoretical underpinnings and framework, and the ability to be comprehensive at a useful level whist maintaining acceptable consistency.

The framework of TRACEr was reviewed to identify the taxonomies of most benefit to TRACEr *lite*, since the complexity of the original TRACEr was partly a product of the number of taxonomies. This was achieved by reviewing the previous NATS projects that had utilised TRACEr, and surveying other techniques used in industry. In addition, operational ATC personnel were consulted to determine which aspects of TRACEr were of most benefit. Some of the names of the taxonomies were simplified for TRACEr *lite* to make them more accessible to incident investigators and ATC specialists. Following this review, the following TRACEr *lite* taxonomies were proposed:

- *Task error.* The task error describes the error in terms of the task that was not performed satisfactorily. This taxonomy applied to *retrospective* use only, since for predictive use the information would be contained within the task step of a task analysis.
- *External error.* The external error describes the observable manifestation of the actual or potential error, based on logical outcomes of erroneous actions, in terms of timing, sequence, selection and quality. External errors are context-free and independent of cognitive processes (e.g. intention), and are used solely as prompts for *predictive* purposes.
- *Internal error (modes and mechanisms).* Error modes describe what cognitive function failed or could fail, and in what way. Error mechanisms describe the psychological nature of the error modes; the cognitive biases that are known to affect performance. These taxonomies are used for both *retrospective and predictive* use. A small number of error mechanisms are not feasible to predict and so were omitted from the predictive version.
- *Information.* Information keywords describe the subject matter or topic of the error, and relate specifically to the error modes. For instance, what information did the controller misperceive, forget, or misjudge, or miscommunicate (e.g. 'heading', 'flight level')? This taxonomy applied to *retrospective* use only, since for predictive use the information would be contained within the task step of a task analysis.
- *Performance Shaping Factors (PSFs)* - PSFs are factors that have influenced or could influence the controller's performance, aggravating the occurrence of errors, or perhaps assisting error recovery. This taxonomy applies to both *retrospective and predictive* use.
- *Recovery* - error detection and correction is considered only for *predictive* purposes since they are more peripheral to incident investigation, and were thought to pose a risk of increasing the complexity of the retrospective method during the implementation phases.

A key process in the conversion from TRACEr to TRACEr *lite* was the simplification of TRACEr's 'internal error modes (IEMs)' and 'psychological error mechanisms (PEMs)' to create TRACEr lite's 'internal errors (modes and mechanisms)'. A card-sorting task was conducted with eight HF specialists and four air traffic controllers (one of whom was also an ATM tool designer). Each TRACEr IEM and PEM was represented on a single card, with the relevant name and a brief description. The participants of the study were asked to perform one or more of the following tasks:

- sort IEMs and PEMs into user-defined piles (i.e. no criteria imposed).
- sort IEMs and PEMs into piles of 1-3 cards on the basis of similarity, within each cognitive domain.
- sort PEMs according to their strength of affect on internal error modes.

The card sorting exercises led to a reduction of approximately 60% in the number IEMs (to an average of 4 per cognitive domain in TRACEr *lite*), and 40% in the number of PEMs (to an average of 5 per cognitive domain in TRACEr *lite*).

Whilst it was not an original aim of the project, parallel study found that TRACEr *lite* may also be capable of classifying pilot errors (Scaife et al., 2001). This study suggested a number of provisional categories that could be added to TRACEr *lite* to enable pilot errors to be addressed.

The TRACEr *lite* prototype technique was presented to operational incident investigators at Manchester Area Control Centre. This consultation produced a number of observations on the format and method of use of TRACEr *lite*, and on a small number of the categories, which were largely seen as useful. Overall, the feedback revealed that the technique appeared acceptable in terms of the incident investigators' requirements.

The TRACEr *lite* Framework

Classifying errors using TRACEr *lite* is represented as four steps for retrospective or predictive use, as shown in Figures 1 and 2. A prototype version of TRACEr *lite* has been represented using Microsoft Excel, integrating both the '*RETRO*' and '*PREDICT*' versions the same package. This contains hyperlinks for navigation and pop-up contextual examples of categories. The final technique may be implemented using a more suitable database platform, and a Web site is being constructed at www.tracer-lite.co.uk. The remainder of this paper focuses on the *retrospective* version: TRACEr *lite RETRO*.

Step R1 - Task Error: The task error taxonomies describe controller and pilot errors in terms of the task that was not performed satisfactorily. Thirteen categories are provided for controller errors, and seven categories are provided for pilot errors. The categories are shown in Table 1 below.
These categories provide a high-level and clearly comprehensible view of error, and a generic structure that will be resilient to changes in the ATC task (e.g. due to increased electronic assistance). Task errors also provide an organising structure that may be required for periodic reports of error trends.

Step R2 - Internal Error (Error Modes and Mechanisms): Error modes and mechanisms describe in more detail how the error occurred. The modes and mechanisms are structured around four error domains:

- Perception - did the controller/pilot mis-see or mishear, or fail to see or hear something?
- Memory - did the controller/pilot forget or misrecall information, or forget to do something?
- Decision making - did the controller/pilot make an error projecting required separation, or make an error of planning or decision making?
- Action - did the controller/pilot perform an action in a way not intended, or inadvertently say something that was incorrect or unclear?

Error modes provide the next level of detail on an error, describing how the controller's/pilot's performance failed to achieve the desired result. One error mode is used for each error identified in the report. For instance, error modes within the 'Perception' error domain include 'mishear', 'mis-see', 'no detection (visual)' and 'no detection (auditory)'. TRACEr *lite* contains 14 error modes - three or four for each domain.
Error mechanisms describe in greater depth the psychological underpinnings of error. Once an error mode has been selected, an associated error mechanism may be selected. TRACEr *lite (RETRO)* contains 21 error mechanisms - five or six for each domain. Example error mechanisms within the 'Perception' domain include 'expectation', 'confusion', 'discrimination failure', 'tunnel vision', 'overload' and 'distraction/preoccupation'. The TRACEr *lite RETRO* error modes and mechanisms are shown in Table 2.
Error mechanisms can better enable the consideration of measures to reduce or mitigate errors, because the internal cause of the error can be analysed. For example, if a controller misidentified an aircraft on radar, this may have been due to 'confusion' (i.e. visually similar callsigns). Such errors could lead to attempts to increase the distinctiveness of lettering, and in the meantime raise awareness of the issue with controllers. However, the error may have been due to 'overload' (i.e. a lot of traffic on radar, and insufficient time to check properly). These types of errors may lead to attempts to filter the amount of information displayed, split the sector, etc. It will often be possible to determine error mechanisms during incident investigation, but when analysing historical reports, it will not always be possible.
A screen shot of Step R2 of TRACEr *lite RETRO* is shown in Figure 3.

Figure 1 - TRACEr *lite RETRO* prototype front-end

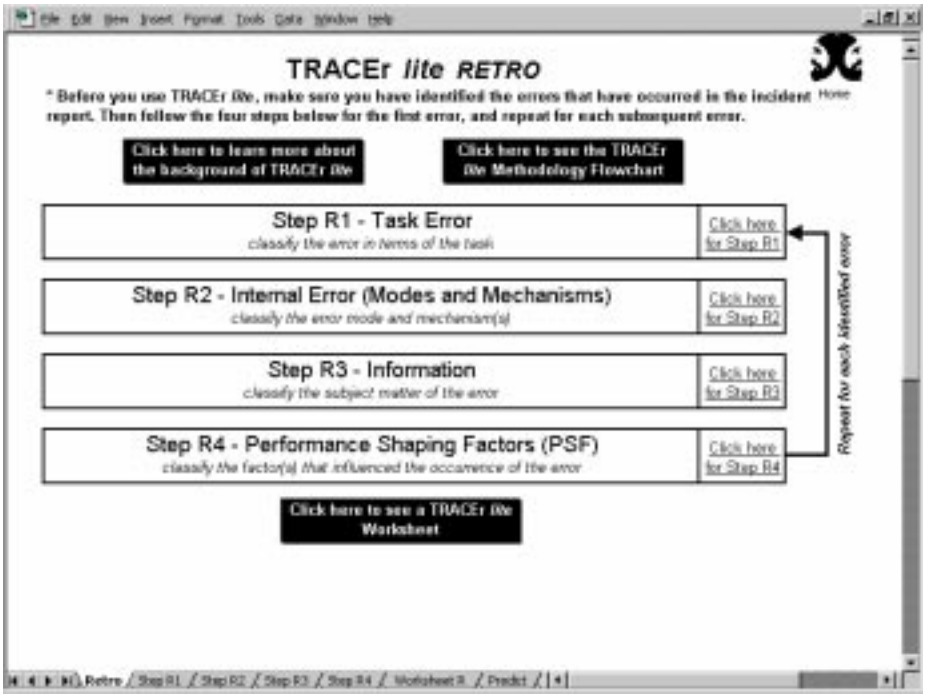


Figure 2 - TRACEr *lite PREDICT* prototype front-end

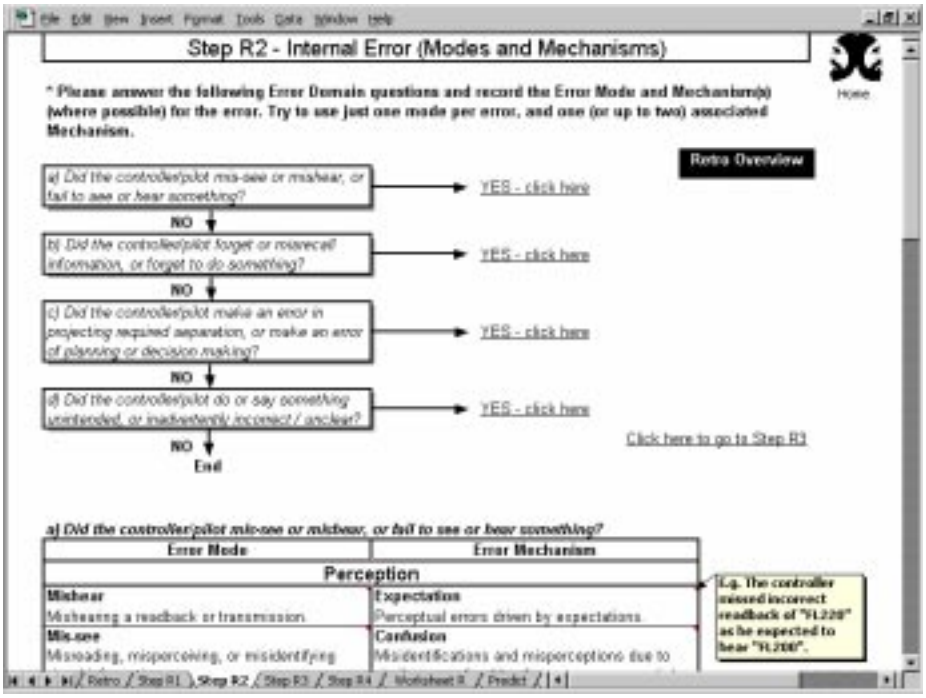
Table 1 - TRACEr *lite RETRO* Task Error list

Controller Task Errors	Pilot Task Errors
Separation error	Pilot-controller communications error
Controller-pilot communications error	Aircraft handling error
Radar monitoring error	Visual observation
Aircraft observation / recognition error	Flightdeck co-ordination/ communications error
Co-ordination error	Operational materials checking error
Control room communications error	Training, supervision, or examining error
Aircraft transfer error	HMI input & functions use error
Hand-over / Take-over error	Other pilot task error
Flight progress strip use error	
Operational materials checking error	
Training, supervision, or examining error	
HMI input & functions use error	
Other controller task error	

Table 2 - TRACEr *lite RETRO* Internal Error (Modes and Mechanisms) taxonomy

Error Mode	Error Mechanism
Perception	
Mishear	Expectation
Mis-see	Confusion
No detection (auditory)	Discrimination failure
No detection (visual)	Tunnelling
	Overload
	Distraction / Preoccupation
Memory	
Omitted or late action	Confusion
Forget information	Overload
Misrecall information	Insufficient learning
	Mental Block
	Distraction / Preoccupation
Decision Making	
Misprojection	Misinterpretation
Poor decision or poor plan	Failure to consider side- or long-term effects
Late decision or late plan	Mind set
No decision or no plan	Knowledge problem
	Decision freeze
Action	
Selection error	Variability
Unclear information	Confusion
Incorrect information	Intrusion
	Distraction / Preoccupation
	Other slip

Figure 3 - TRACEr *lite RETRO* prototype screenshot for Step R2 Internal Error (Modes and Mechanisms)



Step 3 - Information: The information taxonomy describes the subject matter or topic of the error. For instance, what information did the ATCO misperceive, forget, or misjudge, or miscommunicate? The information classification relates specifically to the error mode, e.g.: for "The controller misprojected the required heading to maintain separation", error mode = 'misprojection'; information = 'heading'.

This is an important taxonomy because it highlights specific areas for error reduction. For instance, it is little use in knowing that a large number of memory failures occur if one cannot pinpoint what information is being forgotten. The information taxonomy contains over 50 keywords. The structure of the information taxonomy, with some example keywords, is shown in Table 3.

Table 3 - Information taxonomy structure

Information (structure)
ATC/Pilot Activities and Aircraft Information
1. Controller Materials (e.g. briefing material, flight progress strip)
2. Pilot Materials (e.g. flight plan, charts)
3. Controller Activities (e.g. transfer, co-ordination)
4. Variable Aircraft Information and Pilot Activities (e.g. route, speed)
5. Other
Airspace and Other Keywords
6. Time and Location (e.g. sector, destination)
7. Airport (e.g. runway, ground vehicles)
8. Other

Step 4 - Performance Shaping Factors (PSF): Performance shaping factors (PSFs) are those factors, either internal to the controller or pilot, or relating to the task and operational environment, that affect performance (i.e. 'aggravated' an error), directly or indirectly. PSFs can often be determined following the classification of the error mode and mechanism. Approximately 60 PSFs are included in TRACEr *lite*. The structure of the information taxonomy, with some example keywords, is shown in Table 4.

Table 4 - Performance Shaping Factors taxonomy structure

Performance Shaping Factors (structure)
1. Traffic and airspace (e.g. traffic load, sector design)
2. Pilot-controller communications (e.g. controller RT standards, pilot language or accent)
3. Procedures (e.g. complexity, accuracy)
4. Training and experience (e.g. mentoring, time on sector)
5. Workplace design, HMI and equipment factors (e.g. Mode C/SSR, flight progress strip display)
6. Ambient environment (e.g. noise, lighting)
7. Personal factors (e.g. alertness/fatigue, confidence)
8. Social and Team factors (e.g. team relations, sector manning)

TRACEr *lite*'s modular structure allows the user to describe the error at a level for which there is supporting evidence. For example, if the error mechanism is not known, the user can describe the task error and error mode. When strung together, the classifications from each step form a picture of the situation, and a multi-layered view of controller error. *E.g. The controller was distracted by a phone call and forgot to transfer an aircraft to the next frequency.*

Step	*Taxonomy*	*Classification*
R1	Task error:	Aircraft transfer error
R2	Error Domain:	Memory
	Error Mode:	Omitted or late action
	Error Mechanism:	Distraction
R3	Information:	Transfer
R4	PSF:	Traffic load; Alertness/fatigue

Furthermore, TRACEr *lite* is compatible with TRACEr, such that analyses using the two techniques can be mapped and cross-referenced, and more complex cognitive errors can, if required, be initially classified using TRACEr *lite*, then revisited using TRACEr by an HF specialist and incident investigator to derive additional detail

Future Developments

Retrospectively, TRACEr *lite* is being used by incident investigators to help investigate a number of real incidents in Manchester Area Control Centre. If successful, TRACEr *lite* may be implemented operationally, using a suitable platform. NATS have developed a post-incident checklist for use by operational ATC personnel at the London Terminal Control Centre (LTCC) and London Area Control Centre (LACC) to record information based on the structure and simplified content of TRACEr *lite* to facilitate information transfer. This structure has also been adapted for use during post-incident interviews at both centres.

TRACEr *lite* has been linked with, and used in conjunction with, a safety model for ATM (Scaife, et al., 2001), to analyse future ATM technology impacts. Finally within ATM, TRACEr and TRACEr *lite* are being tested in an evaluation study in mid-2002 to test coding reliability and user opinion.

TRACEr *lite* will be adapted for potential use in the rail sector in the UK, and is open to potential adaptation for other sectors. Meanwhile, adaptations of the original TRACEr are flourishing in European ATM and potentially U.S. ATM (Isaac, et al., 2000, 2002a, 2002b).

A website is currently under construction at www.tracer-lite.co.uk, and expected to be on-line during autumn-winter 2002. In the meantime, interested readers should contact the author for a copy of TRACEr *lite*.

Conclusions

Error classification techniques must redress the balance of criteria if they are to be used by operational personnel and designers without formal training in HF. More emphasis now needs to be paid to contextual validity, flexibility, resource efficiency, and usability. Without this balance and focus on users and requirements, techniques will not have the kind of impact that one might expect, or worse, might not be used at all. Techniques also need to adopt a 'Janus perspective' (Shorrock and Kirwan, 2002), using a common framework and shared taxonomies for prospective and retrospective use, if maximum use is to be made of the feedforward and feedback loops that are available. This paper forms one attempt find the right approach in the domain of ATM. Preliminary feedback from operational users indicates that TRACEr *lite* is reaching the required balance.

Acknowledgements

Some of the work described in this paper was done whilst the author was with National Air Traffic Services (NATS). The author would like to thank Dr. Barry Kirwan and Professor John Wilson in particular for their support and guidance. The author is also indebted to the NATS Human Factors Unit, and the controllers, designers and incident investigators involved in the development and testing of TRACEr and TRACEr *lite*. The views expressed in this paper are those of the author, and not necessarily those of any affiliated organisations.

References

Baber, C. and Stanton, N.A. (1994). Task analysis for error identification: a methodology for designing error tolerant consumer products. Ergonomics, 37, 1923-1941.

Baber, C. and Stanton, N.A. (1996). Human error identification techniques applied to public technology: Predictions compared with observed use. Applied Ergonomics, 27, 119-131.

Cacciabue, P.C. (2000). Human factors impact on risk analysis of complex systems. Journal of Hazardous Materials, 71, 101-116.

Dekker, S., Fields, B. and Wright, P. (1997). Human error recontextualised. Paper presented at Workshop on Human Error and Systems Development. Glasgow University, Scotland, 19th-22nd March 1997.

Embrey, D.E. (1986). SHERPA: A systematic human error reduction and prediction approach. Paper presented at the International Topical Meeting on Advances in Human Factors in Nuclear Power Systems, Knoxville, TN, April 1986.

Evans. A, Slamen A.M. and Shorrock S.T. (1999). Use of human factors guidelines and human error identification in the design lifecycle of NATS future systems. Proceedings of the Eurocontrol/FAA Interchange Meeting, France, 27th-29th April 1999.

Health and Safety Executive. (1997). Successful Health and Safety Management. Norwich: HSE Books.

Hollnagel, E. (1993). Human Reliability Analysis: Context and Control. London: Academic Press.

Hollnagel, E. (2000). The human factors in systems reliability - is human performance predictable? Paper presented at the Human Factors and Medicine Panel (HFM) Workshop held in Sienna, Italy, 1-2 December 1999. NATO Research and Technology Organisation, Meeting Proceedings 32.

Isaac, A., Shorrock, S.T., Kirwan, B., Kennedy, R., Andersen, H. and Bove, T. (2000). Learning from the past to protect the future - the HERA approach. The 24th European Association for Aviation Psychology Conference, Crieff, UK, 4th-8th September 2000.

Isaac, A., Engelen, P. and Polman, M. (2002a). Human error in European air traffic management: from theory to practice. Workshop on the Investigation and Reporting of Incidents and Accidents, Glasgow University, Scotland, 17th-20th July 2002.

Isaac, A., Shorrock, S.T. and Kirwan, B. (2002b). Human error in European air traffic management: the HERA project. Reliability Engineering and System Safety, 75 (2), 257-272.

Johnson, C. (1999). Why human error modeling has failed to help systems development. Interacting with Computers, 11, 517-524.

Kirwan, B. (1992). Human error identification in human reliability assessment. Part 2: detailed comparison of techniques. Applied Ergonomics, 23 (6), 371-381.

Kirwan, B. (1997). The development of a nuclear chemical plant human reliability management approach: HRMS and JHEDI. Reliability Engineering and System Safety, 56, 107-133.

Lucas, D. (2001). Human error prediction and controls: demonstrations made in COMAH safety cases. Proceedings of an IBC Conference on Human Error, London, 27th-28th February 2001.

Nyssen, A.S. (2000). Analysis of human errors in Anaesthesia. Our methodological approach: From general observations to targeted studies in simulator. In: Vincent, C., de Mol, B. (Eds.) Safety in Medicine. Elsevier, Oxford.

Paz Barroso, M., Wilson, J.R. (2000). Human error and disturbance occurrence in manufacturing and a toolkit for practical analysis. Cognition, Technology and Work, 2, 51-61.

Rasmussen, J., Pedersen, O.M., Carnino, A., Griffon, M., Mancini, C. and Gagnolet, P. (1981). Classification System for Reporting Events Involving Human Malfunctions, RISO-M-2240, DK-4000, Riso National Laboratories, Roskilde, Denmark.

Reason, J.T. (1990). Human Error. Cambridge, UK: Cambridge University Press.

Scaife, R., Smith, E. and Shorrock, S.T. (2001). A practical framework for identifying human safety issues in ATM. IBC Conference on Human Error. London, UK, February 2001.

Shorrock, S.T. and Kirwan, B. (1999). TRACEr: a technique for the retrospective of cognitive errors in ATM. In D. Harris (Ed.) Engineering Psychology and Cognitive Ergonomics: Volume Three - Transportation Systems, Medical Ergonomics and Training. Aldershot, UK: Ashgate.

Shorrock, S.T. and Kirwan, B. (2002). Development and application of a human error identification tool for air traffic control. Applied Ergonomics, 33, 319-336.

Shorrock, S.T., Kirwan, B., MacKendrick, H. and Kennedy, R. (2001b). Assessing human error in air traffic management systems design: methodological issues. Le Travail Humain, 64 (3), 269-289.

Shorrock, S.T., Kirwan, B., MacKendrick, H., Scaife, R. and Foley, S (2001a). The practical application of error analysis in UK air traffic management. Paper presented at People in Control: An International Conference on Human Interfaces in Control Rooms, Cockpits and Command Centres, UMIST, Manchester, UK, 18th-21st June 2001.

Shorrock, S.T., Kirwan, B., Scaife, R. and Fearnside, P. (2000). Reduced vertical separation outside controlled airspace. Third Annual Conference on Aviation Safety Management. May 2000.

Stanton, N.A. and Baber, C. (2002). Error by design: methods for predicting device usability. Design Studies, 23 (4), 363-384 .

Swain, A.D. and Guttmann, H.E. (1983). Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications. NUREG/CR-1278, USNRC, Washington, DC 20555.

Taylor-Adams, S., Vincent, C. (2000). Clinical accident analysis: understanding the interactions between the task, individual, team and organisation. In: Vincent, C., de Mol, B. (Eds.) Safety in Medicine. Elsevier, Oxford.

Vanderhaegen, F. (1999). APRECIH: a human reliability analysis method - application to railway system. Control Engineering Practice, 7, 1395-1403.

Vanderhaegen, F. (2001). A non-probabilistic prospective and retrospective human reliability analysis method - application to railway system, Reliability Engineering & System Safety, 71 (1), 1-13.

Safety Reporting and Aviation Target Levels of Safety

G. M. Graham (1), S. Kinnersly (1), A. Joyce(2)

(1) AEA Technology Aviation, Stokes House, 401 The Qudrant, Birchwood Park, Risley, Warrington WA3 6AT, UK.

(2) Eurocontrol, Centre Expérimental, B.P. 15, F 91222, Brétigny s/Orge cedex, France.

Introduction

This paper addresses issues involved in the use of target levels of safety (TLS) in the aviation industry and their implications for investigation and reporting of incidents and accidents. The main focus here is on air traffic management (ATM). ATM is defined formally as "the aggregation of functions, comprising variously those of air traffic services, airspace management, air traffic flow management including their interacting airborne functional capabilities, required to ensure the safe and efficient movement of aircraft during all phases of flight." (EUROCONTROL, 2001). Informally, ATM can be considered as the systems and activities that ensure the safe and efficient movement of civil aircraft from departure to arrival, including any civil-military interface.

The development, quantification and allocation of TLS is currently an important area in aviation. The aviation industry is faced with the challenge of significant increases in air traffic in Europe and worldwide. At the same time, safety must be maintained and, if possible, improved. ATM is at the forefront of this as it is primarily responsible for providing adequate airspace capacity to meet demand safely. TLS and investigation and reporting of incidents and accidents are important contributors to meeting that challenge.

The aviation industry shares some key characteristics with other major industries. However, it has a number of unique or special characteristics that impact on both TLS and the investigation and reporting of incidents and accidents. These characteristics are considered here. In addition, the impact of the challenges, changes and developments facing ATM are considered.

The work on TLS on which this paper is based was carried out as part of the ASTER (Aviation Safety Targets for Effective Regulation) project carried out for the European Commission. A short summary of the ASTER project is given in the next Section.

The ASTER Project

ASTER is a recent research project carried out on behalf of the European Commission (DG-TREN) in response to Key Action 2.2 of the Fifth Framework Programme ('Competitive & Sustainable Growth'). The Research, Technology & Development (RTD) objectives for this Key Action are to promote transport sustainability from an economic, social and environmental perspective, while improving safety and security and optimising the human role and performance. The ASTER project supported the overall Research & Development (R&D) objectives developing a methodology for enabling safety targets to be set and optimised for each of the actors in the air transport system to achieve the optimum level of safety for the system as a whole. A broad range of participants were involved in the ASTER project:

a.	National Aerospace Laboratory (NLR)	Project Co-ordinator
b.	EUROCONTROL	Main contractor (Experimental Centre, Bretigny) and Reviewer (Safety Quality & Standardisation [SQS]) Unit
c.	Meridiana	Main contractor
d.	Airclaims	Main contractor
e.	Netherlands Economic Institute (NEI)	Main contractor
f.	Israel Aircraft Industries(IAI)	Main contractor
g.	Joint Research Centre (JRC)	Main contractor
h.	Federal Aviation Administration (FAA)	Reviewer
i.	Joint Aviation Authorities (JAA)	Reviewer
j.	AEA Technology	Subcontractor to EUROCONTROL

This ensured that a range of different viewpoints and interests (technical, operational and economic) were represented.

The present paper is based on the results of Work Package WP1 (Joyce, Graham, van Eenige, Kinnersly & Roelen, 2001), which was carried out by AEA Technology, EUROCONTROL and NLR. WP1 reviewed existing TLS within the civil aviation community (ATM and the air transport industry), drawing relevant comparisons with other industry sectors such as Nuclear and Rail. It specifically addressed TLS derivation methodologies, particularly in anticipation of future challenges within aviation. The role of stakeholders and the extent of their involvement in TLS derivation processes were also considered.

Characteristics of the Aviation Industry

A Safe Industry: Aviation is a very safe mode of transport. The level of safety has increased steadily since the start of passenger transport. The current situation is that the number of accidents is approximately unchanged from year to year while the number of flights and the total distance flown is increasing. There is, however, concern that the increasing amount of air traffic will result in an increase in the number of accidents per year unless the accident rate per distance flown is reduced. ATM contributes to only a small fraction of the total of aviation accidents. Within Europe, this is of the order of 8% of the total. This corresponds to substantially less than one accident per year due to ATM.

An International Industry: Aviation is intrinsically an international industry. As well as the international market in products and services common to most industries, international air transport necessarily depends on the safe, efficient and cooperative working of systems (for example, communications, navigation, surveillance and supporting software applications) and services (such as airports and airlines) from many countries. For instance, an international flight will be under the control of different national ATM systems at the beginning and end of the journey and may pass through a number of other national ATM systems while en-route. Rapid growth in international travel has emphasised the need for consistent levels of safety across international boundaries. Passengers expect consistent levels of safety, while an accident in one country affects public confidence in others. This is resulting in international harmonisation of safety levels and practices.

Large-Scale Integrated Systems: Many years ago, an aircraft could be sent on its way by its departure controller, fly largely unaided to its destination then land under a local air traffic control unit. Those days are long gone. Aircraft in busy areas such as Europe are tightly controlled from departure to arrival. Furthermore, the large number of flights means that aircraft trajectories are tightly coupled. Thus, a problem with one flight can affect many others - typically, and frustratingly for the traveller, causing many delays and with the potential to affect safety. The tight coupling means that ATM systems must be integrated in order to optimise overall safety and efficiency. This integration is occurring both vertically and horizontally. Systems that carry out different functions, such as flight planning, tactical flow management and ATC operations, are being integrated (vertical integration). Systems that carry out the same functions in different geographical areas (typically by national ATM services) are being integrated (horizontal integration). In addition, new satellite-based systems are being introduced that transcend national boundaries and introduce new functionality that crosses traditional functional boundaries (e.g. GPS applications). Safety levels and the means of monitoring, maintaining and improving safety must be appropriate for the large-scale (and increasingly larger scale) integration of systems. This raises issues such as the partitioning of high level TLS across systems and the distribution of a risk budget to provide lower level targets for existing systems.

Many Stakeholders: Aviation is a complex industry that involves contributions from, and impacts, many different people and organisations. These include technical, financial and social inputs and impacts. The people and organisations range from governments and large, multi-national companies to individuals who may benefit from travelling by air or suffer from environmental nuisance such as noise from passing aircraft.

The complexity of the stakeholder involvement is shown in the following Stakeholder Map (Figure 1) developed in the ASTER project. Each of these stakeholders has a legitimate interest in the safety of aviation and relevant safety targets. The Stakeholder Map is further complicated because each type of stakeholder is likely to be represented in a number of countries so cannot be assumed to be a homogeneous group. Each stakeholder has a different perspective on safety. TLS and the monitoring systems that support them must take the different (and sometimes competing) viewpoints and interests of the different stakeholders into account.

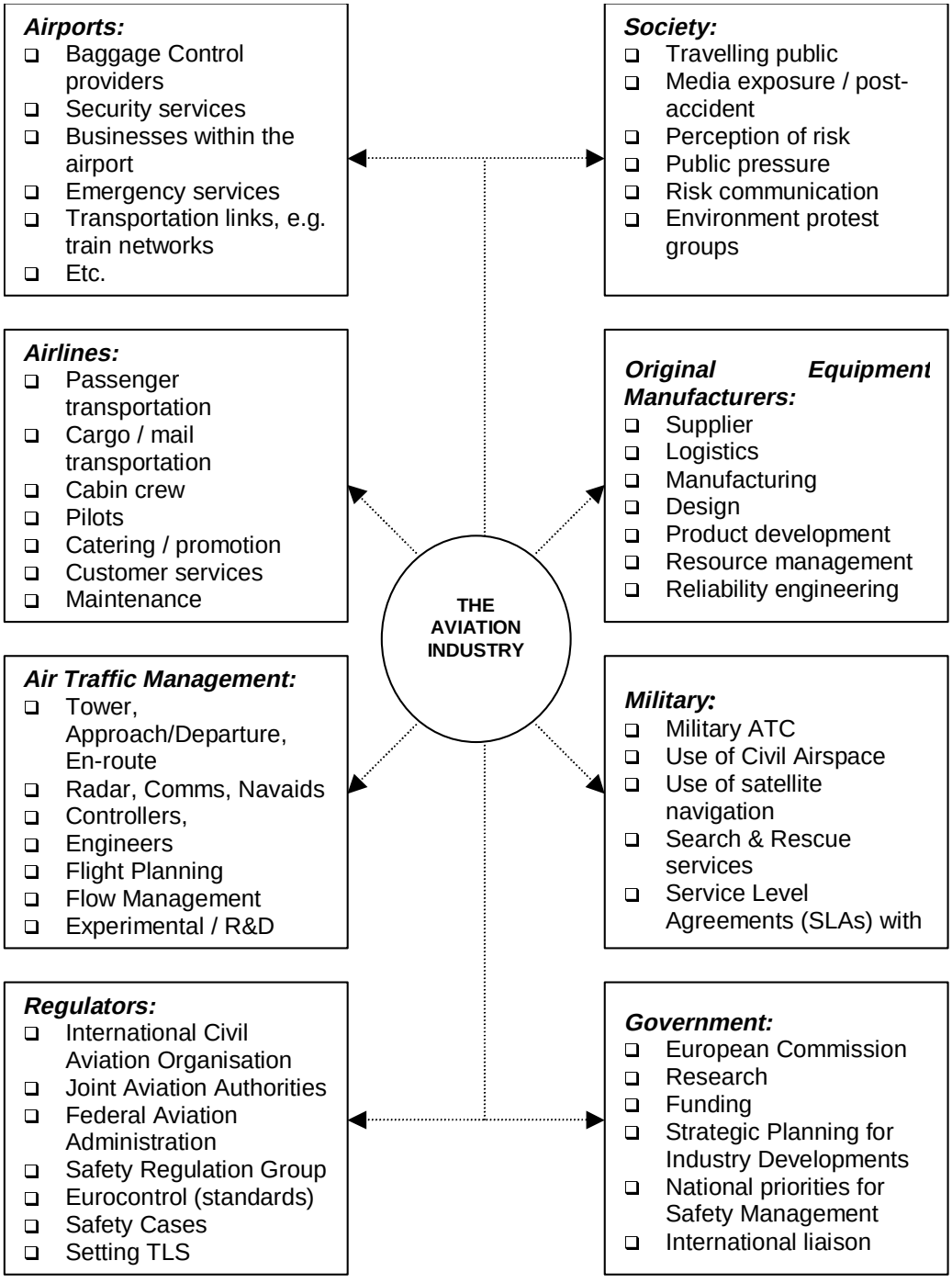


Figure 1 : Stakeholder map for the aviation industry

Technical, Human and Organisational Factors

Aviation, like most industries, depends on technical, human and organisational factors for its success. However, the contributions and balance between these factors is not the same as in many other industries. This affects the way in which safety is perceived, monitored and managed. For instance, a passenger jet aircraft can be considered similar in many respects to a hazardous plant. Safety is built into the design and maintenance. Once it is in operation (i.e. flying), it is controlled by technical systems with limited input from the plant controller (i.e. pilot). ATM, however, is very different. Technical systems provide, integrate and present data (from radar, communications etc.). However, control decisions are the responsibility of human beings (air traffic controllers) who work according to highly developed rules and procedures. Furthermore, the organisation of the airspace (into sectors and types of airspace) is a crucial organisational factor that affects both safety and efficiency (for instance, re-sectorisation can be effective in reducing delays). Aviation TLS and the reporting and investigation systems for incidents and accidents must reflect these different factors and their contributions. In addition, they must take into account the different cultural and socio-political backgrounds in different countries and organisations.

Target Levels of Safety in Aviation

What is a TLS? The concept of a TLS appears intuitively obvious. Most people have an understanding that some things are more or less safe than others. A TLS is therefore the 'amount' of safety that is aimed for. However, it is necessary to consider carefully the implications of the TLS concept in order to be able to use it clearly, unambiguously and appropriately.

A prime developer and user of the TLS concept in aviation has been the International Civil Aviation Organisation (ICAO). Among its other responsibilities, ICAO develops and sets high level requirements for safety in civil aviation. Over the years, ICAO has developed TLS for various safety critical areas, including:

- North Atlantic System (1992)
- All Weather Operations (1994)
- Obstacle Clearance (1980)
- General Concept of Separation (1995)

The latter work offered the following definition of a TLS :

> "A Target of Safety (TLS) specifies an acceptable value of risk which can be used as a yardstick against which the risks associated with a system or procedures can be evaluated. The concept of a TLS is particularly useful when planning changes in safety critical operations such as air traffic control."
> (RGCSP/WG-A/WP/8: 2)

More recently, the Safety Regulation Commission (SRC) of EUROCONTROL is developing TLS for ATM. These TLS are called 'Safety Minima'. They are defined as :

> "A level of how far safety is to be pursued in a given context, assessed with reference to an acceptable or tolerable risk." (Eurocontrol, 2000b)

Risk and TLS

TLS in aviation are, as in other industries, expressed in terms of risk i.e. the combination of harm and the likelihood of that harm. However, the specific definitions of risk are tailored to the aviation application. This results in TLS risk definitions that are significantly different to those in many other industries. For instance :

Harm from an aircraft crash is not partitioned into deaths and injuries. Survivability in an aircraft crash cannot usually be estimated. The worst case assumption is therefore usually made that a crash results in fatalities, although in practice there may be survivors and survivability is a factor in aircraft design. By comparison, fatalities in a train crash are usually limited to a specific part of a train and most, if not all, passengers are expected to survive. Thus TLS for rail accidents can be partitioned into deaths and injuries.

TLS are expressed in terms of aircraft rather than numbers of people. ATM and other aviation operations treat all aircraft equally from the point of view of safety, irrespective of the number of people carried. For instance, ATM does not allow a higher likelihood of loss of safe separation (which may result in a mid-air collision) for a small commuter aircraft compared to a Boeing 747 because it carries fewer passengers. TLS are therefore usually expressed in terms such as accidents per flight or per million flight hours rather.

Different units of risk are used for different phases of flight. A short flight (such as London to Glasgow) and a long flight (such as London to Miami) each have a single take-off and landing although the en-route distance is very

different. Take-off and landing have very different risk characteristics to en-route flight. For instance, the collision risk at landing involves an aircraft failing to land correctly. For en-route flight, it involves two aircraft inadvertently being at the same place at the same time. It is therefore reasonable to have different TLS for take-off or landing and for en-route flight and to express them differently. Typically, take-off and landing TLS are expressed in terms of 'per take-off' (or landing), en-route TLS are expressed in terms of 'per flight hour' or 'per km flown'.
The definitions of TLS risk have a direct impact on the data that must be collected for incidents and accidents in order to determine that TLS have been achieved.

TLS and Tolerable Risk
The concept of ALARP (As Low As Reasonably Practicable) is well known from UK health and safety regulation and elsewhere. It is summarised by the familiar 'carrot' diagram (Figure 2). Aviation TLS correspond to the upper end of the ALARP region – the minimum tolerable level of safety, maximum tolerable risk.

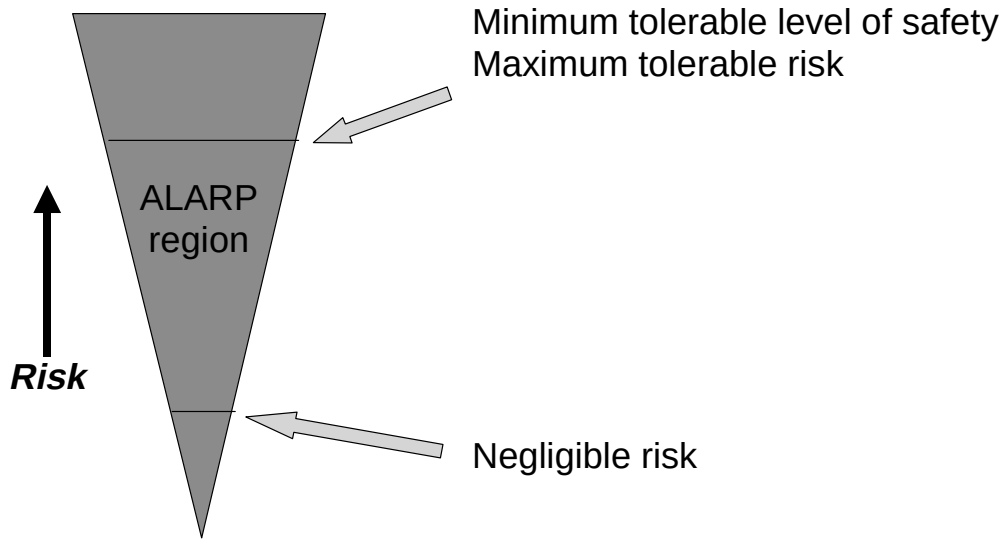


Figure 2 : Risk and ALARP

TLS in aviation typically do not incorporate an ALARP requirement directly (e.g. 'Risk as low as reasonably practicable and no greater than X'). ALARP is, however, frequently used in setting the TLS. Thus, a TLS may be set on the basis of a historic accident rate (which it is clearly reasonable and practicable to achieve) together with an improvement factor to drive future safety improvement. This, in effect, pushes the level of safety that must be achieved downwards through an (original) ALARP region.
An example of the use of historical data and an improvement factor to set a TLS is given in Appendix 1. This summarises the setting of a TLS for North Atlantic flights. More recently, the Safety Regulation Commission of EUROCONTROL has determined Safety Minima (i.e. TLS) for ATM-related accidents in Europe. Historical data were used to determine the current level of ATM safety. An improvement factor was then applied to take into account the projected growth in air traffic in Europe. The result is a TLS that corresponds to no increase in the accident rate (i.e. number of accidents per year) due to European ATM in spite of a substantial forecast increase in the number of flights. Clearly, the TLS corresponds to increased safety per flight. This is stimulating developments to ensure that the TLS is met. Put another way, what is reasonable and practicable is changing as a result of the TLS.

TLS for Incidents
A high-level TLS expressed in terms of accidents has two significant limitations for a very safe industry such as aviation. Firstly, the number of accidents is sufficiently small that statistical fluctuations can make it difficult to determine whether the TLS is being met. Secondly, number of accidents (or accident rate) is not necessarily the best measure of safety performance.

These limitations are particularly significant for ATM. For instance, the EUROCONTROL SRC has recently set the maximum tolerable probability of an accident with direct ATM contribution as 1.55×10^{-8} per flight hour (EUROCNTROL, 2000b). This corresponds to about 2.3×10^{-8} accidents per flight, which is considerably less than one accident per year in the EUROCONTROL region. Data for many years operation are needed to demonstrate that such TLS are being met. A single accident would violate the TLS for that year. Many years without an accident would satisfy the TLS but would not reveal a deterioration in safety prior to an accident. It is therefore important when setting TLS to address not just the TLS itself but also the means by which compliance may be demonstrated.

While number of accidents (or accident rate) is the top-level measure of safety, it is not the most appropriate measure for all aspects of ATM. For aircraft in flight, ATM works on the principle of maintaining safe separation distances between aircraft. Provided safe separation is maintained, aircraft cannot collide. Thus, a more direct measure of the safety of ATM for aircraft in flight is the number (or rate) of loss of separation incidents. TLS can therefore be set with respect to loss of separation. However, loss of safe separation does not mean that aircraft will collide, only that in some circumstances they might. A link between TLS for loss of separation and high level TLS expressed in terms of accidents must therefore be made by theoretical collision risk models. Similarly, TLS can be set for other incidents and related to an accident TLS. The key advantage of using TLS for incidents is that incidents happen more frequently than accidents so conformance with a relevant TLS is easier to monitor. Conformance with the top-level accident TLS, however, now depends on the validity of the model used to link incidents and accidents.

Incident and Accident Reporting and TLS

TLS and Incident and Accident Reporting Having set a TLS, it is clearly important to know whether it is being achieved. This requires the implementation of a safety management system that includes monitoring and investigation of incidents and accidents together with feedback to maintain or improve safety. The generic process is shown in Figure 3. Incident and accident reporting is a key part of monitoring safety.

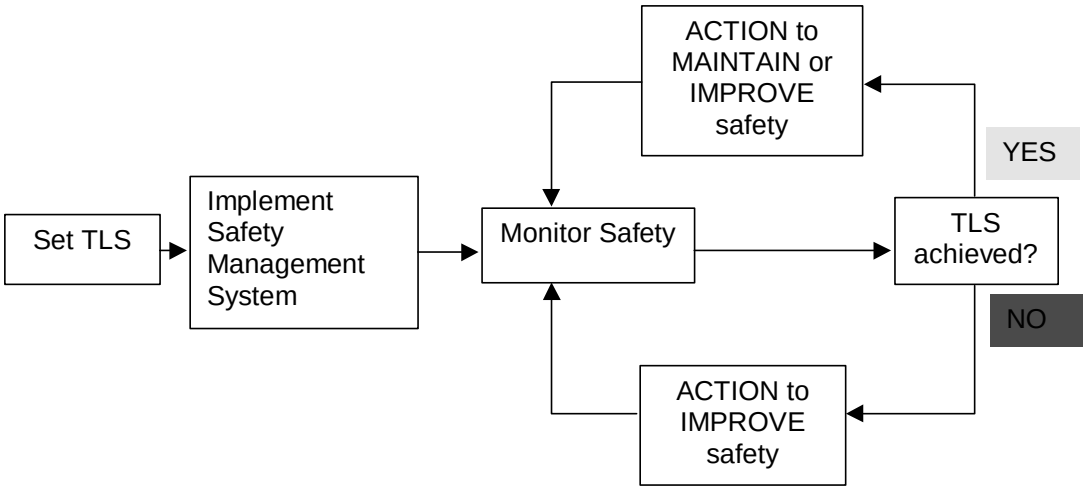


Figure 3 : Safety Management Feedback Loop

For aviation, this safety management process is an international activity which applies to all relevant organisations. Issues of consistency and completeness are therefore very important both for ensuring that overall system TLS are met and to enable meaningful comparisons between different components of the system (for instance, between different types of aircraft or between different countries).

Completeness of Reporting Credible monitoring of safety requires that reporting of incidents and accidents is as complete as reasonably possible. Incomplete reporting results in uncertainty in, or overestimation of, the level of safety achieved. It also makes it difficult or impossible to make realistic comparisons of safety and potentially dangerous trends may not be recognised sufficiently early.

Completeness depends on two factors: recognition that something has occurred that should be reported and then reporting it. When a TLS is expressed in terms of easily detected accidents (e.g. aircraft crash) or clearly defined specific incidents (e.g. loss of separation), the monitoring system can be designed to look for these specific events.

In principle, they can all be detected and reported. However, recognition that something has occurred that should be reported can be impaired if the definition of what is a reportable incident is not prescriptive. A non-prescriptive definition is nevertheless desirable from the point of view of needing to capture all relevant incidents but places a burden of judgement on the person carrying out the monitoring.
An example of both prescriptive and non-prescriptive accident and incident reporting requirements occurs in ESARR 2 'Reporting and Assessment of Safety Occurrences in ATM' issued by the EUROCONTROL Safety Regulation Commission. This gives a prescriptive list of accidents and incidents that is the minimum that must be reported and assessed (Appendix A of ESARR 2). However, there is also a non-prescriptive requirement '… for any person or organisation in the aviation industry to report any such occurrence or situation in which he or she was involved, or witnessed, and which he or she believes posed a potential threat to flight safety or compromised the ability to provide safe ATM services. Such provisions shall not be restricted to the reporting of aircraft accidents or serious incidents, since other types of occurrences could reveal the same types of hazards as accidents or serious incidents'. Judgement is required to decide whether an occurrence or situation posed a potential threat to safety. Different people may interpret a situation in different ways or have different thresholds for what is a threat. Completeness of reporting requires that suitable steps are taken (e.g. training) to ensure that all occurrences and situations that are a potential threat to safety are, in fact, recognised as such by all relevant people.
Completeness can also be compromised if the occurrence of an incident is not noticed even though it would have been reported if it had been noticed. For instance, a busy human being may not notice an incident that is a brief transient (e.g. separation only just lost and for a short period). Computer tools are therefore being developed to address this problem by the automatic detection of incidents (see section on 'Computer-based Tools' for an ATM example).

Consistency of Reporting Unambiguous reporting and investigation of incidents and accidents always requires careful definition of words and terms. This is exacerbated in the aviation industry by the need for consistency across different languages and cultures. There is therefore a strong emphasis on precise, unambiguous terminology. This has resulted in the development of detailed taxonomies for incident and accident reporting.
A recent example of a taxonomy for ATM is HEIDI. The HEIDI Taxonomy is a set of fields and definitions together with a classification scheme supporting the reporting and investigation of ATM accident, incidents and occurrences as defined into the Gate-to-ate Concept. This has been developed by EUROCONTROL as part of the harmonisation of incident and accident reporting in European ATM. HEIDI covers background terms, event types, descriptive factors, explanatory factors, classification scheme and safety recommendations. Standardisation of reporting based on HEIDI will contribute to consistency of safety monitoring. Examples from HEIDI are given in Table 2 and Table 3.

Level	Term Title	Reference	Definition	Explanations	Inputs
3	Short/medium term ATC "Planning"	HEIDI	A situation where a conflict was detected and a plan for a course of action elaborated for some reason was not/ could not be implemented..	Planned not implemented, not planned too late etc…etc… Forgotten, not physically implementatble (e.g. frequency congestion)	Planned And Not Implemented/Too Late/Physically Not Implementable

Table 2: HEIDI Example – From 'Descriptive Factor'

Level	Term Title	Reference	Definition	Explanations	Inputs
2	Accident	ICAO	An occurrence associated with the operation of an aircraft which takes place between the time any person boards the aircraft with the intention of flight until such time as such persons have disembark, in which: - a person is fatally or seriously injured (as a result of..) - the aircraft sustains damage of structural failure (which..) - the aircraft is missing or is completely inaccessible	ICAO Annex 13: - mid-air collisions between aircraft or between aircraft and other objects - collisions on the ground including Controlled Flight Into Terrain or collisions on the ground between aircraft or between aircraft and other objects	Y/N

Table 3: HEIDI Example – From 'Classification Scheme'

The HEIDI taxonomy is publicly accessible via the EUROCONTROL website [EUROCONTROL, 2001].

Learning Lessons The safety management feedback loop shown in Figure 4 requires that action is taken to maintain or improve safety. Investigation of incidents and accidents is an important part of that process. While an incident under investigation might not directly affect achievement of a TLS (because the TLS does not apply to such incidents), the incident might be a precursor to an event that would count against the TLS. Investigation can reveal trends that give early warning of a potential future breach of a TLS. This is particularly important when the TLS is a very small number and the occurrence of very few incidents or accidents (perhaps even one) would breach it. An incident investigation therefore needs to consider the implications of the incident for the specific incidents or accidents in which the TLS is expressed.

Given the international nature of the aviation industry, it is important that incident investigations are carried out to a uniformly high standard and that lessons learned are expressed in terms that can be readily understood and assimilated across the whole community.

Computer-based Tools Computer-based tools can facilitate consistent reporting and investigation of incidents and accidents. In particular, for the multi-national, multi-cultural aviation industry, the use of a consistent set of tools is a means of ensuring that the data from many countries is reported and can be combined consistently and coherently.

Two examples of developments by EUROCONTROL show the type of developments that are taking place in ATM. Firstly, the Automatic Safety Monitoring Tool (ASMT) is being developed for automatic monitoring of specific safety-related events. Automatic monitoring has the advantages of being objective and rigorous in detecting and recording. However, it can only provide facts, and then only those facts that have been chosen for recording. It cannot provide reasons, opinions or judgement. Automatic monitoring also raises human issues relating to 'the spy in the machine' and the use that is made of the data.

The second example is the TOKAI toolkit (TOKAI = Tool Kit for ATM occurrence Investigation). Occurrences may be recorded initially either automatically or by a human-based reporting system. Whatever the source of the incident report, TOKAI provides functionality for:

- Notification of safety occurrences in standard format
- Investigation activities, including data gathering and input, safety occurrence reconstruction, analysis and classification
- Safety recommendations
- Exchange and reporting according to specified requirements

TOKAI uses the HEIDI taxonomy and internationally agreed standard formats to facilitate consistent reporting and data exchange.

Confidential and Non-Punitive Reporting An incident report is almost always ultimately a report on something done (or not done) by a person. Furthermore, the report is usually made by a person (unless there is an automatic reporting system). Thus, human factors enter impact the reporting of incidents and accidents. Many factors can affect how, or even whether, a person reports an incident. Some are cultural (e.g. not wishing to be critical of a superior), others may depend on their personal situation (e.g. impact on career). In order that reporting is as complete as possible, it is important that people do not feel that they or their colleagues would suffer unjustly by reporting an incident or accident. Confidentiality within a non-punitive safety culture has been widely adopted in the aviation industry as a principle of incident reporting. That is not to say that all incident reports should be confidential, but the ability to submit confidential (including anonymous) reports should be available. How this is best done depends, among other things, on national and organisational cultures.

Conclusions

TLS are increasingly being used in aviation including, more specifically, ATM. Use of TLS has significant implications for incident and accident reporting and investigation since they provide the evidence for achievement of TLS. In particular, incident and accident reporting systems must be designed so that the required evidence is obtained unambiguously. The small numbers for TLS (i.e. apparent very high levels of safety) in a very safe industry such as aviation raise important issues in relation to the completeness of reporting, the consistent and correct interpretation of incidents and the appropriate analysis of resulting rates of occurrence.

The international nature of the aviation industry and the use of common TLS across national boundaries means that incident and accident reporting and investigation must be harmonised. Issues such as standardisation of terminology are important. Computer tools can assist but the human element remains of key importance.

References

Joyce, T., Graham, G., van Eenige, M., Kinnersly, S. & Roelen, S. (2001) A Study into Target Levels of Safety (TLS) Within the Aviation Industry, Including Comparative Analyses with the Rail and Nuclear Power Sectors: ASTER Work Package 1, NLR-CR-2001-145.

EUROCONTROL (2000a) ESARR2 – Reporting & Assessment of Safety Occurrences in ATM, Edition 2.0, Safety Regulation Commission.

EUROCONTROL (2000b) ESARR4 – Risk Assessment & Mitigation in ATM, Edition 2.0, Safety Regulation Commission.

EUROCONTROL (2001) HEIDI Taxonomy, Edition 1.1, http://www.eurocontrol.be/src/index.html, Safety Regulation Committee.

Davies, E.H. (1993) Derivation of Flight Target Level of Safety for Use in North Atlantic MNPS0 Airspace, 6th March.

Review of the General Concept on Separation Panel (1995) A Review of Work on Deriving a Target Level of Safety (TLS) for En-route Collision Risk, Working Group A, Brussels, 1st – 12th May, 1995

Review of the General Concept on Separation Panel (1999) Some Thoughts on Setting the Global Target Level of Safety, 17th – 28th May 1999, Working Group A, St. Petersburg, Russian Federation

Appendix 1 – Example TLS : Aircraft Collision During Flight

The ICAO North Atlantic Systems Planning Group (NATSPG) (1992) used historical accident data combined with a collision risk model (CRM) to determine a TLS for lateral separation (specifically for North Atlantic Track separation minima). The methodology is summarised below.

Step 1: Calculate the rate of fatal accidents for jet aircraft, per flight hour, relating to a chosen period for which historical data is available. The original sample used by NATSPG, which was for the period 1959 – 1966, was for all fatal accidents and all causes. This gave 36 fatal accidents in 15.5 million flight hours, a fatal accident rate of 2.34 x 10^{-6} per flight hour.

Step 2: Assign a proportion of the overall accident rate to collisions. A proportion was set at 1 in 10, i.e. a factor of 0.1, for accidents due to collisions. It is noted that an assumption was made that one collision equated to two fatal accidents. Thus the estimated rate of fatal accidents due to collision was assessed to be 2.3 x 10^{-7}.

Step 3: Apply an improvement factor. To turn an estimated historical accident rate into a future target, it was considered that systems planning should aim at improving the historic safety record. As such, an improvement factor was applied to the historic rate, which was initially chosen to be between 2 and 5 over a period of time of 5 years. Thus a TLS in the range 12 x 10^{-8} to 4.6 x 10^{-8} fatal accidents per flight hour due to collision was set.

Step 4: Apportion the overall TLS to three flight dimensions. As the aim of the original process was to establish TLS for loss of lateral separation, i.e. to support the determination of North Atlantic Track separation minima, the overall TLS was divided into three components of flight: lateral, longitudinal and vertical. Risk was apportioned equally between these three dimensions, hence giving rise to a TLS for collision due to loss of lateral separation, from all causes, in the range of 4 x 10^{-8} to 1.5 x 10^{-8} fatal accidents per flight hour.

The method outlined above is simplistic but has been applied successfully over the past thirty years in setting TLS for North Atlantic and En-route airspace, also the implementation of North Atlantic Tracks and planning for (en-route) reduced vertical separation. As this approach has evolved, the use of historical information has been tailored to specific time frames and phases of flight, and also used to challenge the relevance of factors applied.

Such TLS have commonly been derived for a specific sub-set of operations such as Oceanic or En-route, where the route structures are relatively simple parallel tracks and procedural control only is provided. These structures facilitate the modelling of aircraft tracks and their relationship to target probabilities, based upon the utilisation of readily available data. This process of derivation remains, although the data utilised in deriving the TLS has changed over the past year.

Automatic Safety Monitoring' in Air Traffic Control - Achievements and Perspectives

A. Joyce (1) and Christine Fassert (2),

(1) EUROCONTROL Experimental Centre Safety Business Area , France.
(2) CETCOPRA (Paris-1 Sorbonne), France.

Abstract
The ASMT (ATM Safety Monitoring Tool) is a new safety monitoring tool being developed through collaboration between the EUROCONTROL Experimental Centre (Safety Area business) and the Maastricht UAC (Upper ATC Centre). This tool has been implemented in a pre operational stage in Maastricht in May 2000 and is currently under assessment and being parameterised on real traffic. ASMT is also being implemented in 2 other European members states in year 2002. This paper describes the first achievements and the future perspectives of "Automatic Safety Monitoring " with a particular emphasis on sociological aspects, focusing on the perception of Safety by the various ATM actors. The paper draws on the first results of PhD research sponsored by the Safety Business Area in EEC on the sociological aspects in Safety ("transparency" in organizations on Incident reporting, ATM actors understanding of Safety, impact of Regulation, etc.)

What is ASMT?
ASMT is an on line system that detects safety related occurrences, provides appropriate alerts that an occurrence has been recorded and facilitates safety investigation and analysis. The ASMT:

- Acquires relevant data
- Detects safety related occurrences,
- Provides appropriate alerts that data is recorded,
- Stores safety related occurrence data including alerting acknowledgements and archiving,
- Post processes safety related occurrence data, via both automatic and manual processing
- Enables access to safety related occurrence data through viewing and reporting mechanisms.

The present detection of safety related occurrences developed for introduction into the MAS UAC consists of aircraft proximity encounters below specified separation criteria. For the purpose of this paper, the term "ASMT" refers to a tool with this function, which is the only one developed at this stage. Functions currently under development comprise: detection of level busts (any unauthorised vertical deviation of more than 300 feet from an ATC flight clearance), ACAS Reporting (automatic recording of down-linked reports from Airborne Collision Avoidance Systems (ACAS), data relating to Runway incursions, airspace penetrations, and flight below minimum safe level.

Collecting Safety Events
Highly Reliable Organisations have been identified by the Berkeley group (La Porte, Rochlin) as organizations that perform extraordinary well in spite of their complex and intrinsically hazardous technical systems. ATC was identified by this group as a "HRO", and one the characteristics of a HRO is "a system of rewards for reporting and discovering error". (Rochlin, 1999). However, HRO Researchers do not insist on the Learning process of the organization that is based on the discovery of errors. There is nevertheless now a consensus on the need for Safety events to be "known" and used in a Learning process, even if there are still many controversies on the precise use of Safety Events. Reason (1999) emphasises the critical importance of an effective safety information system that is the main pillar of an *informed culture*. This safety information system collects, analyses and disseminates information from incidents and near-misses as well as from regular proactive checks on the system's vital signs.
Airlines have set up for now some years sophisticated means to detect, collect, analyse and process Safety events, for example with the FOQA programmes.
Contrary to this, in Air Traffic Control, the situation varies significantly depending on where and in what country the ATC system is located. In a few cases safety events collection is very much developed, but in most cases, REX, the set of activities that comprises collecting data about an incident, causal analysis to explain how it developed, and proposition of measures to prevent its occurrence in the future, is at a very embryonic stage.
Lastly, the need for Safety improvement is today emphasised by the Eurocontrol Safety policy, together with promotion of collection and dissemination of safety data, and the development of Safety Management Systems able

to address Safety in "an explicit, formal and documented manner operated by trained personal using dedicated methods, procedures and tools". There are already several sources of incident data such as Airprox, Tcas Alerts, and Human Reporting, and ASMT was first identified as a complement for those sources. However, its implementation raised new questions and issues, that are examined here. The following aspects have been identified through the field analysis of the ASMT implementation in Maastricht UAC, and visits and interviews at several European ATC centres. In the first part, some critical aspects are analysed, and in a second part, the perspectives of such a tool are developed.

ASMT does not monitor the whole "Safety"

One of the first problems identified is the ASMT acronym itself. Is there a tool able to monitor safety in an ATC system? Or more simply, where the separation norm in ATC in the European core area is of 5 NM in the horizontal plan and respectively 1000 (below 290) and 2000 (above) in the vertical plan - does it mean that any separation inside these parameters is a “safety event “?. There are for example, visual separation procedures that allow for a smaller separation. Additionally the separation can be on the margins of these parameters without being considered by a controller as a “safety event", insofar that he/she is comfortable and "has the situation in hand". This is perfectly consistent with Amalberti’s findings on "Control of situation", in which he describes how this process is the core of a complex cognitive model.

Of course, a tool like ASMT is not able to differentiate between a situation in which the controller worked "at or just below the margins"- but safely; from a situation where the separation was "at or just below the margins" bur was not controlled or monitored by the controller. Only the controller himself is able to explain what was exactly the situation. Controllers are aware that there are situations where safety can be jeopardised in when there is **no separation infringement**. A Controller can sense a "loss in safety" in situations where no “objective and measurable” indication can be detected, i.e. nothing to be recorded by a system such as a loss of separation: for example, he has forgotten a conflict, but realized in time, and either recovered the situation before any separation infringement or found that separation is maintained. Is this a more valuable occurrence to record than the “controlled “ situation at or inside the set down margins? In conclusion, ASMT, in its first version, detects separation (proximity) occurrences. Whether a particular separation is a "safety event" or not can be revealed in most cases only by the controller.

ASMT may be perceived as part of a "double bind"

A very important concern is the feeling of the controllers that there is a "double bind" to quote a famous psychological concept where a person is requested to do something and its opposite. On the one hand, controllers are encouraged to handle more and more capacity. It some cases it may be very difficult for controllers to maintain the sector capacity without using VMC climb procedures. In densely used airspace, working comfortably outside the required 5 NM standard without, from time to time, working at the very limit of that required standard in order to expedite traffic, would lead to a capacity decrease.

On the other hand, when a tool such ASMT is implemented – and will screen as a routine any occurrence outside its set parameters, controllers might get the feeling that skilled judgment at the margins will now become a safety occurrence.

ASMT or "Big Brother is watching you"

ASMT may be perceived as a reduction of professional privacy. The tool allows for a display of the alert on any computer, and it is then a matter of associated procedures to define *who* would see the alert (supervisor, safety manager? …). It is however understandable that, in organisations with a punitive culture, controllers would fear a use of ASMT focused on individual error detection rather than on organizational learning. Aspects linked to controllers’ liabilities are therefore crucial as explained in the following paragraph on Legal aspects.

Good examples of how procedures and fair definition of agreements on confidentiality aspects can facilitate trust are to be found in airline operations, as mentioned earlier, various means to collect Safety data, that have been implemented in particular, "automatic monitoring", through the Flight Data Recorder. Some airlines (like British Airways) started using flight data recorders 40 years ago. The systematic analysis of aircraft parameters that are recorded during a flight can be accessed easily by the Quick Access Recorder (QAR), fed by one or more Flight Data Acquisition Units. There is as a consequence, a continuous comparison of flight profile, engine and systems operation with a set of parameters Some events are detected on one parameter, while others need to combine several parameters and the use of complex algorithms (as explained in Holton, 1999).

The principle, similar to ASMT, is to define parameters and to record automatically what is outside the parameters envelope. This allows the automatic recording of flight parameters for each flight. The events (abandoned take off, a

too high approach speed, deviation of glide path,) are automatically detected. As with ASMT, this system permits the detection of errors that might have remained hidden, and consequently raise the problem of professional privacy and liabilities. It is then crucial to set up procedures that protect confidentiality, and thus to prevent a punitive use of the tool.
An example is Air France where the crew concerned by recorded flight anomalies receive a questionnaire[3]; The answering of this questionnaire is handled via a confidentiality policy starting with the use of double envelopes. The questionnaire allows the collection of data to understand what happened, and again, only the crew in most cases, is able to give the context of the incident, to explain what were the errors, mechanisms, recovery processes, etc. This would not be possible with a purely anonymous procedure. It is notable that this procedure was the result of an agreement between the airline and pilot professional organizations, and makes possible, in very exceptional cases (no occurrence till now) to remove the confidentiality.

Legal aspects
Legal aspects are very important, because in some places, errors are still considered not only as taboos, but also as *mistakes*, for which a controller may be prosecuted. Here also again, the European laws vary significantly. Amongst them, Danish Aviation Law was until recently particularly severe. In case of an incident, a controller could be prosecuted and may have been required to pay a fine. The loss of his license, and in some extreme cases, jail penalty were possible outcomes of an error.
The consequence was that spontaneous human reporting was almost nonexistent. Very few incidents were reported or reports were restricted to those cases where "the other side" (pilots for controllers, and vice versa) were to “blame”. For a long time, this low level of incidents was considered as a sign of "Safety performance" by the management. No safety events reported = no safety events at all! Controllers agreed that incidents in fact do occur, but were not reported because of fear of being fined. Recent years brought a greater consciousness amongst controllers and management that the situation of being "blind" on incidents was no longer sustainable. A change was initiated by the Danish ATC professional organization and union (Dansk Flyvelederforening) and the law is in the process of being changed. The expected changes have already resulted in a greater amount of safety occurrence reporting, and the possibility to establish a safety information system.
A non-punitive culture is therefore the basis for effective reporting. This was shown by Dyrlov Madsen and Ryan Jensen (2001) who analysed reporting processes in link with the organizational culture. They noticed that, in some cases such as the Danish trains and some hospitals, the removal of punishment very quickly led to a lot of reports. This is of course not the same as developing actual organizational learning, (that comprises an overall frame of a systemic approach to safety analysis, regular debriefings, etc.) but at least a first step. The speed by which reporting develops after punishment has been removed seems to depend partly on the general organisational culture and on the amount of confidence the employees have in that punishment is really (also in non-legal terms) removed. Their research focuses on a comparison between Sweden (non punitive system) and Denmark. In Sweden, the reporting has developed over a wide period, although it is now difficult to realize how much time was needed from the change of Law to an effective reporting process.

Future Challenges
ASMT and "transparency" in ATM organizations: Various other problems linked to the use of ASMT have emerged. We examine below some different roles played by this particular type of "technical object", from a socio anthropological perspective. As explained by Poirot Delpech : "a project is never purely technical. Human dimensions (e.g., political ones) are invariably embedded in its realization. A technical object remains open to uses and appropriations often not predictable. The Social aspect of a project is part of its definition: it is impossible to stop the definition of a project at its initial purpose.

Undoubtedly ASMT makes visible some safety events that may have remained hidden. But the status of this tool varies according to the roles it is required to play. This is very much linked to the procedures associated with its use. Below we have identified 3 contrasting roles, that we will call: "spy", "revealer", and "uncoverer".

*ASMT as a spy***:** in this case, it is considered that ASMT should allow detection of a safety occurrence that has not been spontaneously reported by the controller. But is it reasonable to focus on the detection of one particular incident made by an individual? The role in this case could be viewed simply as a “snitch” tool ensuring that a controller should not be able to hide any error, and might be punished or prosecuted for it. This not the role

[3] Interview with Mr Martegoutte, Flight Data Recording Safety Manager, Air France.

envisaged by Eurocontrol who develop the ASMT concept to contribute to safety management through organisational learning. IFATCA also clearly expresses these concerns in its ASMT policy and states "the ASMT device is something that should be used for the analysis of the circumstances of any perceived loss of separation, rather than as a punitive tool that can be used directly for disciplinary action".

ASMT as a revealer: In use where the focus is not on individual error, but on data collection, it can happen that ASMT *reveal*s something that is anticipated, "known" by the ATM actors (whatever those are: controllers, managers, regulators,), but not objectively demonstrated or quantified. In this case, ASMT plays the role to simply show what was anecdotally already known or in the collective thinking of the organization. For example, many persons in the Maastricht UAC were conscious that, due to the high traffic, it is necessary in some parts of the airspace to operate constantly at the margins of separation requirements. ASMT pre operational implementation on real traffic is able to show how true this feeling was. Even if there is no actual *discovery*, there is however a difference between a feeling, an intuition, and objective records on a computer. The latter enable the raising of some questions: is it OK? What is the "safety status" of this data? On a more general level, ASMT may challenge where it is implemented, the perception and understanding of Safety.

ASMT as an "uncoverer": Here the tool is able to detect something that an individual or a collective of people, as human beings, have not perceived. ASMT may be able to detect some patterns, and, more interestingly, it may be used in a proactive way as a testing of hypothesis such as, for example: "in which airspace do we have more level busts?" For example, in the Maastricht centre, EGATS made some concrete proposals to use ASMT more as a statistical tool and an Airspace complexity Analysis tool (i.e., with other parameters than those linked to STCA). An Operational Evaluation working group (OEG) was formed to evaluate what questions should be asked of ASMT. An Operational Analysis Group (OAG) will start work shortly to analyse the results from ASMT. This can be considered as a positive example of the appropriation of a tool by its users.

Conclusion

In Air Traffic control, Automatic Safety Monitoring is quite a new thing, as is the implementation of ATM Safety Management Systems in many Eurocontrol member States.

The gathering and analysis of safety occurrence data forms the backbone of the learning and safety improvement cycle. Automatic data gathering can greatly enhance the quantity and quality of such data. However, experience from the pre operational implementation of the "ATM Safety Monitoring Tool" in the Maastricht UAC clearly illustrates that no "system" can be restricted to its technical aspects. On the contrary, the use of a tool like ASMT in ATM probes many fundamental non-technical aspects of the Safety Management System such as the organisational culture, controller liabilities, professional privacy, Safety perception, and user appropriation.

References

R. Amalberti : *La conduite des systèmes à risques*, PUF, Le Travail Humain 1996

Marlene Dyrløv Madsen & Thomas Ryan Jensen*: Fejl, ansvar og moral: Behandling af menneskelige fejl og udvikling af en professionsetik inden for flyveledelse og andre sikkerhedskritiske områder.* Technical Report R-1260, Risø National Laboratory, 4000 Roskilde Denmark. 2001.

Holton (Captain Mike Holton). FOQA : Aviation's most important safety tool. British Airways. 52 FSF Annual International Air Safety Seminar, 29th IFA international conference and IATA. November 1999.

Poirot Delpech, Sophie L. : Biographie du CAUTRA. Naissance et développement d'un système d'informations pour la circulation aérienne. Thèse de doctorat de sociologie, Université de Paris I.

Reason, J. : M*anaging the Risks of organisational accidents* , Ashgate, 1997

EPOQUES: Proposing Tools and Methods to treat Air Traffic Management Safety Occurrences

Hélène Gaspard-Boulinc (1), Yannick Jestin (1), Lionel Fleury (2)

(1) Centre d'Etudes de la Navigation Aérienne, France, {helene,jestin}@cena.fr
(2) Axilya, France.

Abstract: This paper presents first results obtained within a research project called EPOQUES, whose aim is to propose tools and methods to treat Air Traffic Management (ATM) Safety Occurrences. The project is based on a bottom-up approach and centred on investigators' needs in the five French Air Traffic en-route centres, plus Paris/CDG and Toulouse approaches. The reported work has been realised by a multidisciplinary team that gathers Human-Computer Interaction engineers, Human Factor specialists, a graphic designer and a safety expert. On the one hand, results concern the job of ATM investigators, as it had been captured in the five French en-route centres, from December 2000 to June 2001. On the other hand, results concern prototypes that have been evaluated by investigators during workshops since October 2001. Participatory design and iterative prototyping are being used to define a set of investigation tools and grant the usefulness and usability of these envisioned tools.

Keywords: scenarios, participatory design, Air Traffic Management, Safety Occurrence, prototypes.

Introduction

The concept of Air Traffic Management (ATM) developed by the International Civil Aviation Organization (ICAO) over the last ten years has been defined as follows: "Air Traffic Management consists of a ground part and an air part, where both parts are needed to ensure a safe and efficient movement of aircraft during all phases of operations". "The main objectives of the ATC service are to prevent collisions between aircraft and between and obstructions on the manoeuvring area and to expedite and maintain an orderly flow of air traffic" [ICAOa]. It is therefore clear from the texts that the ATM objectives relate firstly to safety and secondly to the criteria of efficiency and equity among the users. This paper presents first results obtained within a research project called EPOQUES, whose aim is to propose tools and methods to treat ATM Safety Occurrences. On the one hand, results concern the job of ATM investigators, as it had been captured in the five French en-route centres, from December 2000 to June 2001. On the other hand, results concern prototypes that have been evaluated by investigators during workshops since October 2001.

The Methodology: Analysis of Current Work Practices and Participatory Design

In order to identify the investigators' needs, two complementary methods have been adopted. Firstly, an analysis of existing practices and tools has been realised in situ, through observations, interviews and documents reading. The outputs of this phase are documents that describe existing work and allow to define initial needs. Scenarios of current work practices are identified from stories collected by observations and interviews with investigators and air traffic controllers. These scenarios have allowed designers to identify functionalities and generate first prototypes, ensuring those prototypes are grounded in every day's use.

Secondly, participatory design is used to refine the initial needs. Participatory design, as proposed in [Bjerknes, 1987], is a collection of theories and practices that aims at increasing the role of users as active participants in the design of social systems, including computer systems. In our case, practicing participatory design means the involvement of investigators in the tools design, bringing users and designers together in order to take benefit from their complementary expertise, and obtain useful and usable tools. As proposes in [Holtzblatt, 1998], prototyping is used to drive the design process, and allows designers to iterate, refine and extend the initial needs. Workshops with investigators have been organised monthly, composed of an evaluation part and a conception part. A workshop is a three day session in an air traffic en-route centre, composed of a project presentation, individual trials of prototypes, a collective brainstorming and a debriefing. The objectives of individual trials were to collect feedback on prototypes, and identify new functions and needs through observations and questionnaire. Investigators were asked to use the prototyped reconstruction tools in order to understand facts on one past occurrence. Collective brainstorming gathered the investigation unit and designers to concentrate on one part of their activity, listing current benefits and drawbacks in order to imagine new working practices and tools. A workshop has been

organised in the five French en-route centres from February to July 2002, with an average of 4 individual trials for each en-route centre.

ATM investigators' job

In France, the treatment of ATM safety occurrences is realised by an unit named QS, standing for **Q**uality of service and **S**afety, located in each air traffic en-route centre and approach. A QS unit is generally composed of a permanent head of unit, and of non permanent air traffic controllers that leave the operational room to work at the QS unit during six to twelve months. From the description of the QS formal tasks, we would like to point out the difficulties that face QS units in their daily activities, as it had been observed from December 2000 to June 2001 in the five French en-route centres.

QS Formal Tasks: Considering the description of formal tasks of investigators, the generic phases are common to many occurrence-reporting systems [Eurocontrol, 2000], as shown in figure 1. Occurrence notification is followed by data gathering, in order to allow occurrence reconstruction. The next step is occurrence analysis followed by communication on lessons learnt from occurrence investigation. Recommendations are proposed on the basis of the investigation, and the occurrence is recorded and integrated in database systems.

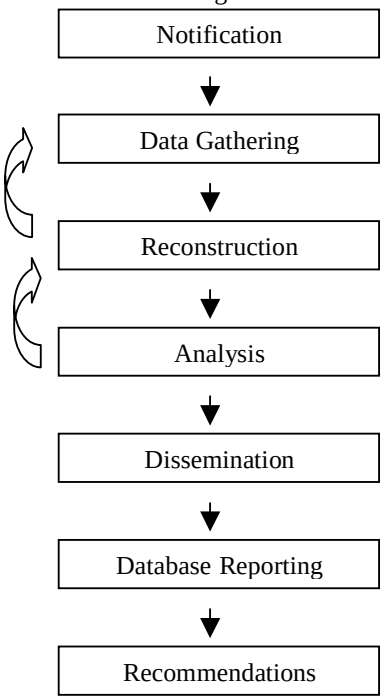


Figure 1: QS formal tasks

However, when analysing the current work practices of QS people, some difficulties in the execution of tasks arise. The following sections detail the difficulties in the notification, data gathering, reconstruction and analysis phases, as it was observed in the five French en-route centres from December 2000 to June 2001.

The notification phase: Concerning the occurrence notification, two aspects can be underlined. The first aspect is the diversity of requests that arrive at the QS unit. In fact, the requests are quite diverse, and imply multiple working situations. Two scenarios can illustrate this diversity, the AIRPROX procedure and the informal request from Air Traffic Controllers. On the one hand, the most procedural request is AIRPROX, that is the report of "a situation in which, in the opinion of a pilot or air traffic services personnel, the distance between aircraft as well as their relative positions and speed have been such that the safety of the aircraft involved may have been compromised" [ICAOb]. AIRPROX are actually only initiated by pilots, providing brief details of the incident in a radio message to the concerned ATS unit. The main steps to process an AIRPROX are the preparation of a draft report by QS people, the submission of the report to a local commission for validation, the submission of the report to a national unit for

verification, and finally the submission to an independent council. On the other hand, Air Traffic Controllers are used to asking QS people to re-visualise a situation that they have lived, to better understand what happened in terms of aircraft separation. These controllers' requests are completely informal but require special attention from QS people, since their ability to provide controllers with adequate answers build a relation of trust with controllers. Therefore, envisioned tools must be designed to take into account this great diversity of situations.
The second difficulty that appears in the occurrence notification phase is the non predictability of the number of requests. The QS unit, with a constant staff, can face an increase in the number of requests at any moment. However, the delay to treat the requests must stay reasonable: when controllers have submitted a safety report, it is important to treat it in a prompt and timely manner [Eurocontrol, 2000], in order to show that their input is taken into account and that lessons can be learnt from their reported experience. Moreover, rapid action must be taken for some data to be available, and thus some data will be lost if the treatment is delayed. The consequence for QS units is some difficulties in work organisation, to ensure timely treatment of requests. The need for a tool to manage requests and work organisation has been identified. To give some idea of their typical workload, an average of two thousands safety reports are submitted by controllers per year for an en-route centre. In the same time, an average of 29000 short term conflict alerts is recorded per year. After an automatic elimination of anomalies and unjustified alerts, justified alerts are automatically classified according to the non-respect of separation minima. Finally, 1000 recorded alerts must be rapidly visualised and inspected by QS officers to select the significant alerts that will be deeply investigated. An average of 15 significant alerts will be selected for investigation and dissemination, on the basis of QS officers expertise.

The Data Gathering and Reconstruction phases: Data gathering, as one can see on figure 1, is crucial since the available data are used to understand what happened. Multiple sources of data must be inspected. There are technical logs from systems such as radar and flight plan information, audio recordings from telephones and radios. There are paper data, such as flight strips and meteorology information. There are declarations from the different people implied in the occurrence. For AIRPROX, a checklist specifies a list of data to be gathered. When current work practices are observed, the QS people spend a lot of time to gather data for an occurrence. In fact, access to the data implies to know how to use multiple systems, with different operating logic. Data are available on different supports (paper, floppy disk, cd-rom, network) and different formats. Therefore, there are difficulties to collect data and to combine them in order to reconstruct the occurrence situation. The result is a lot of time spent on data collection and a resulting reconstruction that is quite different from what controllers have lived. The efforts that require data collection and reconstruction are due to systems weaknesses since the used systems are not initially designed for safety analysis objectives. Moreover, some QS units still have difficulties to obtain controllers' involvement in the occurrence investigation. Controllers can have problems to accept error in both individual and team points of view. Some fear of retribution and legal sanctions can also explain controllers' reluctance [CRNA/Nord, 2002]. Difficulties in data gathering and reconstruction can have an impact on the analysis phase, as shown in figure 1.

The Analysis phase

Analysis aims at identifying the chronology of events, and to explain how or why it has been interrupted before accident. Collected Data are used to reach this aim. Consequently, analysis relies on the selection and availability of data. QS people search for what seems to be abnormal in the situation, make hypotheses and check them with data. Interviews of people implied in the situation are useful in this phase, to complete the occurrence reconstruction in terms of context, intentions and situation awareness. These latter are key elements that help QS officers to make hypotheses and search for pertinent data. Therefore, the lack of controllers' involvement in occurrence investigation can be a real handicap.

From these difficulties observed in QS current work practices during the phases of notification, data gathering, reconstruction and analysis after any occurrence, needs have been identified in terms of functionalities, human-machine interfaces and methodology. Prototypes have been developed and proposed to QS people in order to refine the initial needs and validate the usefulness and usability of proposed tools.

Paper and high fidelity prototypes

From Current working practices to Prototypes: A multidisciplinary team participates in the prototypes development. Ergonomics people and a HCI research engineer have translated the observed working practices and their difficulties into significant scenarios in order to identify adequate functionalities and to ground them in working situations [Mackay, 1997]. Human-computer interaction engineers, combined with a graphic designer, produce intuitive and usable interfaces, that can allow non permanent QS officers such as air traffic controllers, to

become autonomous in the occurrence treatment quite rapidly. This is crucial if we consider the important turn-over in a QS unit. A safety expert guarantees the consistency of vocabulary introduced in the interfaces, avoiding the confusion of different terms such as incident, danger, risk, prevention, etc.
In order to ensure the usefulness and usability of envisioned tools, workshops with QS people have been monthly organised for evaluation and conception. The result is a first set of investigation tools, that are presented in the next section. We will then detail the use of these tools as it has been observed during workshops with QS people.

The main interface paper prototype: In order to support the data gathering phase and potential iterations, a paper prototype has been designed. The "main interface", as shown in figure 2, is inspired from Acrobat reader software, displaying on the left a check-list of actions, and showing what has been done and what is still to be done. The selection of an item of the left sidebar opens dialog boxes. Three main boxes can be displayed. The first one allows to define the type of occurrence. Each type of occurrence is associated with a check-list that defines a set of data to be collected. The second box allows users to define the occurrence characteristics. The occurrence characteristics will correspond to criteria for data search. If characteristics are incomplete, the results of data search will be a list of air traffic situations that match the criteria. The validation of characteristics triggers data searching. A third box displays for each data its source, search criteria and availability. This box will support iteration between the analysis phase and the data gathering and reconstruction phases, allowing user to modify search criteria, or to search for extra data. The visualisation of available data can be triggered from this box, through a "play" button that gives access to reconstruction tools.

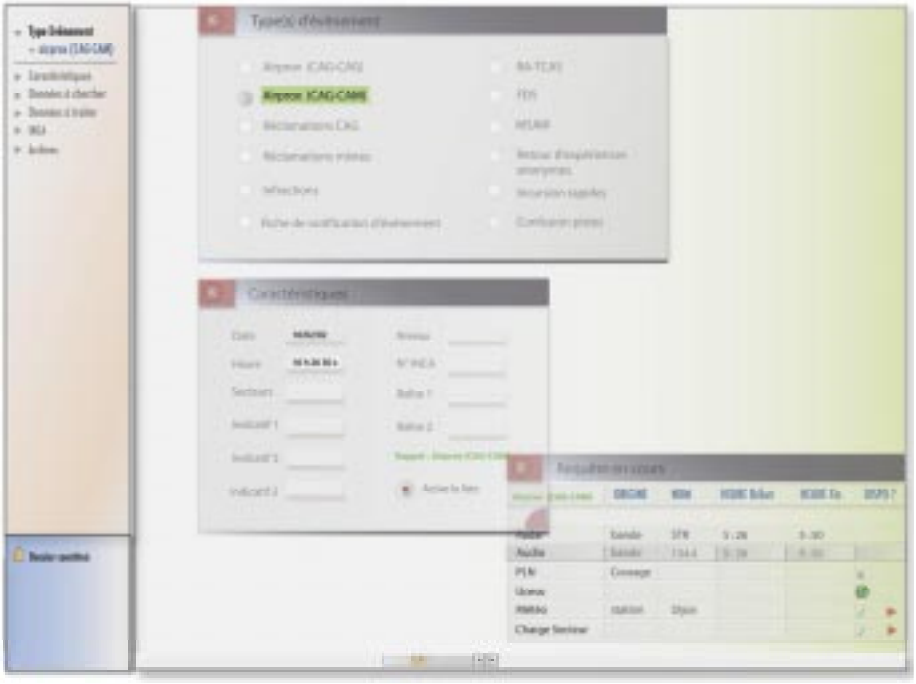


Figure 2: The "main interface" paper prototype

The reconstruction tools

Reconstruction tools are composed of a radar image, plus a set of toolglasses that allows users to navigate in the situation (time and space) and display information and calculation for analysis. A key point can be the deep link between the main interface and the reconstruction/analysis tools. From reconstruction of the occurrence, users can go back to the main interface to constitute the final report.
The radar image is a high fidelity prototype, using state-of-the-art rendering techniques such as animation and transparency. Toolglasses have been first paper prototypes created by a graphic designer, and were then implemented. Figure 3 shows a picture of the paper toolglasses.

The radar image, synchronised with pilot-controller and controller-controller communications, allows QS officers to understand the relevant aircraft behaviour and to capture the chronology of events. It is also used to visualise extra data such as flight plans and trajectories. Concerning the toolglasses, the idea is to be as intuitive as possible, keeping in mind that QS officers are not computer experts and have a short learning period. There is a VCR-like toolglass to pilot the time of the replay, with functions such as play, fast replay, stop, jump to a specified moment or to the time of alert if any. One toolglass is audio-oriented, allowing to listen to radio and telephone communications, detect sound activity and jump from message to message. Another toolglass is geographic oriented, with functions such as zooming, panning. It also allows the centring on the aircraft relevant to the situation and specified before in the main interface, the centring on alert location if any. An information-oriented toolglass allows users to display Air Traffic information on the radar image such as distance markers, waypoints, speed vectors and traffic layers. A snapshot toolglass allows users to print and grab representations to illustrate the final report.

Figure 3: from left to right and top to bottom, the VCR-like, the audio-oriented, the information-oriented, the geographic-oriented and the snapshot toolglasses

In addition to the radar image, two other representations have been proposed to understand the air traffic configuration. The horizontal trajectory representation isolates the aircraft relevant to the situation, displaying radar recorded points and stressing radar points that correspond to a non-respect of separation minima. The vertical trajectory representation shows the evolution of vertical and horizontal separations, and calculates a safety ratio to identify the closest point of approach between aircraft. These two representations also allow users to pilot the time of replay.

These reconstruction tools have been iteratively produced and evaluated during workshops organized with QS people. The next section reports how the reconstruction tools have been used by QS officers to support investigation during workshops.

Firstly, workshops have shown that the designed reconstruction tools require a short learning period: a quick presentation of the tools has been enough to allow QS officers to use the interfaces efficiently. Secondly, workshops have allowed us to capture how the proposed reconstruction tools can be used by QS officers to deal with different working situations. The following examples illustrate how prototypes can meet QS officers' needs.

Prototypes walkthrough

The need to shift from a context-oriented navigation to a focus-oriented analysis has rapidly emerged from the first workshop, and confirmed by the following ones. The design of toolglasses and representations has been driven by this need to first have a global view of the occurrence situation, and then focus on specific points for analysis. The three following examples describe the different uses of the tools we provide.

For pre-analysis

When QS officers have to browse short term conflict alert (STCA) system recordings of one day, they want to rapidly visualise each corresponding air traffic situation in order to select those that are really interesting to investigate. Radar logs and audio recordings are used in this phase of pre-analysis. Several toolglasses have been identified during workshops to support this pre-analysis. For example, the VCR-like toolglass can first be used to go to the time of the next alert. QS officers can then use the VCR-like toolglass or the audio-oriented toolglass to set the clock around ten minutes before the alert. The geographic-oriented toolglass allows users to identify the aircraft concerned by the alert, with the callsign centring function. The VCR-like toolglass fast replay function can be used to visualise the behaviour of concerned aircraft and the air traffic situation context on the radar image. In parallel, the audio-oriented toolglass can allow users to visualise the global activity on the radio frequency, with the number of messages and the occupation rate, and the potential activity on other channels. The VCR-like toolglass normal replay function can be used to listen to pilot-controller communications, with the synchronised radar image. This sequence of actions can be made quite easily using the toolglasses and the radar image, and can allow QS officers to rapidly understand the context of the alert. This context information, combined with their expertise, can support their decision to keep or eliminate the alert for deeper analysis.

For analysis

The designed toolglasses and representations also offer functionalities to partially support the analysis. QS officers can display extra data such as flight plans and real trajectories on the radar image, to detect route deviations. The audio-oriented toolglass can be used by QS officers to make the transcription of selected radio and telephone channels, proposing the past, next and repeat buttons to jump automatically from message to message. The horizontal trajectory representation can be used by QS officers to visualise concerned aircraft radar positions, and to confront different events such as the loss of separation minima, the flash of short term conflict alert on controller's working position and the traffic collision avoidance system on board. This representation has been used by QS officers to pilot the time of replay through the direct manipulation of aircraft. In parallel, the snapshot toolglass can be used to grab representations and other toolglasses in order to make future readers understand analysis elements.

For dissemination

Finally, the use of the proposed tools to communicate on lessons learnt from occurrences has been clearly identified. In fact, QS officers are used to organising safety briefings with controllers to communicate conclusions of occurrence investigations and make them aware of the weaknesses and strength of the system. The video projection of the radar image, combined with the use of toolglasses to pilot the replay and stress some elements dynamically, can facilitate the understanding of the conveyed message. The use of the radar image during different phases of investigation has shown the need for different configurations. When QS officers communicate with controllers during informal requests or safety briefings, it is important for trust and credibility reasons to have a radar representation close to the controller working position radar image. However, during the analysis phase, QS officers need to display extra data on the radar image and build their own representation.

Perspectives

The reported work is a first step based on the practical experience of Quality of Service & Safety units in French air traffic en-route centres. Our hypothesis is that the proposed reconstruction tools should allow QS officers to gain time in the data gathering and reconstruction phases of occurrence investigation, and thus to have more time to spend on the analysis and dissemination phases. Consequently, the second step of our work will concentrate on the evolution of tools and methods to support analysis and dissemination phases. The prototyping environment that has been built will be a good basis to integrate new occurrence representations, such as a timeline to show how an occurrence develops over time [Eurocontrol, 2000]. It will be also a good platform to study how QS officers could spot emerging trends and common features from the increasing number of safety reports submitted by air traffic controllers.

Conclusion

Our approach is multidisciplinary and based on Air Traffic Management investigators' needs captured in the five French en-route centres. The use of scenarios and participatory design have allowed us to propose paper and high fidelity prototypes that meet investigators' needs and integrate their working context.

Incident reporting systems provide an important tool to prevent accidents in many safety-critical industries . The relative frequency of incidents, as opposed to the relative infrequency of accidents, helps to evaluate the efficiency of safety barriers over time and identify dangers [Reason, 1998]. In this context, providing investigators with adequate tools and methods is crucial and should require particular attention from managers.

References

[Bjerknes, 1987] Bjerknes, G., Ehn, P., and Kyng, M. (1987): Computers and Democracy- a Scandinavian Challenge. Avebury, Aldershot.

[CRNA/Nord, 2002] Clamens Laurie, Magerel Julien (2002): Les obstacles à l'investissement des contrôleurs aériens au cours des enquêtes réalisées par les subdivisions QS des CRNA : un frein à l'amélioration de la sécurité, Mémoire de fin d'études ICNA.

[Eurocontrol, 2000] Johnson, C., Le Galo, G., Blaize, M. (2000): Elaboration of guidelines from ATM occurrence Investigation

[Holtzblatt, 1998] Holtzblatt, K., Beyer, H. (1998): Contextual Design: defining customer-centred sysems, pp 370-377, Morgan Kaufmann Publishers, Academic Press.

[ICAOa] Doc 9583, AN-CONF/10, Appendix A

[ICAOb] PROCEDURES FOR AIR NAVIGATION SERVICES, RULES OF THE AIR AND AIR TRAFFIC SERVICES, doc 4444-RAC/501, International Civil Aviation Organization.

[Mackay, 1997] Wendy E. Mackay, Anne-Laure Fayard : HCI, Natural Science and Design: A Framework for Triangulation Across Discipline. Proceedings of Designing Interactive Systems DIS97, pp233-234.

[Reason, 1998] Reason, J., Managing the Risks of Organizational Accidents, Ashgate.

Safety Data Collection In British Airways Flight Operations

Mike O'Leary (1), Carl Macrae (2), Nick Pidgeon (2).

(1) British Airways Safety Services, Compass Centre (S742), PO Box 10 Heathrow Airport, TW6 2JA, UK
mikeoleary@compuserve.com

(2) School of Environmental Sciences, University of East Anglia, Norwich, Norfolk, NR2 7TJ
c.macrae@uae.ac.uk, n.pidgeon@uea.ac.uk

Abstract: This paper describes three safety programmes that provide British Airways Flight Operations with feedback on operational quality. Two of these are the Air Safety Reporting and the Human Factors Reporting programmes and the third is an automated electronic data collection process. These programmes are described to indicate the broad range of data collection considered necessary for effective flight operational safety monitoring and management. Of the three programmes most emphasis will focused on the human factors programme. Its analysis process is fairly novel and a study of flight crew situation awareness described below will indicate its ability to deal with complex issues. A brief discussion of how BA assesses risk in its safety reports brings this paper to a close.

Keywords: aviation, incident reporting, human factors, situation awareness.

Introduction

The traditional method of uncovering unsafe events was to have an accident investigation. That of course requires an accident and largely defeats our purpose. Ideally one should use as many processes as necessary to uncover unsafe events and discover their cause or, more likely, their causes. The more we are able to discover and examine these events we can learn much about their genesis and develop strategies that will help control or eliminate them. Simply stated, we need good operational feedback so that we can apply corrections to the system to maintain its stability. James Reason (Reason, 1990) explained this process elegantly in his book on Human Error. But in fact aviation authorities and airlines had in many cases established safety reporting programmes many years previously - and with great success. However, one point made by Reason has often been overlooked; one feedback loop is not necessarily sufficient. He suggested that feedback was required from all levels of a company and the more variety the better. British Airways (BA) may not be quite as sophisticated as Reason might prefer but over the last three decades we have established three complementary formal feedback loops to monitor Flight Operations safety performance. The three are: (1) Air safety reporting, ASR; (2) Operational quality monitoring, FDR; (3) 'Human factors' reporting, HFR. The three are components of the British Airways Safety Information System (BASIS), a suite of interacting software modules that underpin our safety management.

Figure 1: BASIS Entry Screen

As can be seen from the screen shot of the BASIS entry screen (Figure 1, above) there is a wide variety of modules from which to choose. Functionally they range from administrative functions such as the Audit module to Ground and Air safety reporting modules, and from human factors analysis to highly technical modules dealing with the recording, analysis and visualising of digital flight data. For our purposes here, we will concentrate on ASR, HFR and FDE.

The Air Safety Reporting Programme

British Airways run air and ground safety reporting programmes that require all staff to report safety related events. For flight crew, the events that require the filing of an Air Safety Report (ASR) extend beyond those required by the Civil Aviation Authority's Mandatory Occurrence Reporting programme and include a general requirement to report *anything* that could have safety implications. The net is cast very wide.

The reports are written on a standard form which requests many specific details concerning the flight circumstances such as the time of day, the weight of the aircraft and where the aircraft was at the time of the incident. The BASIS ASR module records all these details and the text description of the incident is also stored verbatim. An analyst encodes the report with a small selection of BASIS References that characterise what kind of event had occurred, and also with a (usually larger) selection of BASIS Keywords that define the event more precisely. The References and Keywords can be used to filter the database for specific types of events and the number and types of events can be graphically displayed over time or location or any one of a number of factors.

The programme is typical of many similar programmes throughout the world. Such programmes are often run by the local aviation authority or, as in our case, by the airlines themselves. Independent organisations can also be involved in this kind of endeavour. For instance, in the USA NASA runs the highly successful national Air Safety Reporting System (ASRS) for the aviation industry. Where and how these programmes are run depends very much on local culture and legal structures (For a general discussion on reporting programme see O'Leary and Chappell, 1996). However, the BASIS ASR programme has the distinction of being the most popular air safety-reporting programme having been adopted by well over 100 airlines around the world.

The ASR programme was the first of the many BASIS modules. Its success is probably largely due to its versatility. It includes basic filing cabinet functions such as storage and indexing; the facility to include analytic 'keywords or 'descriptors' which also provides for a huge variety of search and filtering options; the search / filtering also supports a graphical system to indicate trends over time; and when networked (which is its normal mode) the built in communications processes provide an effective method of 'actioning' departments to investigate specific aspects of an event. As an extension of the ASR module, the Safety Information Exchange (SIE) module allows the sharing of ASR information between other airlines that use BASIS. IATA, the International Air Transport Association, has recently taken on the duty of managing the SIE and has renamed the programme 'Safety Trend Analysis and Data Entry System', STEADES. The expectation of this new arrangement is that STEADES will become a truly global safety information resource allowing access at low cost to all the IATA airlines of any size, huge or tiny.

One of the reasons for the success of the programme, at least within BA itself, lies not in the technology but in the organisational culture in BA. The safety culture that supports such success results from hard organisational factors not (only) relying on the willingness of the flight crew to allow themselves to be permanently under the microscope. The ASR programme also has strong management guarantees to support it. British Airways Standing Instructions No. 4, signed by the CEO, is directly concerned with the reporting of safety incidents. It states:

> 'It is not normally the policy of British Airways to institute disciplinary proceedings in response to the reporting of any incident affecting safety.
>
> *British Airways will only consider initiating such disciplinary action where, in the Company's opinion, an employee has acted recklessly, or omitted to take action, in a way that is not in keeping with his/her responsibilities, training and/or experience.*
>
> *The fact that the employee has fully complied with his/her responsibilities to report the circumstances and to co-operate fully throughout any investigation will weigh in his/her favour in the Company's consideration of the matter.*
>
> However, in the event of an employee failing to report a safety related incident that they have discovered, they will be exposed to full disciplinary action.'

It is clear from the above that management considers that learning from incidents is more important than punishing the 'culprit', and that the worst crime is to attempt to cover up an incident

Within BA the success of the programme is told in two statistics. The ASR filing rate has increased more than fivefold since the beginning of 1991 (when BASIS was first introduced) whilst the number of reports assessed as high risk has decreased dramatically between 1994 and 1999 (1994 was the year when standardised risk assessment was introduced.). Figure 2 indicates that High Risk events contributed 2.7% (123 high risk events) of the 4613

reports in 1994 and 0.2% (19 high risk events) of the 9345 in 2001. Given that the number of sectors flown by BA varied only +/-3% over the 8 year period (with the low point in 201), these data suggest that BA's safety culture has encouraged and improved on what was an already high level of safety reporting (in comparison with contemporary world-wide industry statistics). Furthermore, it indicates that its risk management strategy has been successful in driving down the frequency of high risk events.

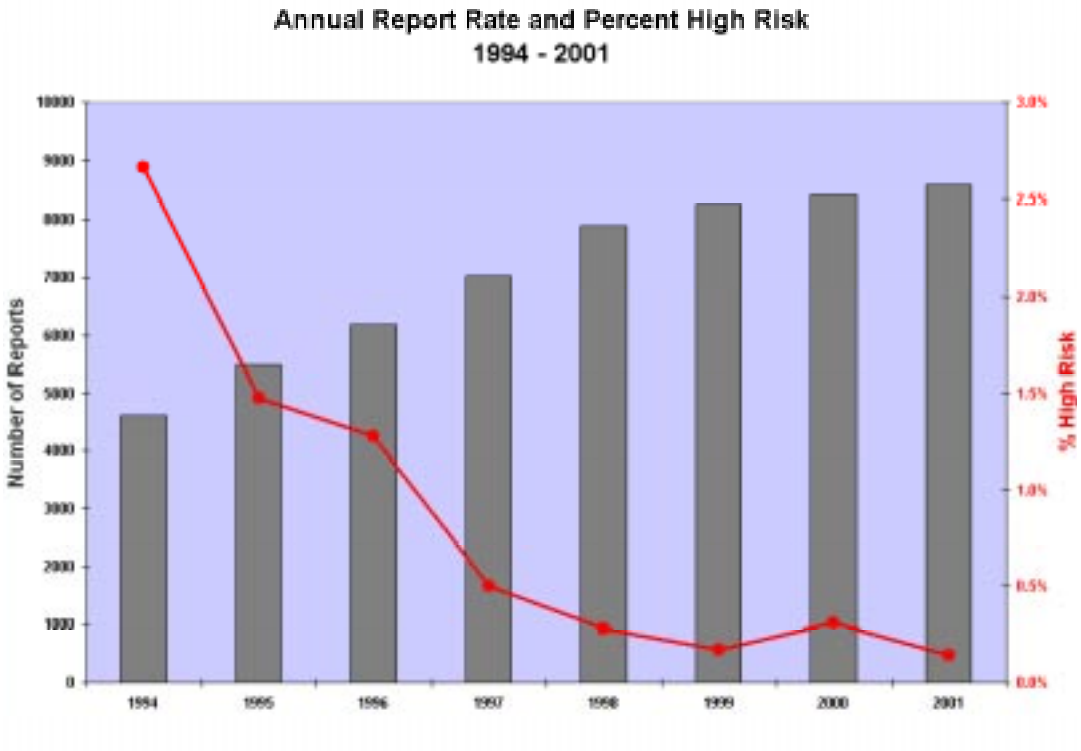


Figure 2: ASR: Report Rates and Risk: 1994 - 2001

The Flight Data Recording Programme

Flight Data Recording is a process where flight (and in-flight engineering) parameters are continuously stored on suitable media (magneto-optical diskettes nowadays) for subsequent download and analysis. Analysis of the data can take many forms but one of the most useful from a safety perspective is the detection of 'events' where flight parameters significantly exceed their safe normal range, indicating that the flight was mishandled by the pilot or was being disturbed by some environmental or technical influence.

The detection and analysis of such events was the function of the original FDR programme called SESMA – Safety Event Search and Master Analysis. The name still persists. SESMA events are discussed at monthly meetings of technical managers and union representatives. If this group considers the event serious enough (particularly if it appears to be a crew caused event) the union representative will be charged with discussing the event with the operating crew. Any information offered by the crewmembers is then reported, anonymously, to the next monthly meeting and used to improve understanding and management of the safety problems involved. The information is also published in monthly fleet newsletters with, of course, appropriate de-identification.

The SESMA programme is supported by a documented agreement between British Airways and the British Airline Pilots Association (BALPA), the pilots union. Despite initial misgivings, both sides keep within the programme guidelines and the proof of both sides' good intent is that no breakdown in the process has occurred in the thirty years of its existence. Both the FDR and ASR programmes have benefited greatly from their respective corporate guarantees and the years of experience with the programmes have created a degree of understanding and trust on both sides that has had enormous benefit for safety.

Beyond detecting event rates the FDR software can calculate and display individual parameters' distributions over time or space (or both). For instance it is a simple mater to calculate and display a distribution of aircraft speeds on approach at 1000' above the ground. This allows examination of how closely a fleet has been following the standard operating procedures, or how well a training initiative has improved flight crew performance (maybe in response to management's dismay at the result of the previous example), or finally, the same distribution could be

plotted for a particular runway at a particular airport in response to pilot reports that ATC were mismanaging approaches at that airport. As a different example, although just as important to BA, the programme can monitor vertical acceleration (g) on take-off and landing to indicate whether uneven or rough runways surfaces were causing more airframe stress than the manufacturer's had anticipated.

The Human Factors Reporting Programme

With a successful FDR operational monitoring programme and a well supported ASR programme, one could reasonably ask why another programme is required. In fact there were two related influences that brought the HFR programme into existence. First, in the late 80's and earl 90's, there was, and largely remains, a worldwide concern that the contribution to accident statistics of 'crew error' remained stubbornly high. Second, whilst the ASR programme was giving excellent information concerning what problems were affecting our flight crew there was little feedback on WHY these problems occurred, nor on how effectively the crew coped with them. Without this kind of feedback safety managers could only react to problems not anticipate them. They realised that some form of proactive safety management was not only desirable but essential. The human factors programme was born of these concerns. At the time of its introduction it is believed that it was unique but more recently other airlines have begun to use similar schemes.

Given the obvious sensitivity of reports that might frequently concern flight crew failures, the HFR programme was made both confidential and voluntary. Consequently, the human factors programme is based in the Safety Services department (not in Flight Operations) and is run by line pilots who are specially trained for the job. Confidentiality is guaranteed. Only the analysts dealing directly with the HFRs know the names of the reporters and these are not entered into the database. The issues raised in the reports are communicated to line management on a regular basis but great care is taken to separate the issues from the incidents in order to safeguard the identity of the reporters.

Most of the HF reports result from the filing of an ASR. When an ASR is received, Safety Services sends a reply to each crewmember of the originating flight. If the report suggests that human factors might have been involved a human factors questionnaire is sent along with the standard reply to the ASR. The reply rate from solicited reports provides further useful information on about ten percent of the ASRs and in some cases HFRs are received from more than one crewmember involved in an event, or in some cases without an ASR having been filed. The next section describes the workings of the HFR programme in more detail. For a tabulated comparison of the three programmes see Appendix A.

Human Factors Report analysis: In contrast to the ASR the HF questionnaire asks how and why the reported event occurred and how the flight crew coped with the situation, or solved the problem. Once the questionnaire is returned to the HF group in Safety Services the report is entered into the BASIS HFR database and analysed using the information from the related ASR when available. Further information about the event may be collected through the process of 'callback' in which the reporter is telephoned by the analyst to confirm the understanding of the incident and to elicit more information where possible. Most likely it will only be by talking to the reporter that any ambiguities can be eliminated before the real analytic process begins.

Each report is subjected to an analytic process that determines the underlying structure of the event. This provides an abstracted description of the event defined by a set of 'Factors' concerning 'Crew Actions' and 'Influences' on those actions. The factors can be assigned in a positive sense, i.e., a safety enhancing sense, or in a negative - safety degrading - sense. Once these Factors are identified they are linked together to create an 'Event Sequence Diagram' (ESD) illustrating the flow of cause and effect throughout the incident.

Factor assignment: There are four groups or categories of factors. The first category is concerned with observable / describable crew behaviour and actions that can be defined as safe or unsafe. Three further categories are devoted to different kinds of influences on crew behaviour. These four categories are briefly described below.

Crew Actions: These are of three distinct types. One type concerns the activities of handling the aircraft and its systems, e.g., 'System Handling'. A second concerns the potential error types reflecting the Reason model of human error (Reason, 1990), e.g., 'Action Slip'. Third is the largest set of factors that are derived from Bob Helmreich's NASA CRM Teamskills (Helmreich, Butler, Taggart & Wilhelm, 1995). These describe a number of activities involved in the safe management of flight, e.g., 'Workload Management'.

Personal Influences: These describe the subjective feelings of physical and mental well-being, emotion, stress, motivation, and attention as described by the reporter. Examples are 'Boredom', 'Personal Stress', 'Tiredness' and 'Mode Awareness'.

Organisational / Informational Influences: These are influences that are directly controlled by the company. Examples are 'Training', 'Technical Support', 'Standard Operating Procedures', 'Navigational Charts'.

Environmental Influences: The final group is composed of those influences over which neither the reporter nor the company has any control. Examples are 'ATC Services', 'Technical Failure' and 'Other Aircraft'.

Crew actions differ from the influences in that they are generally observable and reportable. The majority of the influence factors are not so easily determined. In some cases the influences (and in some cases the actions) can be inferred but it is essential that the inference is based only on the evidence from the ASR, the HFR and callback, not on the analyst's belief about what should or might have happened. Whilst the analyst is expected to have substantial experience of the reporter's aircraft type and operation, this is to help him or her understand the reported information, NOT to allow them to 'put themselves in the reporter's position'. What is required is the application of a common language to encode the reporter's description of the incident.

Factor Linking & 'Operational Problems': Event Sequence Diagrams (ESD) are created for each report in a graphics page in the HFR database. The software allows an interactive trial and error approach to facilitate what can be a difficult task.
The first part of ESD construction is establishing a mental model of the reported event. This helps define exactly what the core problem was. Having established the identity of the problem, the analyst can focus in turn on the causes of the problem and then how the problem was solved, if it was! To help in this process a 'meta-factor', 'Operational Problem', is used and is defined as:

> *'Any situation or event that threatens or could potentially threaten the safety of the aircraft or any of its occupants. An Operational Problem will require the crew to consider the implications of the event and if necessary to act to eliminate or control the threat.'*

The Ops Problem factor helps parse the event into the causes and resolution of the event. Figure 3, below, shows the event structure in the simplest possible case. The arrows are the causal links between the factors. It is important to note that they are intended to indicate the direction of cause or influence, not just chronological relationships.

Fig 3: Basic HF Event Sequence Diagram

Thinking of the event in these simple terms helps the analyst understand the event and the content of the HFR. This is even truer when the reported event contains more than one Ops Problem. In the latter case a methodical approach is essential and the software allows the placement of each Ops Problem on the analysis screen with a brief description of each Problem in a text box associated with the meta-factor.
Finally, it is worth noting the variety of ESD structures. Even with only one Ops Problem the variety of shapes and sizes that the ESD can take is enormous. The structure of an incident can rarely be represented by a single continuous sequence or chain. Some factors might have several causes or precedents and might directly influence more than one other. Thus, we will generally see incident chains with sections that branch outwards or converge. Divergence is less likely than convergence and is most likely to be found in the immediate consequences of an Ops Problem. For instance, after an engine failure there would be at least two distinct sequences. One would describe the handling of the shutdown checklist and another the liaison with ATC negotiating a diversion. Both are consequential on the Ops Problem but they may not necessarily interact causally. Complexity will be multiplied when more than one Ops Problem occurs in the same report. Simplicity often results from the paucity of the information supplied by the reporter. Alternatively, it can result because events constrain the courses of action available to the crew, thus reducing the extent of the ESD subsequent to the Ops Problem.

Human Factors Database Analysis: The process of report analysis described above provides the data for the next stage of analysis. BASIS software facilitates exploration of the database through filtering and trending processes in

a similar way to that used in the ASR modules. This allows both simple and complex comparisons to be made between different parts of the database, e.g., 'Are the Abnormal Checklists more troublesome on the Boeing 777 than on the Boeing 767?' or 'Is situation awareness better on modern aircraft than on older generation aircraft?'. As the causal links between factors is also stored in the database we can also ask questions such as 'What causes problems for flight crew?' and 'What do they do about them?'. Once the time consuming work of report analysis is complete the database analysis can be undertaken with a few key-presses and takes only a few seconds. Only one example is given here, that of a comparative analysis of situation awareness (SA) in the modern 'Glass Cockpit' and conventional, older 'Steam Driven' fleets.

Situation Awareness

The HFR analysis scheme encodes SA with three specific factors in the Personal Influences. These are Environmental Awareness, Mode Awareness and Systems Awareness (the definitions are presented below in Table 1). Environmental Awareness includes aspects of awareness concerning the world outside the aircraft. This includes geographical position and terrain features; awareness of other aircraft and their communication with ATC; and of course meteorological conditions. Mode Awareness relates to the configuration and the flight path of the aircraft. It include Flight Management System modes; Altitude, Attitude, Airspeed, etc, as well as the rates of change of these parameters; and the energy related aspects of the flight path. System Awareness relates to the operational status of the aircraft's engineering subsystems.

One or more of these factors were thought to have been involved in a disquietingly high number of events that had been causing management and pilots some concern. For instance: 1) the occurrence rates of rushed approaches were roughly equal in the Glass and Steam fleets; 2) the rate at which serious Ground Proximity Warning System warning occurred was significantly greater in the Glass fleets; and, 3) whilst more navigation errors occurred in the Steam fleets they were normally picked up by the crew immediately, whereas on the Glass fleets they more frequently resulted in flight path deviations.

The interesting aspect of these three kinds of incidents is that they all involved a loss of situation awareness. Given the nature and relative frequency of these incidents on the Glass and Steam fleets the data seemed to suggest that situation awareness on the Glass fleets had not benefited from the new technology.

Environment Awareness Crew awareness of environment, e.g., other aircraft, communications between ATC and other aircraft, met conditions, geographical position, terrain features and MSA.
Mode Awareness Crew awareness of aircraft configuration, flight and powerplant parameters, flight control system modes, and the dynamic aspects of all of these.
System Awareness Crew awareness of the state of the aircraft's technical systems, e.g., fuel, electric, hydraulic, air.

Table 1 - Definitions of the three Situation Awareness Factors

However, as the FDR programme is effectively an anonymous programme, feedback on the causes of incidents is very scarce. The ASR is somewhat better off in this respect but it is an 'open' programme and reporters tend to offer factual descriptive accounts of incidents and very rarely describe factors such as a loss of situation awareness. Consequently, as the HF programme had been introduced to overcome just this kind of difficulty, Flight Operations turned to the HFR programme for data and analysis.

Situation Awareness Study: One of the factors mentioned above, Mode Awareness, is of particular interest here as, along with Environment Awareness and System Awareness, it is one of the three Personal Influences that are used to encode SA. Personal factors are different from all the others in that they are 'internal' or 'subjective' and in all but a few exceptional cases have to be described in the report or during a callback session with the reporter. In assigning a Personal factor we make no inferences of the 'it must have been' variety. The three factors combine to offer a sufficient representation of good or bad SA as reported in an incident

In this study we were interested only in the relative frequency of positive and negative assignment of the SA factors in the modern (Glass) and older (Steam) fleets. The Glass fleets consist of A320, B737-400, B747-400 and

B757/67/77s. Steam fleets include B737-200, B747-100/200, Concorde and DC10. This study uses data collected since January 1997 and within this period the number of reports was approximately 3700 with over two-thirds from the Glass fleets. From these reports analysts had assigned just 396 SA factors. Unfortunately, it is unlikely that this study will be updated as, in common with most other airlines, the number of 'Steam' aircraft still operating will be too small to make any useful comparisons.

Analysis: Despite having 3700 reports the majority did not report SA directly**.** A total of 396 SA factors were assigned in the analyses of which approximately one quarter was from the Steam fleets. In view of this, no fleet by fleet comparisons were attempted, only between the Glass and Steam groups.

The comparison method is also problematic. Direct comparison of assignment frequency is not possible because of the different sizes of the groups and the different reporting frequency of the factors. Therefore, a measure was calculated by taking, for each of the two groups and the three factors, the number of positive assignments as a percentage of the total number of assignments (positive plus negative) for that group / factor combination. Figure 4, below, shows the percentage positive assignment for the three factors, and a composite of all three, for the Glass and Steam groups separately.

It is clear from Figure 4 that this study offers no evidence that Glass cockpits enhance SA. Overall SA for Steam indicates 79% positive assignments compared to 60% for Glass. That is, 79% of the Steam SA factors were coded positively with 21% negative. For Glass the corresponding figures were 60% positive and 40% negative. At this point it is worth reminding the reader that the HF questionnaires are invariably completed after some problem has occurred. The above figures (hopefully) do not indicate the state of SA during majority of flights in which no reportable problem has occurred. It is, however, quite likely to be indicative of the kinds of issues arising in, or causing, our incidents.

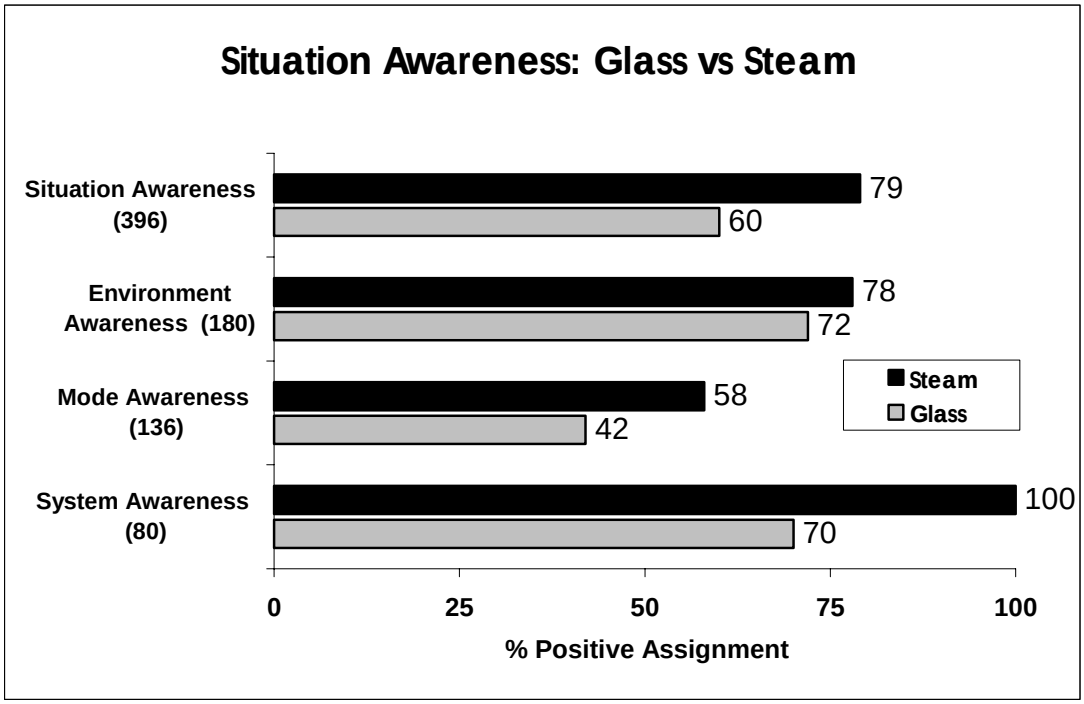


Figure 4: Aspects of Situation Awareness for Glass and Steam aircraft Fleets

The three sub-factors confirmed the above finding individually. Environment awareness was only slightly superior in the Steam fleets but given the benefit of the navigational and environmental displays of the Glass cockpit, the slight difference is not necessarily good news for Glass. The figure for Systems Awareness in the Glass group is relatively high, however it is particularly impressive for Steam with no negative assignments at all for this factor in the group. Here, perhaps, we are seeing the benefit in four out of the five Steam fleets the inclusion of a Flight Engineer as part of the crew.

Finally the Mode awareness scores were high in neither group. However, even here, the Steam fleets do substantially better than the Glass fleets, with scores of 58% and 42% positive respectively. Whilst the immediate

instinct may be to blame the usual problems of hidden FMS mode changes in the modern flight deck, a look at the definition of Mode awareness above will remind the reader that this factor is not solely concerned with FMS modes.
The results of this study suggest that modern technology has not delivered its promise of enhanced SA in comparison with the pre-modern technology. All three SA factors were apparently more effective in the Steam group, with the finding of an overall rate of positive assignment nearly 20% greater than for Glass. Although the majority of the factors were encoded positively in both groups, it is clear that the Steam group consistently reported better SA than the Glass group.
Whilst this data and analysis is intrinsically interesting the reason for its inclusion in this paper is that it is an excellent example of how the HFR programme can address complex cognitive and behavioural problems that are beyond the reach of either the ASR or the FDR programmes. Whilst both of the latter programmes give an indication that there was (is?) a problem, the analysis available from the HFR data goes much deeper. (In particular, a similar analysis of relevant 'Crew Actions' showed mush the same picture adding further support for the above findings.) The data gathered from the HF questionnaire is much richer in detail than in the ASR reports and the more sophisticated coding structure allows us to inspect some more complex factors such as situation awareness. Moreover, the benefit of using both positive and negative factors, absent from the ASR and FDR programme, gives us sounder comparisons between fleets.

Risk and Severity
Naturally, an important strategy for any safety reporting and analysis programme is the assessment of the risk attributable to each incident. A responsible organisation deals with its problems in the order of descending risk. This, of course, presupposes that the organisation has a risk assessment process. The programmes described above, the ASR, FDR and HFR, have, respectively, a risk assessment process, a severity assessment process, and for the HFR programme neither risk nor severity assessment. These will be briefly described in that order.

ASR Risk Assessment: Figure 5 below shows part of the BASIS ASR 'details' screen. It is in edit mode and the small window shows the selection screen for the analyst's risk assessment.
Risk is assessed on a three by three Severity by 'Likelihood of Recurrence' matrix. A Risk shorthand can be seen in the matrix with 'A' indicating the highest risk and 'E' the lowest. The matrix is pragmatic rather than theory based. Particularly when events are coded at the highest risk level, probabilities and severities merge into a practitioner's instinct rather than cold calculation.

FDR Severity Assessment: The FDR analysis differs in many ways from the ASR. It does not estimate the likelihood of recurrence preferring instead to concentrate solely on event severity. A panel of domain experts has assigned a threshold severity value for each of the approximately sixty event types. The experts were mostly highly experienced training or technical flight crew and the process of severity assignment was iterative so that the severity accorded to any recorded event would depend on how much beyond the threshold the actual event strayed. For instance, an altitude deviation would only be recorded as an event if the deviation was more than 400 feet. At that level the event would be assigned the minimum value assignable to the event, say a value of 50. However, if the deviation were greater than 400 feet the value assigned would itself be greater. Whether this was a linear increase or some exponential function would depend on the specific event type.

HFR Risk assessment: At present there is no HFR risk or severity assessment for the HFR programme. However, my colleague, Carl Macrae, is undertaking an intensive study of how we deal with risk in Safety Services and Flight Operations. This is discussed in the next paper in this volume.

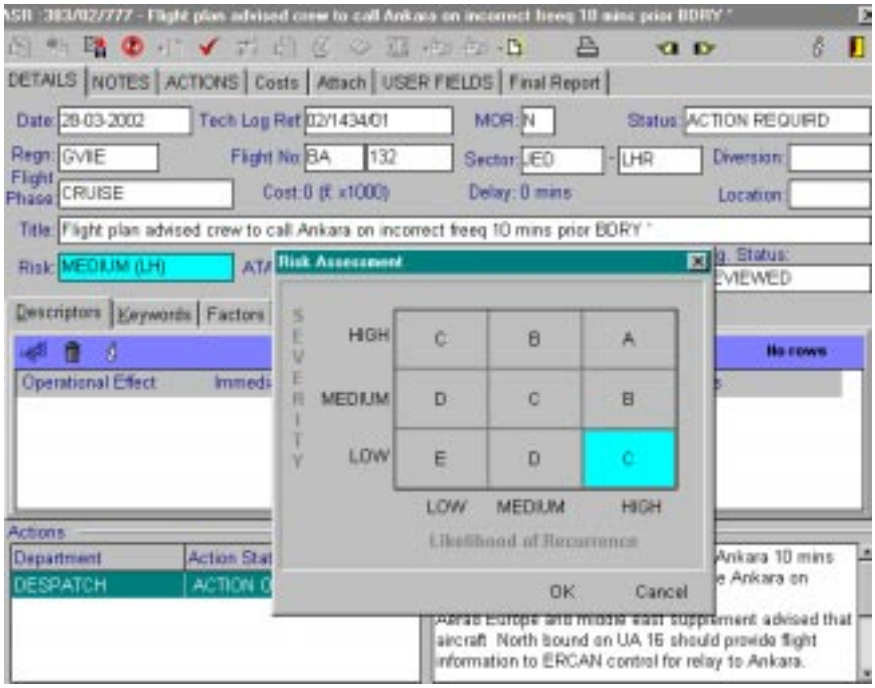


Figure 5: ASR Risk assessment

Summary

Safety management requires many skills and many tools. The diversity of the problems faced by safety practitioners must be parried by a diversity of aids. From the initial launch of BASIS, in the guise of the ASR programme, to the multifunction instrument it is now, see Figure 1, it has coped with the demands of relatively simple Audit programmes, to the highly technical flight data programmes and to the demanding but rewarding human factors programme. Each of these has its particular function aimed at a particular challenge.

BASIS development has responded to many challenges from within our organisation as well as from the environment without. The requirement to provide snapshots of our safety state for senior management is counterbalanced by the need to develop more sophisticated tools to face the more complex environment in which aviation exists and develops. I hope we continue to balance that equation.

References

Helmreich, R. L., Butler, R. A., Taggart, W. R., Wilhelm, J. A. (1995). *Behavioral Markers in accidents and incidents: Reference List*. NASA/University of Texas/FAA Aerospace Crew Research Project. Technical Report 95-1, March, 1995.

O'Leary, M. J., and Chappell, S. L. (1996). Confidential incident reporting systems create vital awareness of safety problems. ICAO Journal, 51: 11-13.

Reason, J. R. (1990). Human Error. Cambridge University Press.

Appendix A

Comparison of organisational structure and process for the FDR, ASR and HFR programmes

	SESMA	ASR	HFR
Responsible department.	'Owned' and operated by Flight Operations in co-operation with BALPA.	'Owned' and operated by the Safety department.	'Owned' and operated by the Safety department.
Data public?	Anonymous.	Open to Management.	Confidential to HF Group in Safety department.
Data format.	Data collection is electronic.	Data collection is by receipt of handwritten report or telex.	Data collection is by receipt of handwritten report (usually after receipt of ASR reply).
Reporting requirements.	Reports are automatic.	Reports are mandatory.	Reports are voluntary.
Report receipt.	Within 24hrs.	Within 24hrs.	Between 24hrs – 2 mths.
Initial data analysis.	Within 24 hrs.	Within 24 hrs of receipt.	Within one month.
Cross program report comparison?	When available, information from ASR used to assist in FDR analysis.	Information from FDR can be used in the analysis of related ASRs. All information available to FDR and HFR programmes.	When available, information from ASR is used to assist in HFR analysis.
Risk assessment.	Within two weeks.	Within two days.	No formal risk assessment but long term 'issue' discovery.
Management action.	Generally within a month but can be 'fast-tracked' if high risk.	Depends on assessed risk – can be immediate.	When analysts establish an issue. – can be many months.

Assessing the Risk of Flight Safety Incident Reports

Carl Macrae (1), Nick Pidgeon (1) and Mike O'Leary (2),

(1) Centre for Environmental Risk, University of East Anglia, Norwich NR4 7TJ, UK
Email: c.macrae@uea.ac.uk

(2) British Airways Safety Services, Compass Centre, Heathrow, London TW6 2JA, UK.

Abstract: The analysis and risk assessment of flight safety incident reports within the British Airways Safety Information System (BASIS) relies on the expert judgement of trained investigators. This paper describes our research that aims to develop a methodology, and corresponding set of risk indicators, to support and extend risk assessment of flight safety incidents. In this paper, three issues will be addressed. First, the application of organisational theories of major accidents to incident analysis. Second, the investigation of incident report analysis in its organisational context as an active process of both sensemaking and safety management. Third, the nature and place of expertise and professional judgement in relation to these former considerations of theory and practice. These issues are discussed as a basis for extending the risk assessment of incident reports, specifically by its explicit integration throughout the entire process of incident report analysis and management.

Keywords: risk assessment, organisational accidents, safety management

Introduction

This paper outlines ongoing research investigating the practical analysis, interpretation and risk assessment of flight safety incident reports in the British Airways Safety Information System (BASIS). The project aims to develop a methodology, and corresponding set of indicators, for the risk assessment of flight safety incident reports.

In dealing here with risks to flight safety we are addressing an operational area in which major adverse outcomes are thankfully rare. If they do occur, they are invariably complex, unique and catastrophic, involving massive and distributed damage as well as social, economic and environmental loss. Investigating and managing safety incidents is one means by which to pre-empt potentially catastrophic ends.

Maximising the efficacy and utility of safety incident management programmes depends on three fundamental cornerstones. These are the application of appropriate models of accident causation, the rapid learning of suitable lessons from near-miss reports and the assessment of risk implied by incidents in relation to safe levels of operation.

The application of appropriate models of accident causation allows organisations to identify relevant aspects of their operations to monitor and to guide incident investigation. Theoretical advances in modelling accident causation have produced frameworks within which to understand adverse events in their organisational context (e.g. Turner and Pidgeon, 1997; Reason, 1997). Importantly, such theories mirror, and help to drive, the development of industrial practice.

In the best of all possible worlds, near-miss reporting allows pertinent and detailed lessons to be learnt rapidly from each brush with danger that the organisation encounters. Both the institutional arrangements surrounding reporting, and the nature of information sought, need careful consideration to achieve as close an approximation as possible to this ideal situation.

Finally, in order to maintain safety, the risk of near-miss incidents—and their causal factors—needs reliable and valid assessment. Further, acceptable risk needs definition and incidents and their implications tested against these criteria. To date, little work has addressed this issue in the organisational and practical context of incident reporting.

Each of these aspects, and their implications, will be discussed. First, two theories of major accidents will be described. Next, the aims and function of incident reporting and risk management will be outlined. Then, incident analysis in British Airways will be discussed, specifically in light of expert judgement and current practice. Finally, implications for assessing the risk of incidents will be considered with regard to theories of accidents and the practice of safety management.

Theories of Major Accidents

Man-Made Disasters: Man-Made Disasters theory (Turner, 1978; Turner and Pidgeon, 1997) provided the first rigorous conception of the causes of major accidents. Rather than bolts from the blue, disasters were conceived as emerging from the social and technical arrangements of organisations charged with managing major hazards. The theory centres on processes of poor management and the mishandling of warning signals.

Disaster-prone organisations were found to be working with incomplete information, poor communication or complacent work practices. Beliefs and perceptions of the organisational operations, environments or vulnerabilities were incorrect, inaccurate, blinkered or resistant to change. Warning signs suggesting the development of a hazardous situation were missed, misunderstood, miscommunicated or minimised. These lead to "the management system losing touch with its operational realities" (Turner, 1994, p. 216). As a result, safety systems degraded in relation to their hazards, and discrepant incidents accumulated unnoticed.

The development of a situation at odds with accepted understandings was termed a disaster's *incubation period*. These were often nurtured by the nature of organisational tasks, involving ill-structured, vague, uncertain and complex problems. Incubation ends when a *precipitating event* forces revisions to the previous misunderstandings and erroneous beliefs. This often brings with it a considerable amount of surprise. In conjunction with, or soon after, this event disaster *onset* occurs followed by *rescue and salvage* efforts. Finally, through inquiry and assessment, sense is made of the events. Understandings of operations, hazards and precautions are adjusted. Ideally this leads to changes in policies, procedures and practice.

Turner's (1978) theory explains the development of disaster with concepts of communication networks, decision making hierarchies and organisational cultures. Formal and informal networks within organisations structure patterns of communication. Shared values, norms and assumptions—or organisational cultures and subcultures—constrain the content of communication and form the basis of understanding and decision making. They guide what information is sought, considered and acted on. Equally, they guide what is ignored and dismissed (Pidgeon, 1998; Weick *et al*., 1999).

Organisational Accidents: Reason (1990; 1995; 1997) developed a theory of *organisational accidents* in complex sociotechnical systems. Organisational accidents involve the catastrophic breakdown of large scale industrial systems with severe social, economic or environmental consequences. They have multiple causes involving many different people spanning many echelons of the organisation—and beyond—from shop-floor operators to management to regulators.

High-consequence sociotechnical systems are conceived as possessing multiple layers of defences and safeguards to control and mitigate hazards. These defences-in-depth may be 'hard', as in physical barriers, or 'soft', in the form of regulations or safe operating procedures. However, such defences will always be partial and incomplete. Reason (1997) outlines the organisational context of defensive failures, proposing two general pathways by which defences are breached.

First, by organisational factors which are generated by actors distant—in both time and space—from the hazards. These *latent conditions* (or the more pithy *resident pathogens*) may lead to weaknesses or holes in system defences due to the unforeseen and adverse consequences of management decisions or organisational processes. Second, organisational factors may also create local conditions that provoke unsafe acts of those with their 'hands-on' the system operations. These *active failures* impact defences directly and may provoke immediate adverse consequences. Ultimately, if active failures coincide with defensive weaknesses in necessary combination, a major accident will result.

To manage safety, Reason (1997) proposes a systemic approach. Practically, for any particular accident, unsafe acts and defensive failures are likely to be both unique and unpredictable in their combination. At this level they are therefore hard—or even impossible—to predict, micro-manage or eliminate. Instead, the model identifies generalisable organisational and local factors provoking defensive failure and unsafe acts. This allows the identification of underlying causes open to remedial management *before* a major accident occurs. Moreover, systemic weaknesses may contribute to a broad range of accidents, being common to many of an organisation's operations. Analysis of minor safety incidents to infer such underlying factors provides a window into organisation-wide safety health (Reason, 1997).

Safety Management and Information Systems

At the heart of managing organisational safety is maintaining awareness and control of operations and environments. Accurate, relevant, information needs to be gained, suitably interpreted and the ensuing lessons

acted on. Moreover, identified hazards need to be prioritised to ensure maximum effort is focused on the greatest threat.

Incident Reporting Programmes: Near-miss incident reporting programmes are an effective safety management tool. Pragmatically, near-misses are, "any event that could have had bad consequences, but did not" (Reason, 1997, p. 118). The analysis of momentary lapses of safety allows both the discovery of ways sociotechnical systems may fail and the monitoring of known hazards (van der Schaaf *et al*., 1991). In practice, however, it's often hard to distinguish these as many incidents are unique, either in their causal factors, failure mode or combination of occurrence. For this reason, the management of individual incidents may involve varying degrees of monitoring and discovery; assessment and action will equally constitute a mix of acting on previous experience and original creativity.

Usefully, incident reporting programmes have been conceptualised by seven functional steps (van der Schaaf *et al*., 1991):

1. Detection and reporting
2. Selection for deeper analysis
3. Detailed description and deeper investigation
4. Classification of causes
5. Recognition of patterns and priorities
6. Interpretation for recommendations
7. Implementation and monitoring

The stages of particular importance to this paper are (2) the selection for deeper analysis, (3) description and investigation, and (5) the recognition of patterns and priorities. These will be discussed later in relation to theories of major accidents and risk assessment.

Risk Assessment and Management: Risk assessment is a broad church. Fundamentally, it represents processes of rational appraisal and decision making regarding the likelihood that adverse events will come to fruition. The basis for this are probabilistic measures of the *occurrence* of events combined with measures of the *severity* of those events' consequences (Royal Society, 1992). Essentially, risk management involves prioritising action according to the greatest risk; and ensuring that current levels of risk entailed by activities are acceptable.

Formal quantitative tools applicable to the high-hazard industries have reached an advanced state of development in such forms as Probabilistic Safety Assessment (PSA) and Human Reliability Assessment (HRA) (e.g. Kirwan, 1994). However, in relation to the practical context of incident report analysis, there are several challenges to their successful application. These are attributable to the dynamic, ill-structured nature of the problems to be managed, the unique catastrophic potential flight safety incidents represent and the organisational context in which these tools invariably have to be implemented.

First, disasters and accidents may only look predictable in hindsight (Turner and Pidgeon, 1997; Reason, 1990). Almost certainly, the people acting at the time are working with a range of, "dynamic situations that consist of complex systems of changing problems that interact with each other"; Ackoff (1979), "call[s] such situations messes." As suggested previously, safety incidents, by their nature, are likely to be unique—if not in their substance then in their structure. Their reporting and the information they contain constitutes a source of weak early warning signals. The selection and subsequent management of such signals is notoriously difficult (Turner and Pidgeon, 1997). Moreover, in a management context there exists the practical imperative to take action. This will often be without the full weight of conclusive evidence. Such certainty only presents itself in the unacceptable form of a bad outcome.

Second, we are concerned here with issues of flight safety. That is, major organisational accidents in well defended systems. By definition, the most relevant incidents to managing these risks will always be towards the high severity, low probability end of the scale. However, quantitative analysis of such small probabilities is often very difficult. As such, extreme event analysis may require extensions to current risk assessment methodologies. What is more, there still remain considerable modeling challenges to be met in capturing the actions embedded in organisational context. These include the possible importance and mechanisms of culture (Pidgeon, 1998), and the means by which latent organisational and management factors shape outcomes at the 'sharp-end' of operations (e.g. Reason, 1997).

Finally, it is of considerable importance that any assessment process is practicable. Not only does a process need to address the assessment problem, but it must meet time and resource constraints; it must be simple enough to be implemented and understood; it must suit and draw on the appropriate expertise; it must

accommodate existing institutional arrangements; and it must be politically acceptable—"recommendations must be sold as well as generated" (Fischhoff *et al*., 1981).

Incident Analysis and Assessment

The British Airways Air Safety Report (ASR) programme is typical of many such programmes operating in safety-critical industries. Here, we will use it as a model around which to frame the discussion of risk assessment. Analysis and assessment of reports is principally reliant on trained investigators' professional judgement. The nature of such judgement has been explored from cognitive, social and organisational perspectives. These findings should help to inform the future shape of appropriate risk assessment methodologies or support tools. First, however, it is worth outlining the current role of professional judgement in assessing ASRs.

Analysis of Air Safety Reports in British Airways: Flight crew are required to report any event which had, or could have had, safety implications. Approximately one hundred and fifty ASRs are submitted per week across all fleets. Once raised, the original reports are converted to electronic format and entered in to BASIS. Professional judgement determines the salient features of each incident, the likely causal factors and the risk to flight safety and operational integrity that each event represents. There are four general task steps.

First, the report's text input and technical details are appraised and reviewed. A brief executive summary of one hundred words or less is produced. This identifies the important elements of the incident, and characterises their relationships and the inferred causal structure of the event. The investigation stops when the investigator is satisfied that the explanation is complete.

Second, suitable keywords are applied to allow subsequent analysis of the incident database. Keywords can specify cause, effect or influencing factor at four levels of description. The first level defines a Major Category for the keyword string; either environmental, operational or technical. Within this major category a second level Primary BASIS Reference is then allocated, such as Security or Stall Warning. There then remain two further levels for the optional specification of First and Second Keywords to provide greater detail (Figure 1).

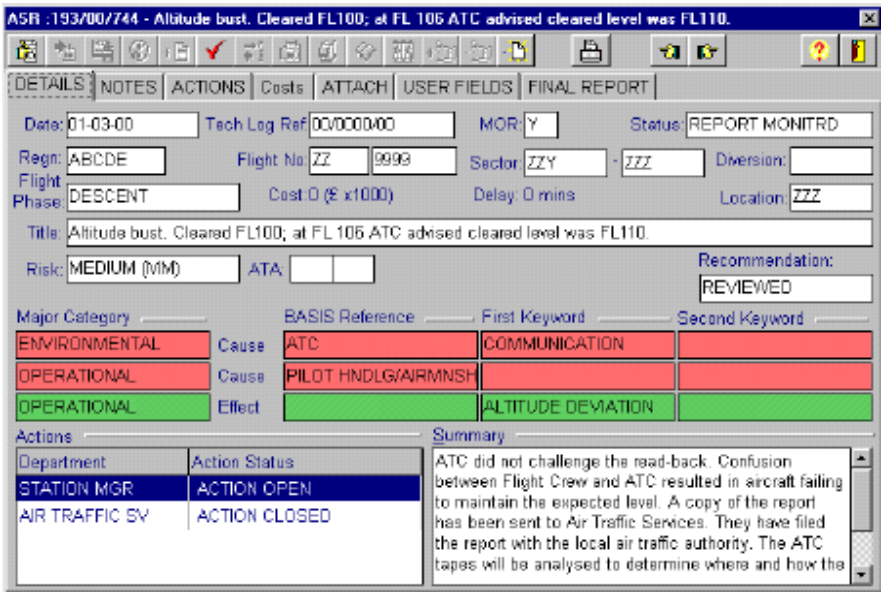


Figure 1 - Example of an Air Safety Report entry in BASIS.

Third, a risk category is assigned to each incident. Explicitly, this allows historical trend analysis and the prioritisation of response to incident reports. Implicitly, it embodies the consideration of each incident's potential impact on flight safety and, more broadly, BA's business as a whole. Currently, assessment is made according to a three by three matrix of *severity* (high, medium or low) by *probability of recurrence* (high, medium or low). This provides an ordinal five point risk scale, A to E (Figure 2). Formally, levels of risk are

primarily defined by company response to the incident (Figure 3), and allocation relies on the expert judgement of safety officers and investigators.
Fourth, if actions are required (for instance, further investigation), they assigned within BASIS, delegated to line management and monitored.

Severity	Probability of Recurrence		
	Low	*Medium*	*High*
High	C	B	A (Severe)
Medium	D	C	B (High)
Low	E (Minimal)	D (Low)	C (Medium)

Figure 2 - BASIS Risk Matrix

Risk	Company Response
A	AAIB or state investigation
B	AAIB or Safety Services investigation
C	Departmental action required
D	No action required by Safety Services
E	Statistical data

Figure 3 - Risk Definitions by Level of Company Response

Expertise and Judgement: The role of professional judgement in risk assessment and decision making has received considerable attention by researchers from a number of disciplines. This is perhaps no surprise in the light of its predominance, as exemplified by ASR management in British Airways in a range of real-world risk management contexts. Previous work in this area have investigated cognitive, social and organisational aspects of judgment.
There is a long tradition in modelling the various statistical and logical pitfalls of the human cognitive system. For instance, very small probabilities are systematically under-rated or entirely discounted (Kahneman and Tversky, 1979) and causes not elaborated in experts models of events receive scant attention (Fischhoff *et al*, 1978). Further, recent work has suggested that , not only do experts from different professions bring different substantive knowledge to a situation, but they may also differ more fundamentally with regard to their conceptions of causality. Gherardi and colleagues (1998) suggested, among other things, that engineers' explanations of safety incidents involved more linear, 'billiard-ball' notions of causality than managers, who viewed situations as a combination of multiple co-occurring events, each necessary but not sufficient.
Equally, experts may posses different tacit models of accident and human error. For example, Lucas (1990) sketched three general models. Those based on traditional occupational safety management may focus on the individual worker, emphasising motivation and coercion campaigns to avoid accidents. Models based on traditional risk management focus on engineering and the man-machine interface to improve safety. Finally, the systems safety approach encompasses organisational and situational induced accident concepts. Such different models may influence risk management, the focus of incident investigation and the nature of risk reduction (van Vuuren, 2000).
Socially, groups of individuals with similar views often over-confidently endorse decisions far more extreme that each individuals' original position (Janis, 1982). Organisationally, the processes by which incorrect assumptions become sanctioned and embedded in understandings taken for granted, and signals of potential danger become expected signals of safety and so accepted as normal, or normalised (Vaughan, 1996), are complex and poorly understood from a risk management perspective. Informing risk assessment and decision support tools directed at the ongoing management of safety incident reports is likely to be of considerable importance.
Notwithstanding the rather negative view of expertise in the foregoing discussion, the application of professional judgement is most certainly not all bad. First, expert judgement is—by definition—often the best we have, and "professional produce answers" in a world where risks have to be managed today (Fishhoff *et al*,

1981). Clearly, they possess rich substantive knowledge to draw from. Moreover, expert decision makers seem to develop special strategies to meet their targets. More often than not, they avoid large mistakes even if they make minor errors (e.g. Shanteau, 1988). Formalising such practical decision rules for the assessment of safety incidents may be a useful approach.

From preliminary work observing and talking to a range of expert assessors at British Airways, two interesting points have immediately stood out. First, incident reports seem to be framed by experts 'catastrophically'. Incidents are frequently interpreted in light of historically severe incidents or accidents which had similar features or consequences, not only from BA's own operational experience but from other operators, press reports or safety bulletins. Second, safety investigators often emphasised the dynamic nature of the assessments in light of current operational safety concerns. The assessors' evaluation of the risk of individual incidents or their underlying factors appeared to be heavily weighted by the recency and frequency of similar events, or current topics of interest each analyst was keeping in mind. Perhaps unsurprisingly, these two features bear direct relevance to the scales of frequency and severity underlying risk assessment matrices. However, the mechanisms and drivers underlying these assessments, not to mention their functionality, would bear closer scrutiny.

Rethinking Risk: Questions of Practice and Theory

The theory and practice of organisational safety management suggest various opportunities to improve the potential of risk assessing flight safety incident reports. First, and generally, the assessment and management of risk provides a framework for the whole incident management process. Second, supporting experts' investigative models and informing the assessment and analysis process with insights from theories of organisational disaster may provide avenues to turn theories of accident into ones of practical safety. Third, criteria of acceptable risk for safety incidents—or perhaps at best, what drives and underlies them—could do with clearer definition.

Practically, it seems that assessments of risk—often tacit and not explicitly stated—tie three central elements of incident safety management together. Risk is used as a pragmatic guide to prioritise management action; this is its explicit function. However, it also seems to guide the level of detail and extent of incident investigation and analysis. Incidents considered of greater threat, either to the safety of the system or to understandings of the safety of the system, would sensibly seem to attract more attention. Further, risk would seem to guide investigators' attention and awareness towards issues of current concern. Investigators do not approach problems as a *tabula rasa*, but rather acknowledge they have several favoured ongoing failure types or issues which they monitor and search for. Making such assumptions more explicit throughout the process should increase the transparency, and so the clarity, of safety management.

Accident theories make clear that much of the real worth of incident reporting comes from the interpretation and understanding of the organisational processes underlying unsafe events. Extensive subject matter expertise underpins professional judgement on these issues. By definition, experts' educated intuitions are better than anyone else's (Fishhoff *et al*., 1981). However, judgements relate to issues that are often empirically unresolvable in practice (Fischhoff, 1989). Instead, models derived from accident analysis provide verified frameworks of generic causal factors. Integrating professional and theoretical models should combine the benefits of both conceptually driven and contextually specific models for the analysis of incidents. Such work has begun in the medical domain (e.g. Vincent *et al*., 2000). Further, Reason (1997) suggests that the severity of individual incidents does not adequately reflect the underlying safety health of an organisation. If so, then the risk of safety incidents should not be directly tied to their severity.

Further, explicit consideration of organisational accident models may reveal a fundamental dilemma. On the one hand the majority of events highlighted by incident reports are likely to be of *low-consequence*. Such incidents will result from well understood mechanisms, each involving perhaps one or two failures and causal factors, in situations where safety defences proved adequate. Such events are most likely to comprise the majority—if not all—of the incidents which cross an air safety investigators' desk in any one week. They should be readily resolvable, either through the accumulated expertise of professional judgement and learning, or through formal risk assessment methodologies. Indeed, the risk matrix shown on Figure 2 is at its most helpful under these routine circumstances.

On the other hand, theories of major accidents suggest that the most important *high-consequence* issues will be unique in their complexity, if not their individual components, and involve many causal factors distributed widely in time and space. These circumstances present particular difficulties of identifying and integrating in foresight the true warning signals—that are always so clear in hindsight—from the available mass of

information. We would argue that such circumstances will be far less amenable to either the routine application of learned expertise or to formal methods of risk analysis.

Accordingly, in high-hazard, complex and well defended industries such as civil aviation, effective safety management demands analysis that accommodates the structured with the ill-structured, the novel with the routine and the familiar with the surprising. Overly rigid problem solving and assumptions regarding organisational safety can lead to the discounting of unexpected and poorly understood information and the normalisation of warning signs (Turner and Pidgeon, 1997; Vaughan, 1996). These tendencies arise from routine and normative organisational behaviour. As such, methodologies are needed to challenge these processes directly. Potentially, this could be achieved by explicitly addressing notions of surprise in the risk assessment process, as a means to recognise opportunities for learning. It could similarly involve informing analysis with practical concepts of *safety imagination* (Pidgeon and O'Leary, 2000) or *collective mindfulness* (Weick *et al*., 1999). Prescriptions of attempting to fear the worst, the generation of 'what if' scenarios and avoiding simplifying assumptions, facilitate the discovery and interpretation of weak signals. In the management of ill-structured, complex problems the creative process of risk assessment often provides as much benefit as the final output. Questions of *how* such beneficial processes could be integrated into incident report management systems remain largely unresolved.

Finally, levels of acceptable risk in relation to assessing incident reports could be more clearly defined. The definition by British Airways of risk by company response (see Figure 3) is perhaps novel only in that it explicitly acknowledges the often undisclosed assumption that judgements of acceptability are bounded by organisational norms, policy and procedure. Nonetheless, "[t]he defense of having adhered to accepted practice only transfers the responsibility to the judgement of others," (Fischhoff *et al*., 1981). Levels of *acceptable* risk can't be predetermined from theoretical or empirical models of accident causation. Risk acceptability, as contingent on many competing demands, is instead an emergent property of the decision process (Fischhoff *et al*., 1981) than a predetermined standard.

The conceptual issues discussed above outline a research agenda addressing three fundamental questions:

1. How should reporting programmes best provide pertinent information to manage the risks of organisational accidents?
2. How can organisational models of major accidents inform, and be integrated with, expert models and understandings of safety and risk?
3. How can risk assessment actively support the flexibility of process needed to deal with ill-structured, dynamic and ambiguous safety information?

The means by which the dynamic control of organisational safety can be achieved through safety incident assessment would benefit from more detailed practical and theoretical specification of both the organisational context driving the risks and the organisational and social context within which the analysis proceeds.

References

Ackoff, R. L. (1979). The Future of Operational Research is Past. Journal of the Operational Research Society, 30:93-104.

Fischhoff, B. (1989). Eliciting Knowledge for Analytical Representation. IEEE Transactions on Systems, Man and Cybernetics, 19(3):448-461.

Fischhoff, B., Slovic, P. and Lichtenstein, S. (1978). Fault Trees: Sensitivity of estimated Failure Probabilities to Problem Representation. Journal of Experimental Psychology: Human Perception and Performance, 4:330-344.

Fishhoff, B., Lichtenstein, S., Slovic, P, Derby, S. L. and Keeney, R. L. (1981). Acceptable Risk. Cambridge: Cambridge University Press.

Gherardi, S., Nicolini, D. and Odella, F. (1998). What Do You Mean by Safety? Conflicting Perspectives on Accident Causation and Safety Management in a Construction Firm. Journal of Contingencies and Crisis Management, 6(4):202-213.

Janis, I. L. (1982). Victims of Groupthink. Boston: Houghton-Mifflin.

Kahneman, D. and Tversky, A. (1979). Prospect Theory: An Analysis of Decision Under Risk. Econometrica, 47:263-291.

Kirwan, B. (1994). A Guide to Practical Human Reliability Assessment. London: Taylor and Francis.

Lucas, D. A. (1990). Wise Men Learn by Others Harm, Fools by Their Own: Organisational Barriers to Learning the Lessons From Major Accidents. In M. H. Walter and R. F. Cox (Eds.), Safety and Reliability in the 90s: Will Past Experience of Prediction Meet Our Needs? London: Elsevier Applied Science.
Pidgeon, N. (1998). Safety Culture: Key Theoretical Issues. Work and Stress, 12(3):202-216.
Pidgeon, N. and O'Leary, M. (2000). Man-Made Disasters: Why Technology and Organisations (sometimes) Fail. Safety Science, 34:15-30.
Reason, J. T. (1990). Human Error. Cambridge: Cambridge University Press.
Reason, J. T. (1995). A Systems Approach to Organisational Error. Ergonomics, 39:1708-1721.
Reason, J. T. (1997). Managing the Risks of Organisational Accidents. Aldershot: Ashgate.
Royal Society (1992). Risk: Analysis, Perception and Management. Report of a Royal Society Study Group. London: The Royal Society.
Shanteau, J. (1988). Psychological Characteristics and Strategies of Expert Decision Makers. Acta Psychologica, 68:203-215.
Turner, B. A. (1978). Man-Made Disasters. London: Wykeham.
Turner, B. A. (1994). Causes of Disaster: Sloppy Management. British Journal of Management, 5:215-219.
Turner, B. A. and Pidgeon, N. F. (1997). Man-Made Disasters. Oxford: Butterworth-Heinemann.
Van der Schaaf, T. W., Lucas, D. A. and Hale, A. R. (1991). Near Miss Reporting as a Safety Tool. Oxford: Butterworth-Heinemann.
Van Vuuren, W. (2000). Cultural Influences on Risks and Risk Management: Six Case Studies. Safety Science, 34:31-34.
Vaughan, D. (1996). The Challenger Launch Decision: Risky Technology, Culture and Deviance at NASA. London: Chicago University Press.
Vincent, C., Taylor-Adams, S., Chapman, J. E., Hewett, D., Prior, S., Strange, P. and Tizzard, A. (2000). How to Investigate and Analyse Clinical Incidents: Clinical Risk Unit and Association of Litigation and Risk Management Protocol. British Medical Journal, 320:777-781.
Weick, K. E., Sutcliffe, K. M. and Obstfeld, D. (1999). Organising for High Reliability: Processes of Collective Mindfulness. Research in Organisational Behaviour, 21:81-123.

Incident Investigation Method for Cooperative Safety Management

Yoshio Murayama (1), Yusuke Yamazaki (2)

(1) Maritime Labour Research Institute
Kaiji-bld, 4-5 Koji-machi, Chiyoda-ku, Tokyo Japan 102-0083
E-mail : JDV00353@nifty.ne.jp, Tel : +81 3 3262 1184, Fax : +81 3 3239 8246

(2) Toyama national College of Maritime Technology
1-2 Neriya, Ebie, Shinminato-city, Toyama-pref., Japan 933-0293
E-mail : yamazaki@toyama-cmt.ac.jp, Tel & Fax : +81 766 86 5234

Abstract: This article discusses to progress an incident investigation system for developing performance of safety management. Authors made the questionnaire for marine incidents and conditions of navigators' Performance Shaping Factor (PSF). The questions were arranged 55 main questions include 45 sub questions in seven sections. There were many respondents who answered incidents positively comparing with former pilot study because of their sympathy to the concepts of this research; checking conditions of the PSF and finding important dangerous relations between the factors and the incidents. The methods of the analysis for the relations were the zero-order and the first-order cross tables. The results showed two important relations that increased problems of crossing ship and of short time awareness for danger. Safety management using this system motivated navigators and safety managers through the process of these investigations and feedback of the results. If safety stuffs recognize the report as respondents' effort to answer, respondents are encouraged to improve conditions of PSF and next investigation.

Keyword: Incident, Marine, Performance shaping factor

Introduction

In order to prevent marine casualties, the investigation of accidents and incidents has been standardized internationally, and cooperation between nations has led to the investigation becoming increasingly feasible (IMO,1997). The subject of such investigations is the determination of the factors which led up to accidents or incidents. In particular, the amendment of the investigation concentrated on human factors (IMO,2000). Both the United Kingdom and the United States of America have created report forms (MAIB,2001) (USCG,2001), and the United Kingdom has begun to collect these reports (Nautical Inst, 2002). Since a few such reports have been collected, EU countries are urged to report incidents and accidents via these forms. A Japanese non-profit organization has investigated the circumstances and causes of incidents which were recognized by navigators (79WGSS,2000). These investigations were reported to the International Maritime Organization (IMO), at which point the IMO suggested the need for more detailed investigations into incidents.

Although investigation reports of accidents have been cumulated every year (MDIA,2000), incident reports are quantitatively and qualitatively insufficient. We have to consider the following difference between an accident and an incident. An accident to be investigated is clear since third party know it. In the meantime, investigation for an incident becomes possible only for a case in which independent report is submitted, because only the navigator experiences an incident has full knowledge of the situation. However, they hesitate to file a report of the incidents, because they are afraid of disadvantages that any failures on their part will be discovered and blamed. In Japan, navigators more hesitate to report an incident because they are not exempted from obligation for obstruction for marine traffic, even if they report. Therefore, as the first step toward implementation of incident investigation, the report need not registration content for their responsibility, and simple form as check-list (Murayama,1998,2000a). The check-list include not only report items concerning the direct causes leading to the incident but also items concerning indirect causes are raised. To analyze these items allows effective safety countermeasures to be developed. We achieved two investigations and analyzed by these methods. One of the result was frequency distributions for situations of the incidents (Yamazaki, 2001), and the other was a statistically analysis using the method of multivariate analysis (Murayama,2002). This article reports further analysis by computer program for safety staffs.

Questionnaire for marine incidents

Measurements of the investigation: The investigation evaluates conditions of the Performance Shaping Factor (PSF) of navigators experienced incidents to improve them. The PSF is comprised of two groups of inside factors and two groups of outside factors (Miller,1987). One group of inside factors is navigator's ability to work that is psychological and physiological ability of a human. Another group of inside factors is the readiness for work that changes according to the conditions. One group of outside factors is the technical conditions that makes task required to perform difficult. Another group of outside factors is organizational matters to cope with psychological influence of environment, structure and functions of organization. The report form was made up of the items for following factors (Murayama,2000a):

Abilities of work: The abilities are comprised of both the psychological and physiological abilities. Psychological ability includes traits such as intelligence, memory, knowledge, character and attitude. Physiological ability is related to disease, recovery for fatigue, and factors such as aging and physical fitness.

Readiness to work: The psychological and physiological conditions of navigators influences specific abilities. The readiness to work is affected by psychological motive and interest, excitement, sense of responsibility, and awareness to safety, and daily schedule has an effect on their physiological rhythm and fatigue.

Technical conditions: Difficult conditions of work might bring about navigators' error. The difficulty involved in maneuvering a ship includes navigation involving difficulties of maneuvers, a number of specified maneuvers, navigation instrumentation, congestion of traffic. In addition, the effects of disturbances, and the level of difficulty associated with the navigation and the route affect to the ability to perform technical skills.

Organizational conditions: Organizational facilitation and norms have a distinct influence on human behavior. Organizational conditions include company demands, indication for employee, company expectations, supervision, job security, and job satisfaction.

Items of the questionnaire: The questions are arranged 55 main questions include 45 sub questions for PSF in seven sections. Those are as follows:

- Seven questions for ability of work: age, carrier as seaman, practice, license, job ranking and etc.
- Seven questions for readiness for work: rest, term on board, working schedule, job rotation and etc.
- Eleven questions include 17 sub items for technical conditions: dangerous situation, obstacle, sea area, traffic condition, environmental conditions and etc.
- Two questions including 10 sub items for readiness for work: working hour, psychological and physical conditions and etc.
- Two questions including 12 sub items for readiness for work: attitude for work, job satisfaction, organizational matters and etc.
- The last question for crew including 6 sub items for technical tendencies: tendencies of maneuvering, desire to improvement of hard-ware, safety management and etc.

Additional twenty-five questions for technical and organizational conditions for captain: navigation equipment, crew and their working condition, company's management style and etc.

Composition of the questionnaire: The accuracy of the answer to the investigation can be affected by memory, consciousness, effort, and the expressive skill of the subjects. Navigators' knowledge, attitude, or expressive skill is found to differ greatly. That is why simplicity is requisite for the investigation. The form of selective questionnaire is adopted that is whenever possible, particularly when the experience, and whoever possible. This requires queries to be formulated such that the desired information is accurately indicated to the respondent and responses to be formulated such that the answers are categories or scales. In addition, the questions are placed in an order such that started from navigators' self and direct matters; age job rank, duty hours, etc., and progressed to indirect and surrounding matters; crew, ship, environment, and etc., and questions for incident and its conditions.

Investigation procedure

Subjects: The investigations were performed twice, once in 2000 (Murayama, 2000b) (Yamazaki,2001) and once in 2001 (Murayama,2002). The samples of the first investigation were 617 coastal ships belonging to nine companies that operated the majority of ships involved in domestic transport, and 34 ferry ships belonging to eight companies. The samples of the second investigation were 1,700 coastal ships. These ships were selected

from among 30% to 80% of the ships belonging to 150 companies which operated more than 10 ships. The subjects of the investigation were the bridge navigators of the ships. The investigations were begun after explaining the goal of the investigation to the safety managers of 28 large companies. The letter requesting cooperation with this investigation was mailed to the other companies along with the questionnaire. The questionnaires were mailed to the captains and distributed to the subjects through the safety managers. After filling out and sealing the questionnaires unless subjects' name, the subjects submitted the questionnaire to the captains, and the questionnaire were returned to the safety managers of their company. The sealed questionnaires were then returned to the researchers.

Process of analysis: There were three procedures for evaluating the PSF and for finding relationship between these factors, which are shown in the figure 1.

Selecting factors: The numerical values of some answers were divided into five scales. The frequency distributions for these scales or categories of items were made. Next we compared the distributions of items that had some relations such as age and career. We regarded the similarity or difference between the frequency distributions contained some characters. The problems to be noticed were set from the relations and recent accidents, changing conditions.

Grouping and calculating odds ratio: These scales and categories of the items were gathered in two groups from the meanings of the items and of categories or scales. The relevance of items was obtained from the odds ratio of a zero-order cross table of pairs of the two groups. When the odds ratio was less than 1, the relevance of the factor was made to be a negative reciprocal. We regarded the items as dependent values that were the noticed problems and were comparatively large odds ratios between other items that were independent values.

Finding relation: The other items (the third factors) which related to the dependent values or the independent values were found out from the first-order cross tables. The first-order cross tables composed of double zero-order cross tables which were two groups divided by the third factors. When difference in the frequencies of one category of the dependent value, the third item effected to the relation between the dependent value and the independent value. Large differences of frequencies of the problems of the first-order cross table by the third step were regarded as important relations to eliminate the problems.

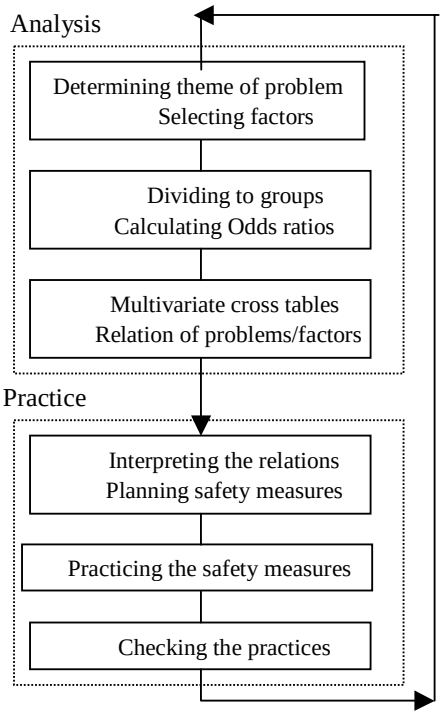


Figure 1. Managing flow of the incident investigation

Process of the practice: We suggested the process of the safety management based on the analysis consists of three steps were as shown in figure 1. The first, safety staffs make plan of safety measures through considerations for the relations why eliminate the problems and what measurers are possible and effective. The second, the planed safety measures must be practiced that reduces the problems by managing PSF. The goal of the reduction of the problems will be informative for navigators and related parties. The last, performance of the practices is evaluated and reexamine the plan. The questionnaire will be useful for the evaluation.

Tool: The computer soft ware for the three analysis steps above mentioned is made by the Excel and its macro command since it is the most popular personal computer utility soft ware, by which many safety staffs can analyze the data from their point of view. The operations include only data entry to the data sheet, level input that divides the categories to two groups, and code input of the factor that verifies the relevance, then they can sea the ratio of the problems that related to items input code.

Results of the investigation

Respondents: Total number of respondents of recent two investigations was 2,831. They had gap between career of seamen and years on board coastal merchant ship, which means many of them have moved from a fishing boat to a merchant ship who were inexperienced in congestion of marine traffic. Their routine generally consists of three months continuous work on board and a month holiday, which were large variation for individual. Sixty percent of all respondents reported performing a single watch keeping on bridge and the remainder reported performing a double watch keeping. Fixed hour watch system was adopted for sixty percent of all navigators, and rotating the system or variable the system was adopted for halves of remainder. Therefore various conditions for career and work might effect their performance.

Incidents: We received 2,165 answers for incidents of the respondents. The period from the day of the incident to the day it was reported was under 10 days for over one-third of all the incidents, and the median value was 24 days. This means that many navigators experienced one incident in every month. The incident frequency according to the time of day increased from midnight to early morning, and short-term fluctuation appeared in four-hour cycles around the time of the watch change. When they experienced the incidents, 1,224 respondents avoided accidents by performing emergency procedure, 564 respondents were forced to perform difficult operation, 305 respondents avoided an accident without performing any action, and 52 respondents experienced minor accidents.

Obstacles of incidents: Half of the obstacles of the incidents experienced by 1,096 navigators was a fishing boat as shown in the figure 2. Next many ships were merchant ships: general cargo ships, gravel cargo ships, tankers. Multiple obstacles and shore were less than these cases.

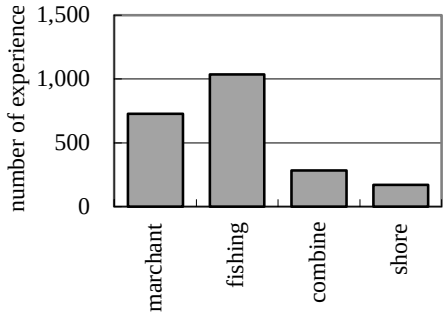


Figure 2 Frequncy to type of obstracts

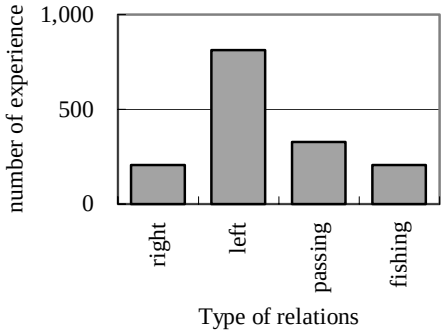


Figure 3 Relation with obstacles

A half of the relations to obstacles was crossing from left side ahead: left crossing as shown in the figure 3. The ratio between left crossing to right crossing is 4 to 1. Navigators have the duty to keep own ship motion for left crossing ship, which caused many incidents.

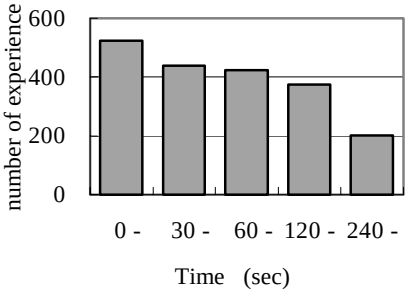


Figure 4 Frequency to dangerous time

Dangerous time: The time until respondents experienced the most dangerous situations after they felt danger for the obstacles: dangerous time, was less than 60 seconds for many navigators as shown in the figure 4. Frequency distribution to the time showed the median value was 50 seconds for fishing boats, 55 seconds for merchant ships, and 25 seconds for wharves or piers. The distance to the object presenting the most dangerous situation: dangerous distance, was in most cases approximately 50 meter. The median value of frequency distribution to dangerous distance was 133 meter for merchant vessels, 90 meter for fishing boats, and 12 meter for wharves or piers.

Incident background: Eight hundred incidents occurred on open sea, 723 incidents occurred in narrow channels, 562 incidents occurred in harbors, and 212 incidents occurred in bay areas. There were many

merchant ships when these incidents occurred: 438 navigators reported three ships, 373, 329, and 182 navigators reported two, four, and eight ships. In addition, navigators reported fishing boats within two nautical miles when these incidents occurred: 271 navigators reported five to six boats, and 248 navigators reported nine to 10 boats, Over 200 navigators reported 20 fishing boats. In several cases both types of ships were reported. They emphasized the congestion of sea-lane by fishing boats and merchant ships that were answered in the scale of the questionnaire also. A half of the incidents concerning improper operation of another ship related to the little allowances for ship handling, and another half of them related to the third ship on the sea area. Although there were not many incidents arose due to special conditions: poor visibility, strong wind, dead angle of visibility, conscious of the third ship, delay of recognition of obstacles, tight navigation schedule, mental tension, and other unusual mental and physical conditions, the cases were relatively many in relation to the obstacle ships and insufficient sea area for ship handling. This means that the basic problem was traffic conditions, which increased the problems under another conditions, so that the problems will be decreased by safety measures for another conditions.

Relationship between incidents and PSF

In order to find out important targets of measures that will enhance safety, we put up theme to effect the problem. The first theme was navigators' brief experience of maneuvering in congesting coastal routes who moved from fishing boats. The second theme was type of obstacles that was mainly fishing boat. These were independent variables. Two problems related to these tendencies that short dangerous time and many crossings ship from left ahead. These are dependent variables. The other items: third factors, related to the independent variables and the dependent variables were investigated odds ratio using zero-order cross table, and the first-order cross table showed effect of the items.

Dangerous time: The dangerous time was found to be very short as shown in the figure 4, which was a serious cause of danger. A ship can overtake a stationary obstacle at a distance of 400 meter in less than one minute, which is insufficient distance to avoid danger. The problem little increased in cases of navigators having long career excepting absent mind. Unexpected obstacles, unawareness and anxiety get worse the problems. Most serious items of these relations are anxiety for long career navigator, and absent mind for short career navigators. If they were not the conditions, about ten percent reduced the problem.

.

Table 1. The ratio of problems of short dangerous time related to the factors and conditions

Third		Career	Short time	n
obstacle	expect	short	37%	199
		long	45%	589
	un expect	short	50%	212
		long	57%	502
mind	aware	short	42%	365
		long	50%	968
	absent	short	60%	48
		long	53%	127
work	no	short	43%	345
		long	48%	928
	anxiety	short	48%	64
		long	62%	169

Relations with obstacles: Other problem was dangerous crossing ships from left. Whether to keep own ship motion or to avoid an obstacle is examined from two standpoints. One is an examination of the reasons for obstacles not changing the relation before other navigator feel danger. The other point of view is an examination of the reasons for they did not avoid before respondents have duty to keep course and speed of own ship. Since many cases were short dangerous time or distance, the results of the analysis shows that serious problem was nearing fishing boats, which increased about twenty percent comparing with merchant ships. The third factors did not totally relate to the problem, but incase of nearing fishing boats on narrow sea

area or they had to pay attention, the problem became worse. Serious problems are crossing fishing boats having duty to avoid another ship, which are inevitable in spite of navigators' attention especially on narrow sea area.

Table 2. The ratio of problems of crossing ship from left ahead related to the factors and conditions

Third		Obstacle	Left crossing	n
sea area	wide	merchant	51%	332
		fishing	68%	389
	narrow	merchant	35%	151
		fishing	79%	118
recognition	no delay	merchant	51%	324
		fishing	75%	346
	delay	merchant	37%	163
		fishing	59%	165

Discussion

The incident investigation method to be efficient use in business is discussed following matters. The first is how to obtain good cooperation with seafarers to reporting incidents that the questionnaire for getting sufficient number of answers. The second is how reflects actual conditions of incidents and PSF. The answer must not be merely the difference of respondent's individual variation, or deflection from self-defense attitude. The third is relationship between actual frequency according to the type of accidents and incidents. The fourth is effectiveness of the third factors in providing effective safety measures. Finally, the management of the incident investigation and make use for safety measures.

Cooperative response: The pilot study was only requested by the letter to safety managers, the respondents were a quarter of the subjects and only a half of respondents answered incident experiences (Murayama,1998). Otherwise, these two investigations got many answers since the investigations were started after explanations for the aim and efficiencies of the process and the results to safety managers from the researcher. The ratio of the answers to the subjects went up to 3 times comparing with the pilot study. This fact appeals the understanding of safety managers is important to obtain incident reports, and at the same time the understanding come down to the subjects. When we explained the meanings of the investigation, we emphasized that they found out factors relating incidents to make safety measures and that the respondents became aware of dangerous factors through the consideration for incidents. And we promised to inform the results of the investigations to safety staffs and to provide the computer program for the analysis.

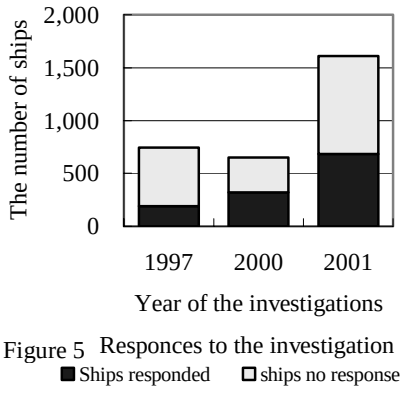


Figure 5 Responces to the investigation

Accuracy of the answers: Although many respondents pointed out traffic conditions; congestions of ships and inadequate operation of other ships, a few of respondents answered to the questions for own peculiar mental or physical conditions. This tendency might be appeared from psychological attitude to put the blame on dangerous factors caused by other person. Some questions were answered to scale values, which were apt to include individual variation. We calculated individual average and standard deviation of the answers to the scale, and evaluated individual variations by these relations; large standard deviation means variety of conditions of incidents, small standard deviation but high or low average means large individual variation. The results show that many respondents' answer included both influence of the conditions and individual variations. We need to recognize the answers include error of the tendency and of the variations. Multivariate

analysis was adopted to control some errors and to make clear the relations between incidents and items (Murayama,2002).

Relation between accidents and incidents: Area, time of day and relations to obstacles of accidents were similar to those of incidents (MDIA,2001) (Yamazaki,1999). Otherwise the ratio of the number of accidents on every type: collision by two moving ships, collision by single moving ships or grounding of passenger ships and other merchant ships was 9:10:16, the ratio in our research was 97:10:4. Among the incidents reported in our survey, collisions were 10 times greater than the report by MDIA, and groundings were a quarter of accidents. Many respondents felt danger for obstacle ships and boats but they did not feel danger for grounding in spite of causing many. An important problem of grounding is awareness for the danger. Therefore, to consider the process of accidents and incidents and their factors effecting them is important for frequency of a specific type of accident.

Relations between incidents and PSF : The questionnaire were consisted of items of incidents and PSF, which related to dangerous situations and PSF. The relations were the targets of safety measures. We suggested the problems of short dangerous time related to navigators career on coastal ship and crossing ship related to types of obstacles. The relations were effected by the third factors of PSF. In the case of short dangerous time, the third factors were expectation of action of another ship, and the navigator's anxiety for other work. In the case of crossing ship and delay to recognize obstacles, difficulty of ship handling were third factors.
These relations show some factors to plan safety measures, efficiencies of which we evaluate from the first order-cross tables. The ratio of the problems in every relation of items is the indicator of the target of effective safety management. It is necessary to change the relations of the conditions of PSF for reducing problems. The conditions are the combined items that shows low ratio of the problem in the first order cross tables. If navigators attempt to reduce problems of short dangerous time, they need to avoid absent mind of for navigators having short career and anxiety for navigator having long career as shown in the table 1. These enable them to reduce about ten percent of the scene. As same that, in the cases of danger to the crossing ship, safety measures for controlling fishing boat in narrow area and not to delay to recognize fishing boat are enable to reduce about 20% of the scene as shown the table 2.

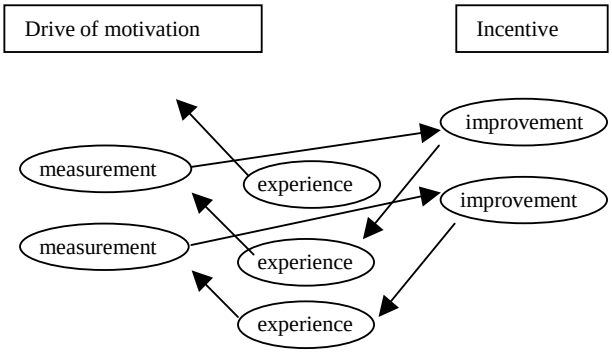


Figure 6. Dynamic process of motivation through safety managers

Safety management process: Although safety managers are interested in useful information to plan effective safety measures, navigators are not necessary interested in the information because they continue same work safely day by day without some exceptions. They need motivations to activate the safety measures. The investigation examined three processes. The first process was to inform meanings of the investigation as communication to improve work conditions for safety between safety staffs, navigators and related parties, which brought them cooperative attitude. The second process was to know what they explained difficult conditions of navigations through the investigations, which brought them incentive to improve the conditions. The third process was to inform the results include suggestion for important problems to make measures to safety staffs. Next they will plan and practice the measures. If improve navigation conditions will be encourage them to develop the managements. We most emphasized in these process were not to find defects but to

evaluate their effort to maintain or improve conditions of circumstances and PSF more safety. The schematic diagram of the process is shown in the figure 6 (Murayama,2000b)..

Summary
To investigate the mechanism of accidents and incidents, and factors relating to the mechanism make clear effective safety measures. We developed the investigation method through some field research.
The goal of the present study is the development of a method to lighten the burden associated with reporting an incident in advance by preparing a check-list by which to collect a great deal of data easily. In addition, the data can be used to analyze the mechanism of an incident.
The category of the checked item was divided into two content independent groups, and zero-order cross tabulation was performed on all items. The odds ratio was then obtained from the matrix of the result.
The cross tables and their odds ratios which seemed to cause the problem examined whether pair of items has relations or not, and we decided items to effect the problems as dependent value and items effect the relations between them.
The computer program is made by the tables and macro of the Excel, which is the most popular micro computer utility soft ware, so that all safety stuffs use and change the program meeting with users intention.
The examples of the results by the analysis are that the frequency of dangerous crossing ships associated with fishing boat and narrow sea area and delay to recognize obstacles, the dangerous time associated with absent mind of navigator having short career and anxiety for other work of navigators having long career.
In addition, the present survey is useful for the improvement of marine traffic control by concerning organizations. We examined the relationship between the characteristics of the sea lane and the navigating ship by investigating the attributes of the ship operator and the work schedule, as well as the personnel management and industrial relations within the company.

Reference:

IMO (1997) . Code for the Investigation of Marine Casualties and Incidents. Resolution A.849(20), IMO (London)

IMO (2000). Amendment to the Code for the Investigation of Maritime Casualties and Incidents. (Resolution A.849(20)), Resolution A.849(20), IMO, (London)

MAIB (2001). Incident Report Form. URL (http://www.maib.detr.gov.uk/)

Maritime disaster inquiry agency (MDIA) (2000). Report on Judge for Marine Disasters, MDIA (Japan)

Miller, D. P., Swain A. D. (1987). Human error and human reliability. (G. Solvandy (ed.): Handbook of Human Factor), Wiley-Inter-science. Prude Univ.

Murayama, Y., Yamazaki and Endo, M. (1998). A pilot Study on Marine Incidents. The journal of Japan institute of navigation, 98, 257-264

Murayama, Y., Yamazaki, U. and Endo, M. (1999). Investigation System for Safety Management Applying Multivariate Contingency Analysis on Human Errors of Maritime Casualties, Proc. Int. Con. On TQM and Human Factors, vol.2, 259-264,(Sweden)

Murayama, Y., Yamazaki et al. (2000a). Consideration in Research for Human Factors on Ship Maneuvering Incidents. The journal of Japan institute of navigation, 102, 173-181

Murayama, Y., Yamazaki, U.(2000b). Performance Measurements to Motivate System Operator, Proc. 3rd Int. Con. on Building People and Organizational Excellence, 418-422,(Denmark), 418-422

Murayama, Y., Yamazaki, U. (2002). Development and Application of Research for Maritime Incidents on Bridge -□, The journal of Japan institute of navigation, 106

Nautical Institute (2002). MARS Search. URL(http://www.nautinst.org/marineac.htm)

The 79th Working Group of Study on the Standard (79WGSS) (2000). A study on Standard for Vessel- Human Factors-, Association of study on ship-building, (Japan)

USCG (2001). National/International Maritime Safety Incident Reporting System. URL(http://www.uscg.mil/hq/g-m/docs/blue.htm)

Yamazaki, U. Murayama, Y. and Endo, M. (1999). A comparison of Maritime Incidents and Marine Accidents. The journal of Japan institute of navigation, 104, 245-251

Yamazaki, U. Murayama, Y. (2001). Development and Application of Research for Maritime Incidents on Bridge, The journal of Japan institute of navigation, 104, 173-178

Integrated Safety Investigation Methodology (ISIM) - Investigating for Risk Mitigation

Marcel Ayeko,

Transportation Safety Board (TSB) of Canada
http://www.tsb.gc.ca (marcel.ayeko@tsb.gc.ca)

Abstract:
Over the past 10 years (1992-2001), an annual average of over 4,200 incidents and accidents were reported to the *Canadian Transportation Accident Investigation and Safety Board,* commonly known as TSB. TSB investigates accidents solely for the purpose of "advancing transportation safety." In accordance with its Classification Policy, an average of 80 to 100 accidents with the greatest potential for advancing transportation safety are investigated each year. Since 1992, TSB has investigated over two thousand such accidents and incidents involving air, marine, railway and pipeline modes of transportation. TSB analyses of these accidents show that transportation accidents, just like those in other industries, are the result of multiple causes rooted in underlying factors; human and organizational factors are implicated in most accidents. Peoples are essentially human system components involved in complex interactions with other system components such as machinery, equipment, procedures, and other humans, in an operating environment, to accomplish a certain mission objective. As such, peoples' actions are often influenced by their mission, goals, and the operating conditions. When a failure occurs and causes are attributed to "human error", the people on the scene are not always at the root of the problem. This suggested a need for a much broader system approach to our investigation practices. Hence, in 1998, TSB developed the *Integrated Safety Investigation Methodology (ISIM)* which includes the analysis of not only equipment functions, but also the interfaces between human and other system components controlled by managerial, design, regulatory and other procedures. When this approach is used as a basis for accident prevention, it can address safety issues in a holistic context. ISIM provides an accident investigator with a framework, discipline, tools and techniques to look for information beyond the immediate cause. ISIM also encourages investigators to assess risks associated with safety deficiencies so that their efforts and valuable resources may be allocated to those safety issues with highest risks. The underlying principle of ISIM is that accident prevention should be based on realistic analysis of risks rather than the strict compliance with prescriptive rules, standards and regulations. TSB believes that while rules compliance is necessary for accident prevention, that alone is not sufficient to maintain or advance safety. ISIM embraces the "defence in depth" philosophy which seeks multiple and diverse lines of defence to mitigate the risks of normal human errors.

Keywords: ISIM, Accident Investigation Methodology, Risk Mitigation, Human & Organizational Factors

Transportation Safety Board and its Objectives
TSB, is a Canadian federal government agency; TSB's sole mandated is to "advance transportation" safety by:

- conducting independent investigations, including, when necessary, public inquiries, in order to make findings as to their causes and contributing factors;
- identifying safety deficiencies as evidenced by transportation occurrences;
- making recommendations designed to eliminate or reduce any such safety deficiencies; and
- reporting publicly on its investigations and on the related findings;

TSB conducts selected investigations into aviation, marine, railway and pipeline accidents and is independent of other government departments that regulate or operate elements of the marine, rail, commodity pipeline, and air transportation systems. TSB investigations are carried out with the prime purpose of identifying *Safety*

Deficiencies[4] in transportation occurrences and to propose corrective safety action designed to eliminate or minimize risks associated with any such deficiencies. It is not the function of the Board to assign fault or determine civil or criminal liability; however the Board does not refrain from fully reporting on causes and contributing factors merely because fault or liability might be inferred from its findings.

Shortcomings in Traditional Approach to Accident Investigations

Accident investigations are carried out for various objectives as follows:

(a) to find out "What happened?"
(b) to determine "Who did it?"; and
(c) to improve safety;

Traditionally, investigations in the past placed more emphasis on (a) and/or (b). Objective (a) will be met if the investigation can just determine the cause of the accident; once the *immediate cause* of an accident is found, the process of investigation often stops without further examining the underlying factors and contributory conditions leading up to that *immediate cause*. Determination of *immediate cause* is useful in identifying who had the last opportunity to intervene and prevent the accident or who failed to assess the risk. However, it does little in terms of developing an understanding of the unsafe conditions, which lead to the accident.

With objective (b), the investigation will be looking for who is to blame with a view to taking deterrent measures as well as establishing damage compensation and punishment such as civil / criminal liability. For example, an investigation might conclude upon determining that a collision occurred because the master of the fishing vessel did not proceed at a safe speed. Possible underlying factors such as the requirement to maintain a tight sailing schedule, to take advantage of a per-trip fishing quota, or the need to work long hours resulting in fatigue due to a small complement, etc. were usually left undetermined. As such, cause determination or apportioning blame by itself would not do much to improve safety except with respect to its deterrent value.

How Do We Improve Safety?

As indicated above, the sole purpose of each and all TSB investigations is to "advance safety". We all have our own understanding of what ***Safety*** is. However, for the purpose of this discussion, let us adopt the Oxford dictionary's definition of "***safety***" as "*freedom from danger or risks.*" Risk has two elements and is commonly defined as the product of the probability of an adverse outcome during a specific period of time and the severity of that outcome.

RISK = PROBABILITY X CONSEQUENCE

If we attach the units of measurement, the Risk equation may be written as follow:

$$\text{RISK}\ \left[\frac{\text{Impact}}{\text{Time}}\right] = \text{PROBABILITY}\ \left[\frac{\text{Event}}{\text{Time}}\right] \times \text{CONSEQUENCE}\ \left[\frac{\text{Impact}}{\text{Event}}\right]$$

Therefore, the improvement of safety means the elimination or reduction of risks. **Risk** can be treated by either reducing **probability** and / or by reducing the **consequences**. To do so, one must understand the causes and underlying factors that contribute to both elements of the RISK equation. If the focus of an investigation is only on the causal factors and on preventing "*recurrence*", it will limit the potential for safety improvement by not considering the second element of the risk equation – i.e. the **consequence.**

Many of us can think of an accident that had factors at play that were not causal, but that contributed to the severity of the outcome; i.e., the consequence. An obvious example would be inadequate lifesaving equipment, deficient search and rescue operations or inadequate emergency response to train derailment involving dangerous goods in a residential neighbourhood. Another could be a design characteristic of an aircraft or a ship that allowed a relatively minor incident to become a serious accident. Eliminating such deficiencies will do nothing to prevent a future accident, but it may significantly improve safety by reducing the severity of the consequences.

[4] *Safety Deficiency*, in the context of ISIM, is an unsafe condition or underlying factor with risk for which defences are less than adequate.

Integrated Safety Investigation Methodology - ISIM

To minimize risks in the transportation system, the TSB's *investigation methodology* places emphasis on the identification of safety deficiencies in the system and the assessment of risks associated with such deficiencies. The ISIM process is made up of eight systematic steps (Figure (1)) commencing with the initial assessment of the accident, whether "to investigate" or "not to investigate", through to the effective communication of the identified risks to those who can affect the necessary change. However, for the purpose of this paper, only the following five important steps will be discussed:

1 Collection of occurrence data.
2 Analysis of "How's" and "Why's" of the accident (determination of sequence of events & identification of safety deficiencies).
3 Risk Assessment.
4 Barrier (Defence) Analysis, and
5 Consideration of Risk Control options.

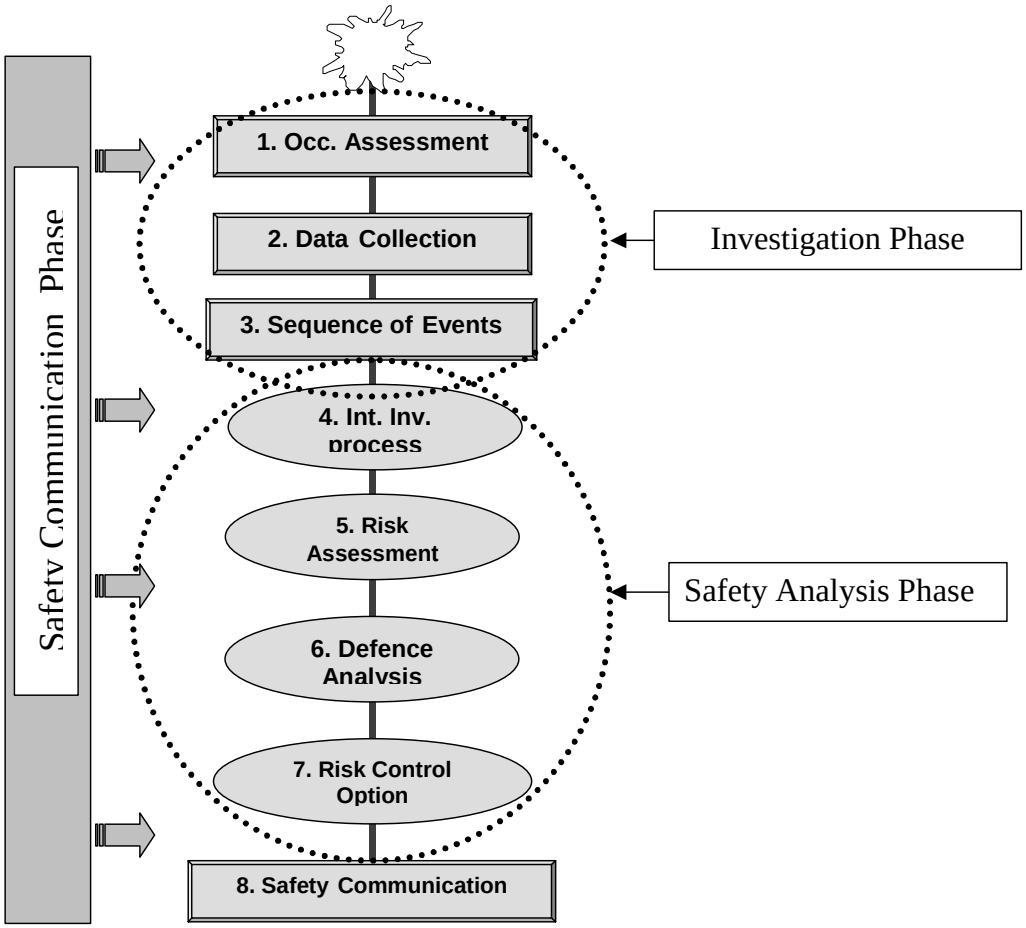


Figure (1) - ISIM Process

Collection of Occurrence Data

The first step in an investigation process is the collection of information regarding personnel, tasks, equipment, and environmental conditions involved in the occurrence. A systematic approach to this step is crucial to ensure that a comprehensive analysis is possible to determine not only "what, who and when" of the accident but also "why and how" the accident happened.

To conduct an effective systematic data collection, the investigator must recognize that regardless of the type of accident, there are five core elements that can play an interactive role in causing the accident; **Human, Machine, Medium, Mission, and Management.** Like any industrial operation, transportation is a complex operation system where **Human**, **Machine** (aircraft, ships, locomotives, equipment, machinery, etc.), and **Media** (external and internal environments**)** interact in a confine of **Mission** (schedules, goals, needs, financial objectives, policies, procedures etc.) and **Management** (organization, regulatory bodies, management, etc.) which creates and maintain the system environment.

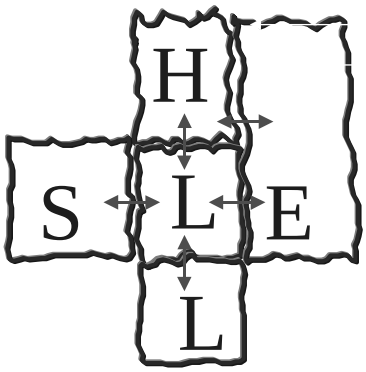


Figure (2) SHELL Model

Often, Mission and/or Management factors influence the way Men interact with Machines in certain system environments that may be unsafe. Analysis of transportation accidents over the past several years indicates that, while one or a combination of the five aforementioned basic risk elements are generally present in each accident, human and organizational elements play, by far, the biggest role in contributing to these accidents. Understanding the interrelationship of these elements can help investigators in determining all the relevant causes and contributing factors of accidents. In a complex system, such as an operation of an aircraft or a ship where there is a broad range of interactions among the system components, with the human operators being at the very core of the system, there is constant danger that critical information will be overlooked or lost during an investigation.

Another technique used by TSB investigators to gain knowledge about the broad range of interrelationship of these components in a system is a framework based on the SHELL model that was developed and later modified by F.H. Hawkins. Although the SHELL model was intended as a tool for human factors studies, it also serves as an effective tool for data collection. The "**L**" block representing LIVEWARE, or the human element, is the centrepiece of the model. The human component interacts directly with each of the other building blocks namely :

- SOFTWARE **"S"** (such as as equipment manuals and instructions, standing orders, rules and regulations, flight operation manuals, charts, etc.) ,
- HARDWARE **"H**" (such as be an aircraft, a ship, an engine, instruments or any other engineering components, etc.),
- ENVIRONMENT "**E"** (environment such as climatic conditions, motions induced by turbulence or sea states, visibility, noise, vibration, fatigue, stress, boredom, etc.) , and
- the second component of LIVEWARE "L" (such as personnel ashore, air traffic control, railway traffic control, vessel traffic control, etc.)

In the transportation environment, the aircraft, ships and equipment should also be designed, installed and maintained for the environmental conditions in which they are required to function. As such, it is essential to consider the interface between Environment and the Hardware.

In the SHELL model, each component has the shape of a block, whose edges are not straight and possess a unique contour, suggesting that these factors may not have a perfect interface. Use of the **SHELL** model as an organizational tool for the investigator's data collection process facilitates the quality and depth of the investigation as it:

- encourages the consideration of the interrelationships between the human and a broad range of system elements; and,
- focuses on the factors, which influence human performance by relating all peripheral elements to the central human element.

At this data collection stage, the investigator initially attempts to answer the more simplistic questions concerning "what, who, and when" and then moves to more complicated questions of "how and why".

Analysis of the "How's and Why's" of the Accident

Having completed the task of collecting all relevant information surrounding an occurrence, the investigators must document pertinent information in a clear and organized fashion so that adequate analysis can be conducted to uncover causes, contributing and underlying factors of the accident. As with any good investigation, the facts and analysis should lead logically to meaningful and supportable conclusions and findings. This process can be better defined in two stages as follows:

- determination of sequence of events of the accident, and
- identification of Underlying Factors and Unsafe Conditions

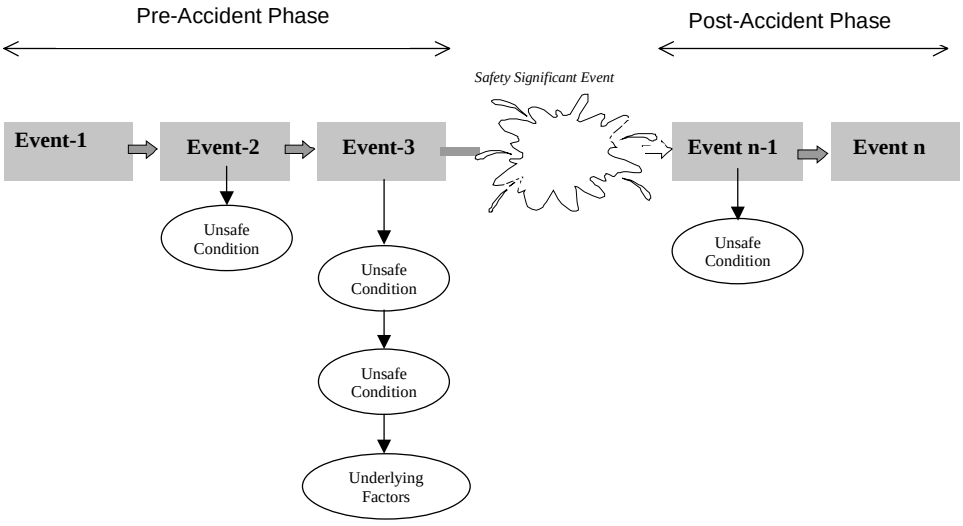


Figure (3) Sequence of Events and Underlying Factors Diagram

Determination of Sequence of Events of the Accident

The concept of graphically representing the sequence of events leading to and following the accident is an effective way of documenting observed and reported accident information; it also facilitates organized thinking and helps identify information deficiencies. This concept is not new and is based on the principle that accidents rarely result from a single cause or event; rather, they are generally multi-factorial and develop from defined sequences of events. The events are portrayed graphically by arranging them, often chronologically left to right in rectangles, in a logical flow indicating 'what' happened. For a comprehensive investigation, the entire sequence of events can be built from any suitable starting event originating as far back as necessary before the accident event. This sequence of events often, if not always, extends beyond the time of the accident event to include circumstances of the post-accident phase. (Figure (3)). A good practice in developing the sequence of events diagram is that each event describes a single, discrete happening or an action step in a sequence of happenings/actions that lead to the occurrence. Each event block should contain the time and source of information whenever applicable. If time is estimated, it is important that this fact be clearly denoted. This technique aids investigators in ensuring the completeness of the investigative logic through the identification of each event deriving logically from the one preceding

it. A clearly defined sequence of events diagram will help investigators in determining *safety significant events* worthy of further investigation/analysis to uncover unsafe acts, unsafe conditions, contributing and underlying factors. A *safety significant event* is an event, which has the potential to reveal unsafe conditions and underlying factors.

Identification of Underlying Factors and Unsafe Conditions

Once the sequence(s) of events diagram is complete, *safety significance events* can be easily identified for further investigation/analysis to determine "why" it happened. In order to understand the "how and why" of the accident, the investigators need to identify the relevant contributing conditions and underlying factors of such *safety significant events.* In many situations, investigators cannot confirm, with certainty, why an accident happened. However, short of determining "why" it happened, we can often find information that can be used to reduce risk. Investigation can, for example, determine unsafe conditions and underlying deficiencies, which in turn can often lead to safety actions that mitigate risks.

Since its inception in 1990, the TSB has systematically analysed its investigative findings to understand contributory conditions and underlying factors of unsafe acts leading to the accidents. TSB consistently found that most accidents could be traced back to compounded human and organizational factors. This finding is consistent with findings from many investigation and safety agencies around the world in that human and organizational factors are recorded as causal or contributing factors in most accidents. According to the UK P&I Club of Insurers report on the "Analysis of Major Claims - 1992", *human error* was the main cause of half the cargo claims, half the pollution claims, 65% of the personal injuries, 80% of the property damage, and 90% of the collisions. According to the 1995 report of the Accident Prevention Advisory Unit of the UK Health and Safety Executive, "*.....human error is a major contributory cause of 90% of accidents, 70% of which could have been prevented by management action....*"

Today, several models, tools, and techniques exist for human factors investigation and analysis. A brief description of such models/techniques often used by TSB investigators is given below.

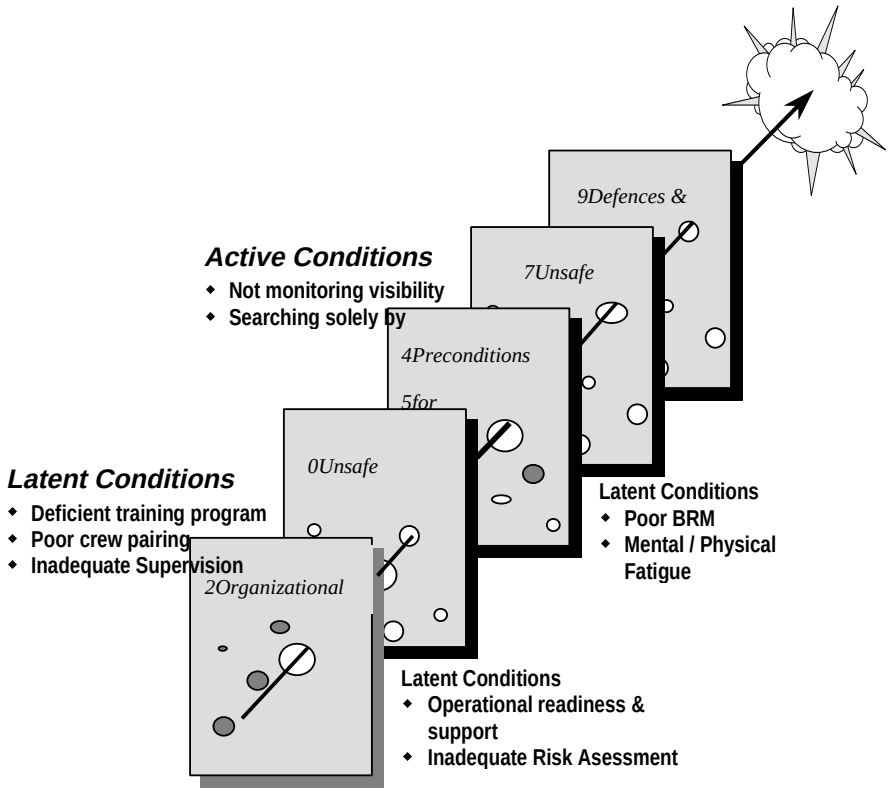


Figure (4) Reason's Model

Reason's Model:
"Reason's Model" was developed by Dr. James Reason of the University of Manchester (Figure (4)). Reason's model encourages us to look beyond the immediate circumstances of the accident. It forces investigators to examine all the preconditions at the time of the occurrence and latent conditions created by line management as well as high level decision-makers and organizational factors such as regulators, owners, the designers, manufacturers, the unions, and management, etc.

Generic Error Modeling System (GEMS) for Investigating Human Performance:
Human performance specialists at TSB use the integration of two models – SHELL and Generic Error-Modeling System (GEMS)[5] to identify underlying human factors by identifying error types, failure mode and behavioural antecedents (Figure (5)). However, to improve their investigation skills, TSB investigators / analysts are being gradually trained to become adequately familiar with evaluating human behaviour and human performance.
For the scope of this paper, it is sufficient to recognize that to uncover the underlying causes behind the decision of an individual or group, it is important to determine if there were any factors in the work system that may have facilitated the error or the unsafe act.

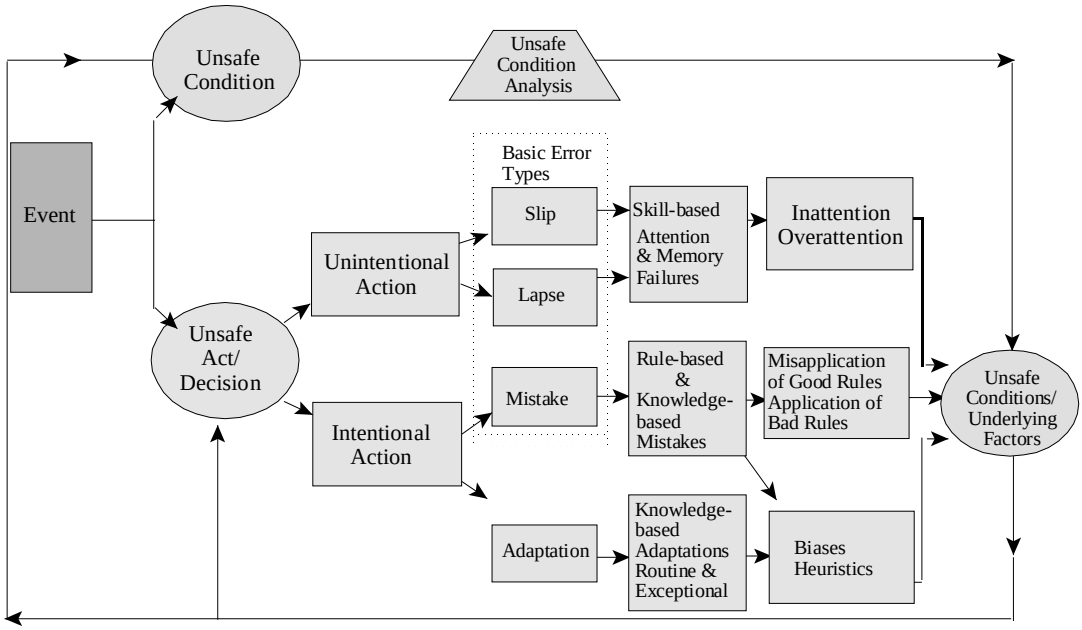


Figure (5) Generic Error Modeling System (GEMS) for Human Factors Investigation.

Other Analytical Techniques:
A few other techniques, such as MORT (Management Oversight Risk Trees Analysis, Fish Bone Analysis, Systematic Cause Analysis Technique (SCAT) and Change Analysis, have been tried at the TSB for certain accident scenarios. Several others techniques and models have been known to many for root-cause(s) analysis; these include, inter alia, Prof. Chris Johnson's work on CAE diagram, Prof. Peter Ladkin's "Why-Because Analysis", TapRoot Analysis, etc.
Fault Tree Analysis (FTA) is perhaps the most common concept in root-cause analysis. FTA is a technique, either qualitative or quantitative, by which conditions and factors that can contribute to a specified undesired event are deductively identified, organized in a logical manner, and represented pictorially. The faults identified in the tree can be events that are associated with component hardware failures, human errors, or any other pertinent events that lead to the undesired event. Starting with the top

[5] James Reason 1990

event, the possible causes or failure modes on the next lower functional system level are identified. Often, a simplified concept of FTA is applied by treating any *safety significant event* identified in the sequence of events diagram as the "top event" and by repeatedly and systematically asking "WHY" until the appropriate level of underlying factors or unsafe conditions are identified.

Risk Assessment

Once underlying factors and/or unsafe conditions related to the *safety significant events* of the occurrence have been identified, it is important to assess the level of risk associated with such unsafe conditions or underlying factors.

Risk is a consideration in every decision made regardless of the role of the person making the decision. The purpose of risk analysis is to estimate and evaluate risk potential associated with the identified unsafe conditions/underlying factors. Some of the unsafe conditions, underlying factors identified above may be neither causal nor contributory to the accident. It is particularly true if such underlying factors are derived from post-accident events. Nonetheless, the potential risks to lives, property and the environment must be assessed and addressed. Evaluation of risks is undertaken using available data, supported by judgements on the severity of potential adverse consequences and the probability of those consequences during any defined period of time. TSB has developed its own risk matrix, which is used by investigators in assigning the level of risk associated with identified unsafe conditions and underlying factors. The level of risk should be one of the important criteria throughout the entire investigation process in allocating investigative efforts and priorities for corrective actions.

Probability Assessment:

In evaluating the probability, the investigators will consult various TSB databases to determine if there is a history of similar occurrences or if it is an isolated occurrence. The answers to some of the following questions can assist investigators in assessing the probability of adverse outcome:

- Is there a history of occurrences like this one or is this an isolated occurrence?
- How many similar occurrences were there under similar circumstances in the past?
- What system defences need to fail for the adverse consequence to be realized?
- How many pieces of equipment or vessels are there that might have similar defects?
- How many operating or maintenance personnel are following or are subject to the practices or procedures in question?
- To what extent are there organizational, management, or regulatory implications, which might reflect larger systemic problems?
- What percentage of time is the suspect equipment or the questionable procedure or practice in use?

Consequence Assessment:

For the second element of the risk equation, the negative impact of unsafe conditions and underlying factors leading to a particular accident, on people, property, environment, and often on commercial and other intangible elements, must be considered; e.g.:

How many persons could be affected by the risk?

- Fare-paying passengers?
- Transportation employees?
- Bystanders or general public?

Property:

- What could be the extent of further property damage?
- Direct property loss to the operator?
- Damage to adjacent infrastructure?
- Third-party collateral damage?

Environmental:

- What could be the environmental impact?
- Dangerous commodity spill?

- Physical disruption of natural habitat?

Commercial:

- What is the potential impact on carriers:
  - On commercial operations?
  - Corporate viability?
  - Financial markets?

Others:

- What could be the public and media interpretation?
- What might be the implications:
  - Internationally?
  - Nationally?

Once the probability and severity of adverse consequences has been considered, investigators can evaluate the risk. Various agencies and industries use numerous qualitative as well as quantitative criteria against which a level of risk can be estimated. Qualitative risk analysis uses expert opinion to evaluate the probability and consequences. The qualitative method offers analysis without detailed information; however, the intuitive and subjective processes may result in differences by those who use them. The quantitative analysis generally provides a more uniform understanding among different users, but requires quality data for accurate results. In TSB investigations, where human reliability issues are involved in most accidents, it is difficult to quantify risk. Qualitative analysis is considered sufficient for the purpose of TSB investigations. According to one risk management practitioner, the process of carefully weighing available information and critically seeking to understand how much or how little we know about any set of risks, is definitely superior to relying on sheer intuition or self-serving belief.

Defence (Barrier) Analysis

Defence analysis is based on the principle that the absence of adequate barriers (whether they be physical or administrative) for preventing any harmful "contact" between *hazard*s and *vulnerable person*s or *property* is found in every accident. The purpose of the Defence analysis is to examine the status of barriers and to identify those that are less than adequate. Defences/barriers, in the context of TSB investigations, are barriers/guards that isolate and protect persons, property, and the environment (targets) from hazards. Defences may be divided into two categories:

- **Physical defences / barriers:** (such as guards, personal protection devices, liferafts, lifejackets, oxygen masks, etc.), and
- **Administrative defences / barriers:** (such as training, safety regulations, policies, procedures, supervision, inspection, maintenance, safe system design, system support services, operational & personal readiness, etc.)

Practically anything that does not fall into the Physical Defence category can safely be considered as an Administrative Defence. TSB investigators are provided with a worksheet as a job aid to identify the status of barriers before and after the accident; (e.g. Were barriers provided? Were they used? Did they fail or did they function as intended? Has their presence been "advertised" to users and operators?)

In fact, some degree of barrier analysis should be done at all level of the investigation process. The information on the status of both physical and administrative barriers must be collected during the data collection as well as during the analysis phase. Analysis of adequacy of defences will lead to a better understanding of unsafe conditions and underlying factors and will be particularly useful in considering the risk mitigation options discussed in the following section.

If one recalls that the "Safety Deficiency", in the context of ISIM, is defined as "an Unsafe Condition or Underlying Factor with risks for which there is no adequate defence(s)", then by the end of this defence analysis, the investigators would have uncovered Safety Deficiencies and associated risks that warrant corrective and control action.

Consideration of Risk Control Options
Once the Safety Deficiencies have been identified, investigators can begin to device safety action in order to mitigate or control associated risks. There are normally control options available for any risk situation. However, the most important aspect of risk management is to ensure that the full range of possible control measures is considered and that the optimal trade-off between measures is made. Obviously, natural phenomena such as weather, waves, wind, lightning - are elements outside our control. The principal elements that can be controlled are the safety features of the system and the safety procedures used in its operation and support. Some control measures are more effective than others.

In assessing the options for controlling risk, investigators would consider one or more of the following strategies:

- Reducing the probability of similar accidents (e.g. designing for minimum hazards, modifying human behaviour, improving human performance, safe operating procedures, etc.)
- Reducing the severity of consequences (e.g. personal protection equipment, emergency response and preparedness, etc.)
- Segregating the risk from vulnerable objects
- Adding redundancy in the safety system (e.g. procedures for safeguarding sigle-point errors such as monitoring of operator's action, crew resource management practices, etc.)

It is obvious that preference should be given to developing safety measures that will completely eliminate the deficiencies to prevent similar adverse consequences in the future. Regrettably, such solutions are often the most expensive and are often impossible. In such situations, some organizations may decide to tolerate a certain degree of risk as a result of hazard analysis or as a result of a cost-benefit analysis. In such cases, an investigator may need to determine the adequacy of the rationale and the extent of risk that is assumed by the organization. Risk can also be transferred to someone else, such as an insurer, for a price. But, as a safety organization, TSB is not interested in transferring risks to someone else. In most cases, where the risk associated with potential safety deficiencies cannot be eliminated in a complex system, the risk to the system may be treated by building one or more of the following defenses/barriers in the system[6]:

- Designing for minimum hazards that induce errors (Both Equipment and Human Factors)
- Installation of safety devices to guard against errors and hazards
- Establishment of procedures and practices
- Provision of warning devices, signs, placards, etc
- Provision of training and awareness.
-

It is important to note that the sole use of administrative interventions, such as procedures and training, may not provide an effective hazard control, especially when the level of risk is high. Rather, the use of administrative interventions in conjunction with engineering interventions, such as designing for minimum hazards, would be more appropriate. TSB generally believes in the "depth in defense" philosophy, particularly for complex systems such as transportation, where multiple and diverse lines of defense are considered desirable to mitigate risks.

As with many similar transportation safety boards and agencies around the world, TSB does not have the mandate or authority to implement specific corrective actions. Such actions are taken by the regulatory agencies, and by the industry such as manufacturers and the operating companies, etc. The TSB's role is to identify safety deficiencies and potential risks associated with system deficiencies and make a convincing argument for others to take corrective actions.

[6]While there are some disagreements as to the order of effectiveness in intervention (known as "safety precedence sequence"), safety professionals are unanimous in proposing these defences/barriers.

Training the Investigators

In 1998, the ISIM concept was developed internally by TSB to prepare its own investigators to better understand the concept of *safety deficiencies* and of risk assessment, and to look for deeper systemic causes so that appropriate remedial measures can be implemented. ISIM approach to accident investigation required new knowledge, skills and attitudes on the part of the investigators. Hence, to assist with the delivery of the two-and-a-half-day ISIM training course, the services of a consultant with a good working knowledge of designing and delivering adult learning program were initially retained. To date, all TSB investigation personnel and over 20 non-TSB personnel have participated in the training program.

The training efforts focused on the practical application of the methodology to the investigations. Since people tend to retain 10% of what they read, 20% of what they hear but 90% of what they say and do, the training was structured for high participants' involvement; as such much of the course is spent in syndicates, working on practical case studies.

At TSB, most accidents are investigated by personnel who, through extensive training and experience, have developed the skills necessary to investigate highly technical and operational aspects of accidents. Partly for that reasons, some investigators were initially cynical towards the practicality of the use of human factor investigation models, the safety deficiency analysis concept of uncovering deeper causes, underlying factors and risk elements, etc. Realizing that the possession of theoretical knowledge by the investigators will be of little use if they lack the skill and confidence to apply these concepts in their daily work, TSB management facilitates a continued learning through reinforcement on the job to hone their newly acquired skills. The management also continues to encourage investigators to use all the resources available to them; for example the services of human performance specialists and the services of those who are proficient in the practical application of ISIM to real investigations. Over the period, TSB has witnessed major attitudinal shift among many of those skeptical investigators; they realized that the ISIM is not something that you add-on to a normal investigation but an integral part of every phase of the investigation. They recognized ISIM as a tool to help them in managing their investigations more effectively by being able to better organize the information, to better identify and prioritize safety significant issues thus enabling them to better allocate scarce investigative resources. However, it is recognized that follow-up on the job training and continued reinforcement on the job will be required to ensure the effectiveness and lasting application of the ISIM's systems approach to accident investigation.

Conclusion

Accident prevention depends to a large degree on lessons learned from accident investigations. But to learn a lesson from an accident one must understand not only the immediate cause but also contributing factors and underlying conditions of the accident. Accidents are rarely, if ever, the result of a single cause. It is my belief that, when we seek "cause" rather than "information about cause" in an investigation of an accident, direction of the investigation often veers towards elements that are more likely to be linked to blame rather than the mitigation of risks.

It has been my experience at the TSB that human and organizational factors are implicated in most accidents. Meaningful analysis of the people and organizational part of the system can help us understand underlying factors so that appropriate safety action can be taken to minimize the human contributions to risk. However, humans have more failure modes and are far less predictable than the machinery or equipment they operate; performance of each individual human is shaped by numerous and varied factors. Accurately determining human reliability is extremely difficult. Yet, accident prevention is critically linked to the investigation of issues related to human interaction with other system components as discussed in this paper. I believe that a systematic and holistic approach to investigation through the use of progressive techniques, such as ISIM, will help us improve the safety of life, property and the environment around the world.

References

[1] Hawkins, F.H (1987)*Human factors in flight*. Aldershot, UK: Gower Technical Press

[2] Reason. J (1990) *Human error*. New York: Cambridge University Press

[3] Transportation Safety Board of Canada (1997*) Integrated process for investigating human factors*

[4] Marcel Ayeko (September 1997) *Knowledge, Skill & Training – Need Analysis Based on Transportation Safety Board's Experience,* IMLA Conference, St.John's, Newfoundland, Canada.

[5] Marcel Ayeko (October 1999).*Integrated Safety Investigation Methodology - Systematic Approach to Learning from Accidents,* International Conference on Learning from Accidents , RINA, London, UK

Reporting Adverse Events in Hospitals: A Survey of the Views of Doctors and Nurses on Reporting Practices and Models of Reporting

Henning Boje Andersen,[7] [(A)] Marlene Dyrløv Madsen,[(A)] Niels Hermann, [(B)]
Thomas Schiøler [(B)] and Doris Østergaard [(C)]

[(A)] Risø National Laboratory, Systems Analysis Dept., Roskilde, Denmark
www.risoe.dk/sys/spm
[(B)] DSI - Danish Institute for Health Services Research, Copenhagen, Denmark
www.dsi.dk
[(C)] Herlev Univ. Hosp., Dept. of Anaesthesiology and Danish Inst. Med. Simulation, Herlev, Denmark
www.anaestesiologi-herlev.kbhamt.dk/startsim.htm

Abstract: This paper summarises some of the results of a questionnaire-based survey of the views of more than 2000 hospital doctors and nurses carried out in Denmark in January-March 2002. The questionnaire was developed as part of a patient safety project sponsored by the Danish Ministry of Health aimed at reviewing the advantages and disadvantages of different models of incident reporting, eliciting doctors' and nurses' requirements to a reporting system, and identifying barriers to incident reporting. The final output of the project is a set of recommendations for a model of incident reporting covering national and local efforts at learning from adverse events to reduce their number and control their consequences. The questionnaire-based survey was motivated by the fact that little is known about doctors' and nurses' requirements to a system to which they may report incidents and, in particular, about their views of options involved in anonymity, confidentiality, and reporting criteria. At the same time, the questionnaire sought to elicit the views of health care staff about barriers that they perceive against bringing up incidents with colleagues or leaders or informing the patient involved. The paper focuses on three types of results of the survey: First, respondents' reactions to adverse events as described in terms of the actions they report they would take if they were themselves to be involved in specific events; second, respondents' views of reporting models and the options associated with these; and third, their views of possible barriers against reporting.

Keywords: Incident reporting, truth telling, incident reporting systems, adverse events, patient safety.

Introduction

The survey study reported in this paper was conducted as part of a patient safety project sponsored by the Danish Ministry of Health and supported by the Copenhagen County Health Care Administration. The primary aim of the patient safety project was to produce a set of recommendations based on (a) a review of international and Scandinavian experience with incident reporting systems in hospitals and, in general, efforts to implement organizational learning based on experience from adverse events; (b) focus group interviews with Danish hospital doctors and nurses and (c) a questionnaire-based survey of the views of doctors and nurses of key characteristics of reporting systems as well as their perceptions of barriers against bringing up adverse events with colleagues and leaders as well as informing patients involved.

Four counties (Copenhagen, Roskilde, Frederiksborg, and Ringkøbing) participated in the survey and questionnaires were distributed to a random sample of doctors and nurses drawn from a selection of hospitals in the former three counties to ensure that university hospitals, central hospitals and local hospitals as well as the selected specialties were all represented (see below). In the fourth county, Ringkøbing, questionnaires were distributed to all hospital doctors and nurses in the specialties selected. The specialties of the departments and wards comprised:

- Surgery (general / other surgical specialties)

[7] Corresponding author: henning.andersen@risoe.dk

- Anaesthesiology / Intensive care medicine,
- Gynaecology / obstetrics,
- Internal medicine and other medical specialties,
- Orthopedic surgery

In this paper we present, for obvious reasons of space, only a small sample of results from the survey. In particular, results provided here describe differences between doctors and nurses but not between specialties, age groups, types of hospitals or between the Danish and the Japanese survey using parts of the same questionnaire (Itoh et al. 2002), nor have we yet analysed the free-text comments offered by respondents to open-ended questions. The authors are currently preparing publications that will describe results in greater detail.

The Questionnaire

A basic input to the creation of the questionnaire were the results of four two-hour focus group interviews with groups of doctors and nurses. The interviews revealed quite divergent views and attitudes to incident reporting and, in general, to bringing up incidents with colleagues and leaders; at the same time they were useful in prompting us to be especially careful and explicit in defining key properties of possible reporting systems such as 'anonymity', 'confidentiality', 'feedback' to reporting staff.

The questionnaire was developed iteratively using pilot tests in which about 35 nurses and doctors answered and commented on the questionnaire. In addition, human factors and medical colleagues commented in detail on the pilot versions.

The questionnaire has seven parts preceded by a short Foreword describing the purpose of the questionnaire, emphasising that the survey is strictly anonymous and specifying the (two-week) deadline for returning responses. In addition, the questionnaire items are prefaced by a short glossary defining what a reporting system is and the notions of 'adverse event' and 'error', reproduced in Table 1.

Table 1 - Glossary from the Foreword of the questionnaire

A reporting system is a system designed to receive, for the purpose of learning from past mistakes, accounts by health care staff of both "*adverse events*" and "*errors*" that have or might have caused patient injury.

An *adverse event* is an event that actually or potentially involves or leads to a patient injury and that is not caused by the underlying disease of the patient. Adverse events include complications as well as *errors*. An *error* is an act that is not carried out as the agent intends or that follows an incorrect plan to achieve the agent's goals.

The first and introductory part of the questionnaire contains four fictitious cases, each of which is followed by questions about specific actions. The cases formed the first part of the questionnaire for two reasons: first, the focus group interviews had made it obvious that health care staff had quite different notions of adverse events and errors. By having concrete cases presented to respondents, it was hypothesised that this would make the abstract notion of adverse events more concrete and apparent to respondents. Second, the cases were used to elicit respondents' opinions about their own likely reactions if they were themselves to be the health care person involved in the incident.

The first case, however, was somewhat different being a translation of a short case used in a survey published by Hingorani et al. (1999) - confer below. The reason for including this case as the very first was to obtain a point of reference allowing us to compare directly between our sample and the interesting results of Hingorani et al. The next three cases were designed to illustrate a near-miss event (the error is captured before any consequences develop or even contact is made with the patient); a case involving a relatively mild injury and, finally, a case involving a relatively severe injury.

The second part seeks to elicit from respondents their requirements to the basic properties of a reporting system[8], asking them about their preferences with respect to anonymity, confidentiality, organizational affiliation of person receiving reports etc.

In part three respondents are asked to indicate, for each of a range of potential barriers (a total of 13 are offered) against bringing up an incident with their colleagues or leaders, their agreement or disagreement; and, similarly, they are asked about (a) their perception of possible reasons, if any, their colleagues may

[8] Currently, there is no national system for reporting adverse events in Denmark. But a prototype system has been introduced in a few hospital departments and, notably, in the Copenhagen hospitals.

have if they withhold information from patients; (b) how they expect their current leader to react if they report on errors; and (c) their opinions about the most likely general causes of adverse events in hospitals.
Then, in part four, respondents are asked to express their beliefs about what *patients* want if they have become victims of an adverse event[9]

Part five deals with reactions, feedback and care offered to colleagues who have been involved in adverse events and it asks respondents how they think their current department or ward deals with adverse events and errors.
In part six respondents are asked to indicate their [dis]agreement with a range of questions that serve to indicate the extent to which they acknowledge human fallibility, probe their job satisfaction and motivation and their perception of patient safety as a relevant issue in their current work, and finally, the extent to which they find it distressful to consider that they may injure patients.
Finally, in a demographic section respondents provide information about, *inter alia*, their age group, specialty, job function, the extent to which supervision is used in their current ward or department. In order to ensure anonymity, no question was included that could serve to identify a respondent's current department or ward.
Most of the questionnaire items were phrased as closed questions with Likert-type response options (strongly [dis]agree / [dis]agree somewhat / neither agree nor disagree), a few as yes / no/ [don't know]. In addition to the closed questions, there were 13 open-ended questions, usually asking respondents to provide an additional or "other" reason (for instance, for not reporting) than the ones offered in the closed questions.
Finally, the questionnaire was developed in a doctors' version and a nurses' version, the two differing only in trivial aspects.[10].

The survey
The questionnaire was distributed to doctors and nurses in the four counties in Denmark described in the Introduction in February-March 2002. A total of 4108 copies were distributed and 2031 were returned. When subtracting the copies (89) that were returned with addressee unknown/moved, this yields a modest rate of response of only 51% (response rate of doctors: 46%; nurses: 53%). The deadline for returning responses was set at two weeks after receiving the questionnaire.[11] Data entry has been made 2008 questionnaires (703 doctors and 1297 nurses) - a few completed questionnaires (23) arriving too late to be included.

Reactions to cases of errors and/or adverse events
Four cases were prepared for the questionnaire, the first being adopted from the study published by Hingorani et al. (1999). The remaining three differed in terms of outcome. The Hingorani et al. case is the following

Table 2 - the Hingorani et al. case
Please read the following story (which is typical but fictional):

[9] Results of this section will subsequently be compared with results of a planned questionnaire directed at patients and the general public about their wishes and requirements in similar situations).
[10] In addition, one of the cases (case C) was developed in two versions as well, randomly distributed within the groups of doctors and nurses: one version described the error as occurring due to a busy on call and the other due to simple distraction or absent-mindedness. The details of this variation and the results will be described in future publications, space not allowing us to detail this any further.
[11] It is a matter of concern to the authors that the response rate is so relatively low, notwithstanding that we have no prior expectancy of response bias. We have subsequently received a number of informal comments from doctors and nurses who received the questionnaire, and it appears that the single most important reason why many potential respondents did not complete the questionnaire is that is quite long (12 pages, 108 individual items) and requires quite some effort to complete. So while strenuous efforts were made to make the questionnaire as simple as possible, the "assurance" in the Foreword of the questionnaire - that it will take 30-60 minutes to complete - was an estimate at the lower end of a realistic range. Unfortunately, much of the data collection also coincided with the one-week winter vacation period (schools are closed and families often go on vacation).

Mrs Brown has an operation for cataract. During surgery, there is a complication. The lens capsule breaks and the surgeon has to make a bigger cut than planned, use stitches and put in a different style of lens implant. There is approximately a 1 in 10 chance of her vision being affected by these changes.
The next day, she sees well and is pleased.
Should Mrs Brown be told about the surgical problem? **Yes / No**
If yes, do we discuss the possible consequences? **Yes / Only if she asks / No**

Below we present the results of the Hingorani et al. study (including data from personal communication, Hingorani, 2002) along with the survey results from the Danish hospital doctors and nurses.

Table 3 - data from Hingorani et al. and this survey		**Yes**	**No**	
Should Mrs Brown be told about the surgical problem?	Danish nurses	99%	1%	
	Danish doctors	95%	5%	
	UK ophthalmologists	60%	40%	
	UK patients	92%	8%	
		Yes	**No**	**Only if the patient asks**
If yes, do we discuss the possible consequences?	Danish nurses	92%	0.4%	7%
	Danish doctors	89%	0.4%	11%
	UK ophthalmologists	33%	21%	46%
	UK patients	81%	3%	16%

It is of considerable interest that the results obtained with the Danish sample of doctors (and nurses) are quite different from Hingorani et al.'s results obtained from 48 ophthalmologists attending a regional meeting. We have no idea of what lies behind the fairly large difference in responses between UK ophthalmologists and Danish hospital doctors[12].
In the following we present the three cases that differ in terms of outcome. In the first case, there is a near-miss event and no harm at all to the patient. In the second case, the patient is somewhat harmed, but only temporarily; whereas in the final case, the harm to the patient is probably severe and permanent.
First we list the cases and then we summarise results.

Table 4 - a near miss, a relatively mild outcome and a relatively severe outcome case

Case B: A patient in the internal medical ward has an I.V. in his left arm providing an infusion of isotonic glucose. When you are about to give antibiotics you realise that the I.V. has become blocked. You now want to rinse the I.V. as the infusion is not running and you want to flush the i.v. using saline which you draw from a capped vial into a 20-ml. syringe, placing it on the tip of the venflon. You are just about to rinse it when you look once more at the label on the vial and realise that it contains potassium chloride and not saline. You are aware that this dose of potassium chloride would probably have killed the patient
Case C: A 53-year old male (married, 2 adult daughters, self-employed truck driver) is hospitalised for elective surgery (cholecystectomy). Before his operation the patient will receive a prophylactic anti-coagulant injection as a matter of routine. There are an excess number of patients in the ward, so it is a busy on call. [Doctors: When you are dictating the case notes, you are interrupted several times due to emergency situations. You forget to dictate the anti-coagulant for the 53-year old patient] [Nurses: When you are on your drug round, you are interrupted several times due to emergency situations. You forget to include the anti-coagulant for the 53-year old patient]. The patient develops a deep thrombosis in his leg. He therefore has to remain hospitalised an additional week and will be on sick leave from work longer than planned. It is very unlikely that he will have permanent impairment from the thrombosis
Case D: A 42-year old woman (married, one child, school teacher) is hospitalised in order to receive chemotherapy. The drug has to be given as a continuous infusion intravenously. There is no pre-mixed infusion available in the department and you have to prepare it yourself. While you are preparing the infu-

[12] We are currently planning a study to find out if a respondent's medical specialty may play a role. Possibly Danish ophthalmologists might resemble, in terms of responses to this case, their UK colleagues more than do Danish doctors in general. Thus, it might be speculated that a doctor will assess risks within his or her own specialty differently than risks in other specialties.

sion, you are distracted. By mistake you prepare an infusion with a concentration 10 times greater than the prescribed level.
You do not discover the error until you administer the same drug to another patient later that day. By this time the 42-year old patient has already received all of the high concentration infusion. You are aware that in the long term the drug may impair cardiac functioning. You realize that there is a significant risk that the patient's level of functioning will be diminished and that she probably won't be able to maintain her present work

For each of the cases respondents were presented with a set of statements describing possible actions with regard to (a) reporting the incident, discussing it with colleagues, ensuring that the patient gets required attention etc, and (b) patient information and interaction. The questions were, as far as possible, identical

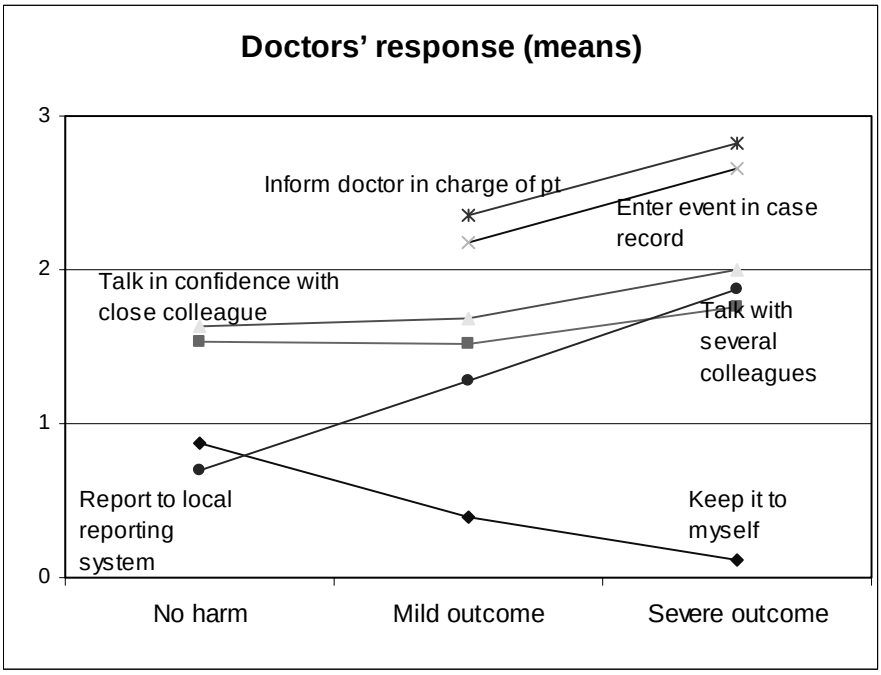
Table 5 - Possible actions after an adverse event

"*Each of the following statements describes a possible action. Please indicate for each item whether you will carry out the action".*				
▪ Keep it to myself that the patient has received 10 times the prescribed level				
▪ Talk in confidence with a colleague about the incident				
▪ Talk to several colleagues about the incident				
▪ Write in the patient's case-record that the patient has received 10 times the prescribed level				
▪ Inform my superior or the doctor in charge of the patient				
▪ Report the event to the local reporting system [mark this item *only if* you do have such a system]				
"Please indicate your response by marking one of the following options:"				
Yes, definitely	Yes, probably	Probably not	Definitely not	Don't know

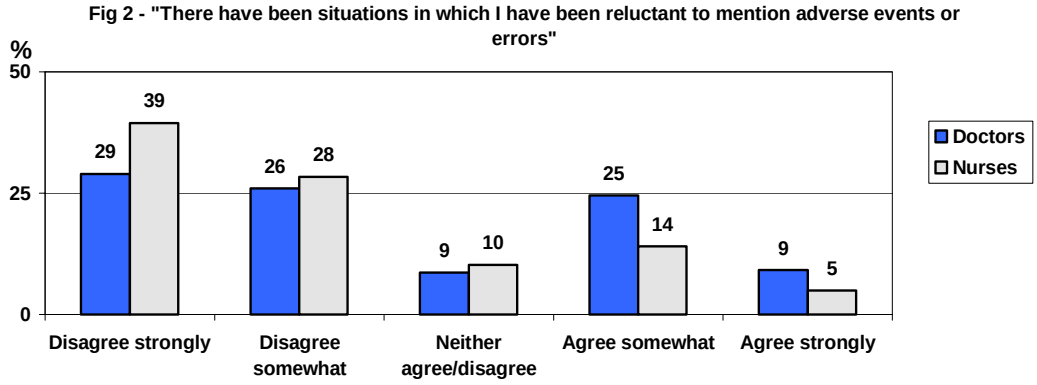
across the cases, but some of the possible actions were inappropriate for Case B involving no injury. For reasons of space we here reproduce only the questions of Case D - see Table 5. Each question was to be answered by marking one of the response options.
Below are shown the mean responses of the doctors (nurses' responses, being largely comparable are left out for reasons of space). The numbers 1-4 on the y-axis correspond to the four response options excluding "Don't know". In general, doctors as well as nurses show a marked willingness to bring up the event with their leader or the doctor in charge of the patient when the patient is affected. In addition, there is a significant trend that the more severe the case is, the greater is the willingness not to keep it to oneself and to discuss the event with colleagues. We also asked respondents whether they agreed or disagreed that there have been situations in which they have been reluctant to mention an adverse event or error (in which they had been involved). In Figure 2 below we show the results comparing nurses and doctors. The difference between the two groups is highly significant (Mann-Whitney's test using asymptotic inference, $p<0.0000$). The difference between the two groups may possibly be ascribed to nurses' greater tendency to discuss events with their colleagues and to the fact that doctors' tasks will tend to bring them into adverse events more often than nurses.

Reasons for not reporting
Below we show respondents' answers to proposed reasons for *not* reporting an adverse event or error. Respondents were asked to indicate their agreement or disagreement with the suggested reasons for not reporting by marking on a 5-point Likert scale (plus an additional "don't know" option). As will become apparent, neither doctors nor nurses rate *any* of the 13 proposed potential barriers very highly, in fact the *highest median rating* on any of them is "Neither agree nor disagree". For the sake of overview we provide mean ratings made by doctors and nurses (but the test of significance between the groups on individual items [Mann-Whitney] is rank based and, therefore, makes no assumptions about equal distance between response options).

Figure 1 - Doctors' responses across three cases



0 = definitely not; 1 = probably not; 2 = yes, probably; 3 = definitely yes



These results differ in several respects from those published by Vincent et al (1999). Thus, Vincent et al. found that the four most potent reasons their respondents (doctors) gave for not reporting were:

1. Often the outcome makes it unnecessary to report
2. Reporting entails increased workload
3. Junior staff may be blamed
4. The doctor involved is too busy or forgets to report.

Much further down the list in Vincent et al.'s study do we find barriers that refer to personal reasons such as "worry about litigation", "my colleagues may be unsupportive", "worried about disciplinary action".
In contrast, our respondents consider that the risk that the press might write about it as the most important barrier (doctors significantly more so than nurses). And not far behind is cited the fear of "reprimand" (where the small observed difference between nurses and doctors is non-significant).

Table 6 - Potential reasons for holding back on reporting an adverse event

Suppose that you became involved in an adverse event or an error. Which among the following factors might lead you to holding back on reporting? (Mean rating of [dis]agreement is shown)	Doctors	Nurses
It might get out and the press might start writing about it ***	2,9	2,7
We have no tradition in my department for bringing up adverse events/errors ***	2,8	2,4
I do not wish to appear as an incompetent doctor [nurse] ***	2,7	2,5
I might receive a reprimand	2,5	2,4
It might have consequences for my future employment or career	2,5	2,5
I don't know who is responsible for bringing up adverse events/errors *	2,4	2,1
It is too much bother bring up adverse events/errors ***	2,4	1,9
The adverse event/error may become reported to the medical licensing board ***	2,4	2,5
One does not feel confident about bringing up adverse events/errors in our department **	2,4	2,0
The patient may file a complaint ***	2,4	2,2
Bringing up adverse events/errors will not lead to any improvement in our ward ***	2,2	1,8
When I am busy I forget to bring up adverse events/errors ***	2,1	1,8
It wouldn't help the patients that I bring up my own events/errors ***	1,8	1,7

Mann-Whitney's test using asymp. inf.: * = $p < 0.05$; ** = $p<0.01$; *** = $p<0.001$.

Respondents were asked to rate the importance of each of the potential reasons provided on a Likert-type 5-point scale going from Strongly disagree to Strongly agree but including also "Don't know" option. The ratings have been transformed to numbers so that Strongly disagree = 5; slightly agree = 5; neither agree/disagree = 3; slightly agree = 2; and strongly disagree = 1.

To illustrate the extent to which respondents are unwilling to name *any* of the suggested reasons as *a* reason for not reporting we list the proportions who agree and disagree with the *most* as well as the *least* compelling reason for not reporting.

Table 7 - the most and the least compelling reason for not reporting

Potential reason for not reporting	Doctors		Nurses	
	% Agree	% Disagree	% Agree	% Disagree
The press might write about it	25	60	25	58
It wouldn't do the patient any good	11	82	6	84

But, it might be objected, perhaps the reasons offered as potential barriers are not particularly apt or they do not match those that doctors and nurses really have for not reporting, when they do in fact refrain from reporting. Yet, only 10% of the nurses and 13% of the doctors have suggested other reasons for not reporting in the open-ended question about "other reasons". Moreover, our results do match, in terms of rate of agreement and therefore estimated strength of proposed barrier, those obtained by Vincent et al.

We seem compelled to conclude, therefore, that a large majority of doctors and an even larger majority of nurses do not find that there are *any* reasons or factors that do in fact act as significant barriers against

reporting. Respondents' replies to the question cited in Figure 2 seem to support this conclusion. Thus, most doctors and the vast majority of nurses do not think they are reluctant to bring up adverse events or errors.

Preferred Models of Reporting

One of chief difficulties in developing a questionnaire for eliciting attitudes and requirements to models of reporting is that several of the issues are inherently complex. Thus, the focus group interviews (see introduction) showed that subjects had difficulties in imagining to what extent a confidential system might protect their anonymity outside their department. Moreover, potential respondents will have different understandings of key concepts and we should not expect respondents to have the same notion of, say, "anonymity", "reporting", "error", and "adverse event".

Therefore, when seeking to elicit respondent's views about models, it was unavoidable to insert a few terminological statements. Thus, in the following (admittedly rather wordy) Table 8 respondents are asked to comment on models of reporting that are distinguishable in terms of the protection of the identity of the reporting nurse or doctor.

Table 8 - Models of reporting

Please indicate in the following table for each of the tree models below whether you find it acceptable as the basis for a future reporting system. In questions 11 and 12 you will be asked to indicate which person you wish to be the "recipient". (Insert one tick for each item below)

Model of Reporting System	Degree of anonymity or confidentiality	Possibility of clarification of event/error beyond initial report	Possibility of personal feedback to reporting doctor/nurse	**0Please indicate your degree of acceptance**			
				0Highly acceptable	Somewhat acceptable	Somewhat unacceptable	Not at all acceptable
1. Anonymous (reporter unknown)	*Name and identity unknown to everybody*	*Clarification not possible due to anonymity*	*Feedback not possible due to anonymity*	☐	☐	☐	☐
2. Confidential (reporter known only by recipient)	*"Recipient" knows identity of reporter and may not transmit this to others*	*Clarification possible*	*Personal feedback*	☐	☐	☐	☐
3.Conditionally confidential (confidentiality is broken in case of gross negligence)	*"Recipient" knows identity of reporter and will transmit this to regulators only if there is a violation of criminal law*	*Clarification possible*	*Personal feedback*	☐	☐	☐	☐

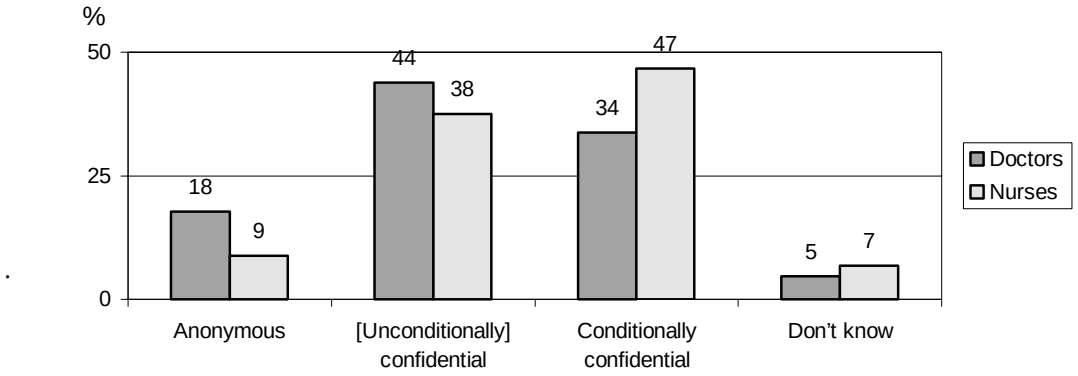
The purpose in eliciting from respondents their response - in terms of acceptability of each of the three models described - was, first, to assess the degree of resistance which health care staff might have with respect to each of models, and second, to introduce in a perspicuous form the key characteristics of each of the models. In addition, respondents were asked to indicate their preference and simply state their choice of model - see Table 9.

Table 9 - Your preferred model of reporting

Which of the above three models would you prefer?	*Anonymous*	Confidential	Conditionally confidential	Don't know
	☐	☐	☐	☐

Results of the respondents' preference are given in Figure 3 below, showing that the anonymous model is by far the *least* preferred model, whereas the doctors and nurses are somewhat divided with respect to which confidential model they prefer.

Figure 3 -"Which of the above models would you prefer?"



.

At the same time, a majority among nurses and about half the doctors expressed the view that the anonymous model was not at all acceptable or somewhat unacceptable.

Table 10 - respondents' resistance against individual models

%	"Not at all acceptable / somewhat unacceptable"		
	Anonymous model	[Unconditionally] confidential model	Conditionally confidential model.
Doctors	49%	13%	32%
Nurses	69%	11%	17%

Conclusion

The survey indicates that Danish doctors and nurses consider all of the reasons offered for not reporting as having little or no weight. At the same time, they express a marked willingness *not* to keep adverse events to themselves if they have an impact on the patient. Their willingness to report on adverse events therefore stands in need of a system of reporting and organizational learning, which is currently lacking in most parts of Denmark. However, experience from recently introduced pilot schemes (in particular the recent patient safety programme of the Association of Copenhagen Hospitals) is encouraging - as indeed are current efforts at establishing an efficient national reporting system.

At the same time, results of the survey have shown that doctors and nurses are overwhelmingly in favour of a confidential type system - and not a strictly anonymous one - that permits feedback to the individual doctor or nurse and allows interviews and interaction with the person reporting. It is noticeable, however, that doctors tend to favour an unconditionally confidential system that requires that all information about adverse events be entirely de-identified in all cases with no exceptions possible.

In future publications, we plan to review differences in terms of age groups and specialties with regard to these and other aspects covered by the questionnaire, space being strictly limited on this occasion.

Acknowledgements
The authors gratefully acknowledge useful comments by Henriette Lipczak, DSI, Morten Freil, Glostrup Hospital, and Jørgen Hansen, The National Board of Health, to early versions of the questionnaire. We also thank the health care administrations in Copenhagen, Roskilde, Frederiksborg and Ringkøbing counties as well as the doctors and nurses and the ward and department leaders who have contributed to the study and whose efforts have made the study possible.

References

Hingorani, M, Wong, T, and Vafidis, G (1999) Patients' and doctors' attitudes to amount of information given after unintended injury during treatment: cross sectional, questionnaire survey. *BMJ.* 1999;318:640-641.

Itoh, K, Abe, T. and Andersen, HB (2002) A Survey of Safety Culture in Hospitals Including Staff Attitudes about Incident Reporting. Workshop on Investigation and Reporting of Incidents and Accidents
17th - 20th July 2002, University of Glasgow.

Vincent C, Stanhope N, Crowley-Murphy M. (1999) Reasons for not reporting adverse incidents: an empirical study. *Journal of Evaluation in Clinical Practice* 1999; 5(1):13-21.

Development of a Region-Wide Process for the Investigation of In-Hospital Deaths

J. M. Davies (1) and B. Young (2),

(1) Department of Anesthesia, University of Calgary, Foothills Medical Centre
(jdavies@ucalgary.ca)

(2) QIHI, Calgary Health Region, Calgary, Alberta, CANADA T2N 2T9
(Barb.Young@calgaryhealthregion.ca)

Abstract: This paper describes a new method of mortality review now in use in acute care hospitals throughout the Calgary Health Region (CHR), as well as a description of the evolution of the CHR's mortality review process. The reasons for investigating in-hospital deaths are described, as are certain aspects of procedural importance and analysis.

Keywords: mortality reviews, review process, investigations

Introduction - Calgary Health Region

Background: Calgary is the largest city in the province of Alberta and one of the fastest growing cities in North America. Before 1985, Calgary's economy revolved around the single industry of gas and oil, with changes in this industry following a pattern of boom and bust. However, economic diversion following the Winter Olympics in 1988 allowed Calgary to break from its reliance on gas and oil, and become an urban centre linked economically, socially, and culturally to the world.

Healthcare regionalisation: This economic 'revolution' was followed by major changes in healthcare in the province. The greatest change came through regionalisation, a process that included closure of many large and small hospitals. At the time of regionalisation in 1994, the Calgary Health Region (the largest in the province) had seven acute care hospitals. Each organization had:

- separate medical staff, with doctors tending to practice at only one or two sites;
- different structures, policies, procedures and practices;
- different cultures; and
- different, incompatible and outdated information technology systems.

There was neither coordinated planning nor good data about population health needs or even about current services. Most available data were based on acute care services and there were no means of measuring the efficiency, let alone the effectiveness of most services. Health services in Calgary bore little resemblance to any form of an integrated health care system.

To meet the requirements of operating within available funding from the Alberta government, the CHR had to take drastic action. The initial business plan proposed:

1. merging all independent organizations under a common structure, including Administration;
2. closing three of the seven acute care hospitals;
3. focusing on moving and increasing services to the Community; and
4. creating a regional medical staff structure by bringing together the doctors of the seven different medical staff associations and establishing one medical leadership team.

Despite the magnitude of these proposals, all were implemented without interruption of health service delivery to the region's population. Although the legislative and administrative restructuring required by regionalization have been completed, the 'cultural' changes of regionalization are still on going, as for example, has occurred with mortality review

Mortality review - History of the process in Calgary

Before regionalisation: Originally, each hospital's medical staff carried out mortality reviews - as the members saw fit. Not only were the reviews confined to each acute care facility, but also each site had a different questionnaire and process, with no comparison of results with the other acute care sites. This was true even for similar clinical divisions, with each death in each Clinical Division viewed separately. For example, the Cardiology department at one acute care site did not compare their data with other Cardiology departments within the city. Data were not collected, aggregated or analyzed. There was no effort to identify common trends in deaths or even any common contributing factors as most data were collected on paper and the results were usually not computerized. As one Clinical Department Head said "Our forms collected dust, not data".

Of the seven hospitals, the Foothills Medical Centre, the major trauma and tertiary centre, had the most refined process for mortality review. Originally a Death and Tissue Audit Committee, the process at the Foothills included review not only of deaths but also of various tissues removed at the time of operation, e.g., the 'normal' appendix' rate. Over the years, the goals of the committee evolved, as its role in reviewing tissue decreased. This change was based upon the recognition that tissue review was best done by those who removed the tissues, e.g., the surgeons, and those who then examined them, e.g., the pathologists. At about the same time, the importance of systemic factors as contributors to deaths became increasingly recognized. The Committee developed and used a new questionnaire to track human and system factors for every death in the hospital. Although the results were not collected electronically, a database was developed to track trends.

After regionalisation: When regionalisation was first undertaken, the Medical Advisory Board of the CHR recognized the need for central data collection and asked that a region-wide method for death review be developed for all acute care hospitals. Initially there was no consensus for a regional process amongst the Medical Staff. This was in part due to an intellectual concern about the value of mortality review, and in part to cultural differences among doctors from the different sites. With continued pressure from the Chief Medical Officer's office, a new questionnaire was developed over the course of a year, with input from members of other hospital mortality committees. The Medical Advisory Board approved the questionnaire in mid-2000. At that time, there were major changes at the senior management level, including a restructuring of the Chief Medical Officer's office. Implementation of the new form did not advance for over a year. During this time, divisional Mortality Reviews continued, using the old forms and processes. In July 2001, Quality Improvement Health Information (QIHI), a portfolio of the Chief Medical Officer's office, assumed responsibility for mortality review. The new questionnaire was then implemented, with full support from the Clinical Department Heads and Division Chiefs. In the autumn of 2001, more revisions were made to the questionnaire, in the form of inclusion of addition demographic and statistical information, such as length of stay and a Code (resuscitation) level for patient care. The questionnaire was fully implemented in April 2002.

Objectives: The goal of mortality review is to collect and analyse data that reflect the contribution of systemic factors to the death, with the aim of improving care for other patients. A questionnaire is used to ensure that the review is conducted systematically, with the inclusion of previously identified factors. A region-wide questionnaire is to ensure that problems related to the structure and process of care in one department are not treated as isolated problems, and common only to that department. The role of individual caregivers is minimised, as issues of competence are dealt with at a departmental level.

Process: When a patient dies, a decision is made as to whether or not the review should be carried out under the auspices of the Critical Incident Review Committee (CIRC), a sub committee of the Medical Advisory Board. The criteria for a CIR review include:

- any process variation resulting in a death; that is, performance that varied significantly from established guidelines or standards for the implementation of care or a service;
- any death where care was provided by more than one department or clinical program/ area.
- any unexpected and unusual occurrence, including death

If the case does not fit those criteria, then the process of mortality review continues. Most departments review deaths on a monthly basis, although the division of General Surgery carries out weekly reviews. The first step is the generation by the Health Records Department of a list of cases to be reviewed. Included in this step is the determination of which department (or division) the patient 'belonged', i.e., the 'most

responsible' department. This will either be the service that cared for the patient the longest, or the service that used the most resources in caring for the patient during the admission. The list is then forwarded to the Division or Department Chiefs. The next step is the departmental (or divisional) process of review. The Chair of the Division or Department Audit Committee contacts the Health Records Department to have the charts pulled and ready for the monthly meeting. The review is carried out in whichever way the department has determined best suits its needs. This includes asking departmental specific questions. For example, the Department of Family Medicine includes in its review consideration of charting of the doctor's interaction with the family. The Department of Anesthesia uses the death review process as part of a rotation in Quality for trainees, who are encouraged in their review to 'look beyond the doors of the Recovery Unit' and to review the postoperative care of the patient on the ward. The Department of Emergency Medicine reviews all cases where a patient died within 72 hours of treatment in the Emergency Department. After completing this department-specific review, the third step is completion of the region wide questionnaire, which is then forwarded to QIHI for data entry, analysis and evaluation.

Questionnaire: The top part of the single-page questionnaire allows the recording of a number of details. These include: the patient's name, hospital number, age, sex, length of stay, Code level (1-3), date of death, date of review and 'most responsible' medical or surgical department.
In addition to the demographic details, four questions are asked. The first question asks about the autopsy – if one was requested and if one was carried out. In addition, reviewers are asked if the cause of death, or diagnosis at the time of death and the cause of death according to the autopsy coincided.
The second question asks if death was expected at the time of admission and at the time of death. (Reviews of stillbirths exclude this question.) This question forms a triage-point in the review. If death was expected on admission to hospital and at the time of death, and if the reviewers did not have any concerns about system issues in the delivery of care in the acute care setting, then no further questions need to be answered.
The third question concerns the preventability of the death. Reviewers are asked to consider preventability if inpatient care had been different during the admission, for example, if an operation had not been performed, or if pre-hospital care had been different, for example, if the family had been able to call the ambulance earlier.
The fourth question asks for identification, if any, of 'deficiencies'. These deficiencies relate to:

- the patient (pre-hospital condition, age, sex, co-morbidities);
- the personnel (number, training, experience);
- the equipment (number, presence/absence, functioning);
- the workplace environment (size, lighting, ventilation, workspace-layout);
- the organization (corporate structure, budget, policies and procedures); and
- the regulatory agencies (budget, rules & regulations).

A fifth question is planned, which will address the appropriateness of care at the time of death. In other words, 'did the patient die well?'. This question is particularly important for the review of care provided by the Division of Palliative Care and the Department of Family Medicine.

Data analysis: In April 2002, the Calgary Health Region implemented a new software program for Risk Management activities, including the Mortality Review process. This program facilitates the calculation of trends in mortality rates – among different hospitals and within clinical divisions. It is therefore possible to determine, for example, if the mortality rate for a given cause of death is higher at the tertiary centre or at one of the secondary centres, or if there is a seasonal component (see Figures 1 and 2). However, because one of the hospitals within the Region is a Children's Hospital, these trends are therefore compared only amongst the three adult sites.
Both Figures 1 and 2 show that there was a higher than average mortality rate between November and January. This was considered to be related to complications of influenza.

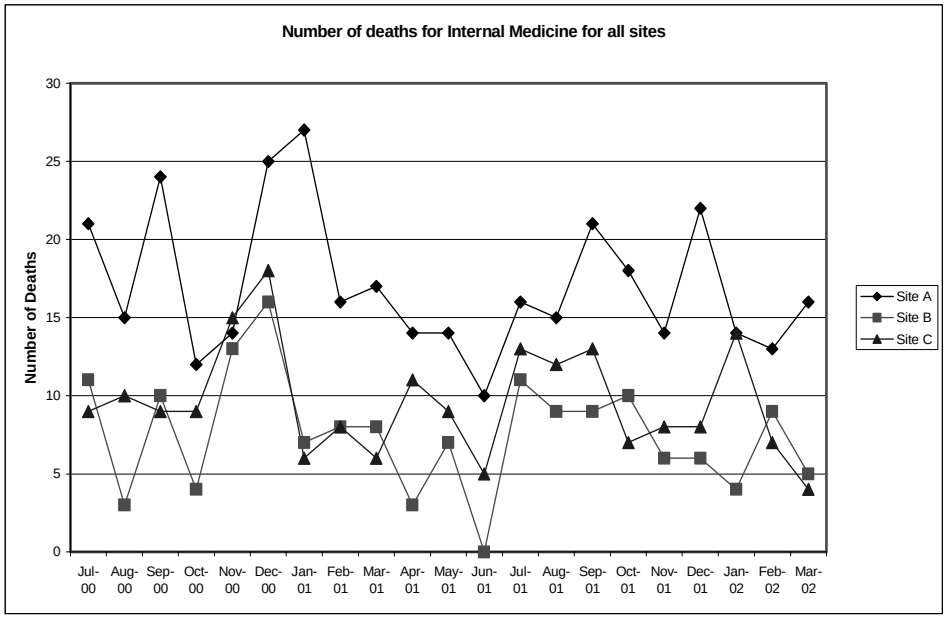


Figure 1 - Internal Medicine Mortality Rates by Hospital Site

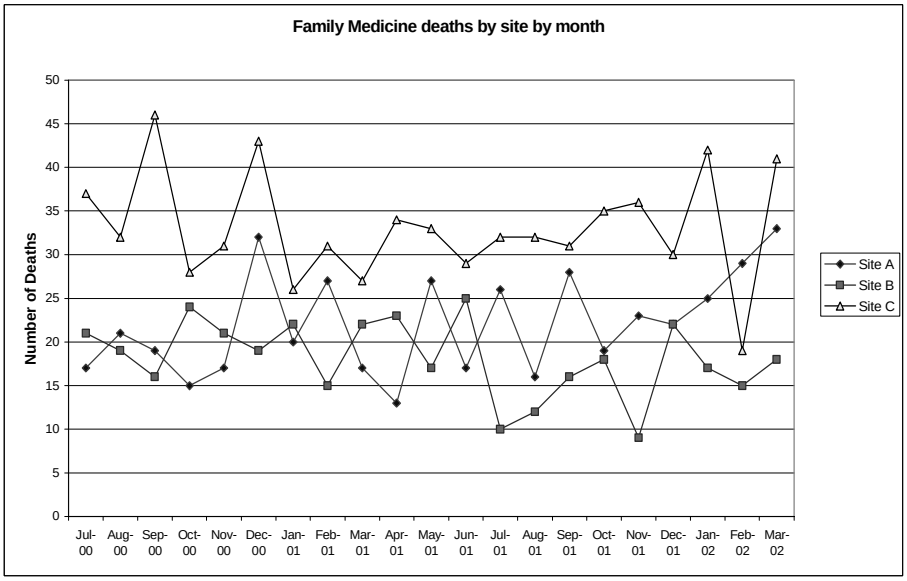


Figure 2 – Family Medicine Deaths (Includes Palliative Care Patients) by Hospital Site

Future data analysis will include making comparisons within departments over time and using control charts to identify normal and special variance. If a special variance is identified, then more intensive analysis will be implemented to determine all of the contributing factors. These factors may include:

- budget cuts;
- new equipment and drugs; and
- new employees, including trainees, residents and consultants.

Also of interest is determining if there is any relationship between the implementation of new procedures / programs, or changes in current procedures / programs, and mortality rates. Table 1 outlines the total population for the CHR since regionalisation, as well as the total number of inpatient deaths each year. Although the population has increased by 15%, the mortality rate for inpatients has remained relatively constant.

Table 1 – CHR population, number of inpatient deaths, and crude mortality rate (CMR) (1995-2001)

Year	CHR population	# inpatient deaths	CMR
1995	816,936	2002	0.245%
1996	832,030	2204	0.265%
1997	856,413	1939	0.226%
1998	887,933	1993	0.224%
1999	916,481	2009	0.219%
2000	936,205	2012	0.214%
2001	953,895	2005	0.210%

In addition to tracking the total number of deaths, the autopsy rate for each clinical department is also tracked. Figure 3 shows that the autopsy rate for the Department of Family Medicine averaged approximately 5% over 21 months, whereas the rate for Internal Medicine remains at approximately 25%. It should be noted that for the first quarter of 2000/01, there were no deaths in the Division of General Surgery.

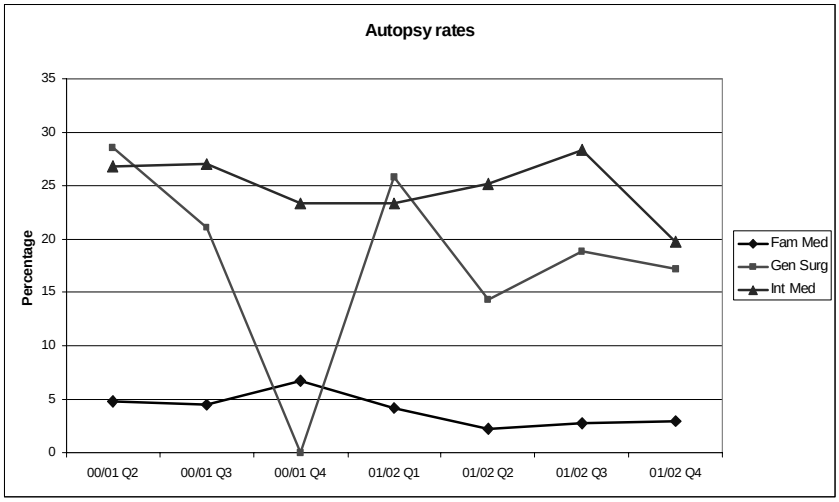


Figure 3 - Autopsy Rates for Three Departments

Data relationships

Another benefit of the newly acquired software is the ability to cross-reference each death to regional incident and critical incident reports. The software also allows tracking of the patient's name and hospital number through the patient complaint database. This function facilitates determining if concerns were expressed about the patient's care and treatment before death. In addition, comparisons are possible between the CHR's mortality rate and provincial and national rates.

Discussion

Why carry out mortality review?: Mortality review can be considered to be the most traditional form of medical audit. For centuries, doctors have looked to their dead and dying patients in the hope of learning from them. However, over the past few years there have been intermittent suggestions that any improvement in the quality of care is better served by review of morbidity or by focused audits of particular

aspects of care. While this argument carries a certain validity, mortality review does hold some advantages over that of morbidity. One of the most important is that definitions of morbidity are debatable but death, or "flower-bedecked failure", is not (Owens & Spitznagel, 1980). Thus, mortality review avoids the problems associated with the variable definitions for morbidity. At the same time, mortality review provides one form of measurement of outcome of various interventions that can then be compared with outcomes for similar interventions elsewhere (Bayly, 2001).

Investigation of all deaths: By reviewing all deaths, the problem of the zero numerator is also avoided. For certain specialties, the mortality rate maybe so low as to appear to be zero, for example, elective cosmetic surgery. Interpretation of 'zero numerators' can erroneously lead to the assumption that 'if nothing goes wrong then everything is all right'. Hanley and Lippman-Hand (1983) have emphasized that zero numerators neither imply 'no risk' nor make inferences about the actual size of the risk. Part of the 'problem' with zero numerators lies with human psychology. For many, a theoretical risk is not one which has not yet occurred but one which will never occur. In addition, when given an expected rate, the numerator will be examined strongly and the denominator ignored. Collecting longitudinal data for every death in every department will help to ensure accuracy of mortality rates.

Classification of deaths: One major consideration when investigating deaths is how to classify them. In 1949, Edwards, Morton, Pask and Wylie were appointed to a committee to study voluntarily submitted reports of anaesthetic-associated deaths. One thousand reports were received over sixty-six months. Some centres contributed regular reports while others did not submit any, although reports about deaths from these institutions were received indirectly. Perhaps the most lasting contribution of this study has been the 'Edwards' classification of perioperative deaths (Edwards et al, 1956). This classification (Table 2) has been used by a number of other investigators, in multi-institutions over short and long periods (Holland, 1970; Holland, 1984; Holland, 1987; Warden & Horan, 1996; Cohen et al, 1992a; Cohen et al, 1992b).

Table 2 - Edwards Classification of Operative Deaths (Edwards et al, 1956)

1.	Where it was reasonably certain that death was caused by the anaesthetic agent or technique of administration, or in other ways coming entirely within the anaesthetist's province.
2.	Similar cases, but in which there was some element of doubt as to whether the agent or technique was entirely responsible for the fatal result.
3.	Cases in which the patient's death was caused both by the surgical and anaesthetic techniques.
4.	Deaths entirely referable to surgical technique, e.g., uncontrolled haemorrhage.
5.	Inevitable deaths, e.g., cases of severe general peritonitis, but in which the anaesthetic and surgical techniques were apparently satisfactory.
6.	Fortuitous deaths, e.g., due to pulmonary embolism.
7.	Cases that could not be assessed despite considerable data.
8.	Cases on which an opinion could not be formed on account of inadequacy of data.

As shown here, the classification focuses on the contribution of the doctors involved in the care of the patient at the time of death, as well as the patient's underlying condition. Thus, as one would expect from a classification of this vintage, there is little recognition of the role of any latent conditions (Reason, 1997), such as missing equipment, inadequate numbers of nurses or organisational culture. By considering the underlying system contributors, seeking out the 'bad apple' (Berwick, 1989) should be avoided.

Legal protection: Finally, mention must be made of the legal protection granted to mortality reviews under the Alberta Evidence Act (Alberta Evidence Act, 2000). Specifically, Section 9 states that a witness cannot be asked and does not have to answer questions about quality assurance activities, nor produce any written material from those activities. Under the terms of the legislation, Quality Assurance is a planned or systematic activity for the purpose of study, assessment or evaluation of the provision of health services with a view to the continual improvement of quality of health care or health services, or the level of skill, knowledge and competence of health service providers. This legislation allows healthcare providers to review a case without the fear of being subpoenaed for their opinions and comments and is a major contributor to the success of mortality review in the Calgary Health Region.

Conclusions**:** The mortality review process in the Calgary Health Region has progressed from seven separate site-specific initiatives to one coordinated regional practice. Aggregated data allow better analysis of system issues related to patient care. Although for the most part mortality review is currently paper-based, advances toward an electronic patient record will facilitate an electronic review process. This will simplify transfer of patient details from the Health Records Department, to the reviewing physicians, and then forwarding of the completed reviews to QIHI for analysis. However, no matter what the format, this new region-wide mortality review process will help in the identification of system-based problems and improvement in the care of future patients.

References

Alberta Evidence Act (2000), RSA 2000, Section 1, Chapter A-18, Competency of Witnesses. 9. Quality Assurance Records. Government of Alberta, Queen's Printer, Edmonton

Bayly PJM (2001). Resuscitating Audit. Anesthesia 56:717-9

Berwick DM (1989). Continuous Improvement as an Ideal in Health Care. New England Journal of Medicine 320:53-6

Cohen MM, Duncan PG, Tweed WA, Biehl D, Pope WDB, Perry M, Merchant RN (1992). The Canadian Four-Centre Study of Anaesthetic Outcomes: I. Description of Methods and Populations. Canadian Journal of Anaesthesia 39:420-9

Cohen MM, Duncan PG, Pope WDB, Biehl D, Tweed WA, MacWilliam L, Merchant R.N (1992). The Canadian Four-Centre Study of Anaesthetic Outcomes: II. Can Outcomes Be Used to Assess the Quality of Anaesthesia Care? Canadian Journal of Anaesthesia 39:430-9

Edwards G, Morton HJV, Pask EA, Wylie WD (1956). Deaths Associated with Anaesthesia: A Report on 1,000 Cases. Anaesthesia 11:194-220

Hanley JA, Lippman-Hand A (1983). If Nothing Goes Wrong, is Everything All Right? Journal of the American Medical Association 249:1743-5

Holland R (1970). Special Committee Investigating Deaths Under Anaesthesia. Report on 745 Classified Cases, 1960-1968. The Medical Journal of Australia 12:573-94

Holland R (1984). Anesthesia-Related Mortality in Australia. *In:* Analysis of Anesthetic Mishaps. Pierce E.C. & Cooper J.B. (eds). Little, Brown and Company, Boston. International Anesthesiology Clinics 22;61-71

Holland R (1987). Anaesthetic Mortality in New South Wales. British Journal of Anaesthesia 59:834-41

Owens WD, Spitznagel EL Jr (1980). Anesthetic Side Effects and Complications: An Overview. *In:* Anesthetic Side Effects and Complications – Seeking, Finding and Treating. Owens WD (Ed.) Boston: Little, Brown and Company

Reason J (197). Managing the Risks of Organizational Accidents. Aldershot: Ashgate Publishing Company

Warden JC, Horan BF (1996). Deaths Attributable to Anaesthesia in New South Wales, 1984 – 1990. Anaesthesia Intensive Care 24:66-73

A Survey of Safety Culture in Hospitals Including Staff Attitudes about Incident Reporting

Kenji Itoh[(1)], Toshiko Abe[(2)] and Henning Boje Andersen[(3)]

(1): Tokyo Institute of Technology, Tokyo, Japan
E-mail: ken@ie.me.titech.ac.jp
(2): Tokyo Medical and Dental University, Tokyo, Japan
E-mail: abet.ns@tmd.ac.jp
(3): Risø National Laboratory, Roskilde, Denmark
E-mail: henning.b.andersen@risoe.dk

Abstract: The present paper reports the results of a questionnaire-based survey of *safety culture* in hospitals including the attitudes and perceptions of medical staff's about the reporting of adverse events. Approximately 600 responses have been collected from doctors, nurses and pharmacists working in five hospitals in Japan. The questionnaire has been adapted from Helmreich's "Operating Team Resource Management Survey" and contains, in addition, questions about respondents' reporting of their own errors and their information to patients who have suffered adverse events. Doctors are significantly more willing to report in the severe outcome case than in the milder one whereas nurses' willingness to report hardly changes at all and matches that of the doctors' in the severe case. Moreover, doctors express a slightly greater willingness to inform the patient in the severe case. We have compared rates of incident reporting and questionnaire responses and have found that several safety cultural aspects, particularly acknowledgement of human errors and power distance, are correlated with the actual reporting rate of incidents. Based on these results, we suggest that realistic recognition of human errors and a small power distance are of critical importance for actual reporting behaviours and, in turn therefore, for patient safe.

Keywords: safety culture, patient safety, incident reporting, and questionnaire-based survey

Introduction

It is widely recognised that human error is the predominant cause of accidents not only in human-machine system operations in industry, e.g., aviation, maritime operations, and in nuclear power plants, but also in health care and in particular in hospitals (Kohn et al., 1999). Similarly, in recent decades, organisational factors have been recognised to be of great importance for safe operations (e.g., Reason, 1993). Thus, it has been observed that organisational problems are frequently latent causal factors that contribute to the occurrence of human errors made by frontline personnel; and similarly, it has been pointed out that the dominant type of *contributing* causes of major accidents involve the organisations that themselves shape the safety culture or climate within which the employees operate (Hee et al., 1999; Reason, 1997).

In an influential publication by the IAEA, *safety culture* was defined as "….. that assembly of characteristics and attitudes in organisations which establishes that, as an overriding priority, ….. safety issues receive the attention warranted by their significance" (INSAG, 1993). This publication goes on to observe that safety culture "….. is attitudinal as well as structural, relates both to organisations and individuals, and concerns the requirements to match all safety issues with appropriate perceptions and actions". The characterisation of safety culture in terms of management policies and commitment, organisational structures and employee and group attitudes has been widely adopted. Thus, ACSNI defines the notion as "….. the product of individual and group values, attitudes, perceptions, competencies and patterns of behaviour that determine the commitment to and the style and proficiency of an organisation's health and safety management" (ACSNI, 1993). In other words, safety culture is coupled not only to management's commitment to safety, its communication style and the overt rules for reporting errors but also to employees' motivation, morale, perception of errors and attitudes towards management and factors that impact on safety, (e.g., fatigue, risk taking and violations of procedures – cf. Andersen, 2002).

To study the relationship between safety culture and operational safety, a number of projects have been conducted to uncover individual organisation's safety culture in high-tech industries such as aviation, maritime, railway and process control industries. Such studies are typically based on the assumption that the quality and safety with which operators accomplish their tasks are affected not only by their

professional technical competence and skills but also by their attitudes to and perceptions of their jobs, their organisation and management (e.g., Andersen et al., 1999; Helmreich & Merrit, 1998). For example, operators' attitudes have been found to be important indices of safety performance having been shown to correlate with incident/accident rates in railway operations (Itoh & Andersen, 1999; Itoh et al., 2000; 2001). In addition, since attitudes may be measured before accidents take place the method of measuring safety attitudes may well be an important proactive means of ascertaining risk levels, especially in domains where incidents and accidents are rare.

Moving to the hospital domain, one would expect that medical activities share many characteristics of the above-mentioned high-tech human-machine system operations, and that *patient safety* will similarly be affected by safety culture. Therefore, it would seem useful to adapt some of the research methods and survey techniques that have been developed for application in the high-tech human-machine system domains to investigating human factors aspects of patient safety.

In the present study, a *questionnaire-based survey* was performed to identify characteristics of safety culture in hospitals. The questionnaire responses have been compared with those of ship officers that have been collected in previous studies using a similar type of questionnaire. As part of the present survey, we seek to uncover doctors' and nurses' attitudes to reporting incidents and own errors and to informing patients who have been injured by medical error. These data have in turn been compared with the data on the other elements of safety culture as well as with the incident reporting rates, i.e., rates of staff's adverse incidents and rates of their informing the patient about an event, obtained independently from one of the hospitals surveyed. Based on the integrated results of the questionnaire survey, we briefly discuss some current issues of safety culture in Japanese hospitals as well as factors that jeopardise patient safety.

Questionnaire and Respondents

The questionnaire comprises five parts and has an additional demographic section where respondents fill in their department or ward specialty, position, experience and age group. Four of the five parts of the questionnaire have been adapted from Helmreich's "Operating Team Resource Management Survey" (Helmreich & Merritt, 1998). The Helmreich questionnaire has several derivatives focusing on specific domains and allows us to compare the results with ones derived from other domains, e.g., maritime operations and aviation (e.g., Andersen et al., 1999; Helmreich & Merritt, 1998; Itoh & Andersen, 1999). We have transformed terms and statements from the original "Operating Team Resource Management Questionnaire" to fit the working situation of doctors, nurses and pharmacists working not only in the operating room but also in other types of departments and wards, keeping the same meaning and intention for each question. Finally, the questionnaire has been translated into Japanese.

The present paper focuses on results from only the first two parts of the questionnaire. Part I contains 57 questions about perceptions of hospital management as well as general questions (e.g., training) that may have a correlation with safety performance. Respondents are asked to rate each question on a five-point Likert scale between 1 and 5 (from 'strongly disagree' to 'strongly agree'). These question items can be largely classified into several groups in terms of organisational and human aspects that form safety culture. In this study, with reference to the original classification by Helmreich & Merritt (1998), we arranged all the items into nine categories of distinct "safety culture aspects": (1) power distance, (2) communication, (3) teamwork, (4) recognition of own performance under high stress, (5) stress management for team members, (6) morale and motivation, (7) satisfaction with management, (8) recognition of human error, and (9) awareness of own competence.

Each category includes several items. For example, the category, power distance comprises twelve items among which the following examples illustrate the format and style of the questions: "The senior person should take over and make all the decisions in life-threatening emergencies"; "Senior staff deserve extra benefits and privileges"; and "Doctors who encourage suggestions from team members are weak leaders."

In the second part of the questionnaire, respondents are asked about their behaviour and actions in terms of reporting own errors and in terms of interaction with patients that have been victims of such errors. Respondents' reactions are elicited as responses to two fictitious adverse events – one in which the patient the patient suffers a relatively severe outcome and the other a relatively mild outcome. The respondents are asked to read each case and subsequently to rate his or her certainty likelihood of engaging in various actions described in the questionnaire. The likelihood rating is made on a five point Likert-type scale ranging from 'definitely yes' to 'definitely no'. The cases and questions have been adapted from items used in the Danish survey of doctors' and nurses' attitudes also reported at this workshop (Andersen et al., 2002). The two fictitious cases were the following:

Case A: A cancer patient is hospitalised in order to receive chemotherapy. When preparing the infusion liquid you become distracted and you mistakenly mix a dosage that has a concentration ten-times greater than the prescribed level. You discover the error several hours later when you administer the same drug to another patient. By this time the patient has received all of the high concentration infusion liquid. You know that the patient now has a risk of developing heart problems later.

Case B: [doctor's version; a slightly modified version was made for nurses adapting to differences in their professional tasks] A patient is hospitalised for planned elective surgery. Before his operation the patient will as a matter of routine for an elder or middle-aged patient receive an anticoagulant injection as a prophylactic against thrombosis. When dictating to the case notes, you are interrupted several times due to patients suddenly getting ill, and you forget to include the anticoagulant for the patient. He develops a thrombosis in a vein in his left leg. He therefore has to remain hospitalised an additional week. It is very unlikely that he will have permanent impairment from the thrombosis.

For each case, respondents received five questions about their attitudes to reporting. They were asked to state the likelihood of the following actions:

- Keep it to myself that I had a mistake,
- Talk in confidence with a close colleague to get support,
- Enter this event into patient's case record,
- Inform my leader about the incident, and
- Report the event to the local reporting system.

There were six additional questions about their possible actions with respect to patients:

- Inform the patient about the adverse event,
- Explain to the patient about the future risk,
- Explain to the patient that the event was caused by your mistake,
- Encourage the patient to apply for compensation from hospital's insurance,
- Explain event to the patient's family, and
- Express regrets about the event to the patient.

The questionnaire was distributed to doctors, nurses and pharmacists working in five hospitals located in different areas in Japan. A total of 66, and 486 and 43 responses were obtained from doctors, nurses and pharmacists, respectively. The mean response rate was 90.7% across the three groups. Among doctors, 33 respondents were heads of department, 22 consultants or doctors after residents, and 9 residents. In the nurse group, responses were collected from 32 matrons and 97 deputy leaders while 354 were from ordinary nurses. In the pharmacist group, samples came from two leaders, 11 deputy leaders and 30 from staff.

Professional Culture of Medical Staff

Category-based responses: Percentage agreement and disagreement as well as mean scores for each of the safety culture aspects mentioned in the last section are shown in Table 1 across the three professional groups. The percentage [dis]agreement is defined as the following rate: the nominator represents 5 and 4 responses, i.e., "strongly agree" and "slightly agree" [the 1 and 2 responses, i.e., "strongly disagree" and "slightly disagree"]; and the denominator represents the total number of responses for the specific items of each aspect. Before calculation of these indices, items that represent negative meaning in terms of the aspect have their ratings of agreement reversed, i.e., 5 and 4 responses, reversed to 1 and 2, and vice versa. This table includes significance levels (chi-square value) of differences between the professional groups.

We have performed similar surveys in the maritime domain using an earlier, derivative version of the questionnaire of the present study, the SMAQ (Andersen et al., 1999; Itoh & Andersen, 1998). Integrating the data collected from seafarers using the SMAQ, we have compared responses concerning safety culture aspects between medical staff and ship crew. In the SMAQ survey, we collected samples from Scandinavian and Japanese ship companies (7 in total, comprising 2,600 responses). The SMAQ samples included 444 Japanese officer responses, 667 Danish and 387 Asian (non-Japanese) officer responses. Table 2 shows comparison results between medical staff and ship officers as well as percentage [dis]agreements of these professional groups. In these comparisons, only the same items between the SMAQ and the present questionnaire were used for each safety culture aspect, and therefore its set of representative items is somewhat different from the one behind Table 1. There were no shared items for the aspects of teamwork, satisfaction with management, and awareness of own competence.

The main general results show that hospital staff as well as ship officers have a relatively high morale and motivation, they exhibit good awareness of communication among teams, members and their organisation. Their satisfaction with teamwork is also relatively high; and in particular, nurses' perception of the value of teamwork is the highest, two thirds of nurses having a positive attitude to this aspect. Compared to these three aspects, percentage agreement in terms of satisfaction with management is not high, and doctors' satisfaction is significantly the lowest of the three professional groups in health care.

One of the safety culture aspects is *power distance*: this refer to the psychological distance between leaders or superiors and subordinate members: A smaller distance reflects, for example, that leaders and their subordinates have open communication initiated not only from leaders but also, more critically, from juniors. The results shown in Table 1 indicate that a relatively small power distance seems to exist in Japanese hospitals; in addition, there is no significant difference in perception of this aspect between doctors, nurses and pharmacists. This result does not match intuitive expectations however. Thus, in Japan, the medical field is widely regarded as having one of the most authoritative and bureaucratic professional cultures in the country. At the same time it is well known from several studies (e.g., Spector et al., 2001) that the Japanese are around the "upper middle" when compared with other nations in terms of power distance – so, while not at the extreme high end (with, e.g., Arab countries and Malaysia) the Japanese are not at the extreme low end either (with, e.g., Denmark and Ireland). Why do we then obtain this result where the measured power distance is relatively low? We offer the following tentative explanation of the data: In Japan there are two contrary concepts covering the expression of values and attitudes: one is the tacit (non-verbalised) disposition for behaviour – which one might call the "real" or unedited meaning – and the other is the "official principle" which is a somewhat idealised stereotype. We are inclined to believe that the results on the power distance measure may include a portion of such an "official" representation.

Table 1 – Percentage (dis)agreement and mean scores for safety culture aspects

Safety culture aspects		Doctors	Nurses	Pharma.	Total	χ_0^2
I. Power distance	% agree.:	30.4%	21.8%	27.6%	23.2%	
	%disagree.:	59.7%	60.4%	59.2%	60.3%	0.88
	Mean score:	2.54	2.43	2.44	2.45	
II. Communication		88.1%	85.9%	89.4%	86.4%	
		4.9%	3.8%	2.9%	3.9%	14.75**
		4.37	4.27	4.41	4.29	
III. Teamwork		57.6%	65.0%	55.2%	63.5%	
		26.0%	15.7%	24.8%	17.5%	16.17**
		3.44	3.68	3.43	3.64	
IV. Own performance under high stress		49.2%	41.0%	42.6%	42.0%	
		38.1%	35.7%	32.9%	35.8%	3.92
		3.14	3.07	3.15	3.08	
V. Stress management for team member		69.5%	69.4%	66.8%	69.2%	
		19.8%	15.8%	21.6%	16.6%	5.12
		3.73	3.75	3.55	3.73	
VI. Morale & motivation		72.9%	65.7%	65.9%	66.5%	
		16.0%	15.1%	18.5%	15.4%	14.75**
		3.91	3.73	3.73	3.75	
VII. Satisfaction with management		45.5%	51.3%	51.7%	50.7%	
		39.6%	28.8%	31.7%	30.1%	10.40**
		3.07	3.30	3.31	3.28	
VIII. Recognition of human error		60.6%	60.7%	55.4%	60.3%	
		26.3%	21.3%	28.6%	22.4%	2.32
		3.56	3.64	3.49	3.62	
IX. Awareness of own competence		58.2%	44.8%	40.2%	46.0%	
		27.1%	24.8%	30.9%	25.5%	17.52**
		3.46	3.28	3.12	3.29	

**: $p<0.01$, *: $p<0.05$

Table 2 – Comparisons with ship officers in percentage (dis)agreement for each safety culture aspect

		Medical staff			Ship officers		
Safety culture aspects		Doctors	Nurses	Pharma.	Japan.	Asian	Danish
I. Power distance	% agree.:	5.7%	8.6%	4.8%	8.3%	17.3%	8.2%
	%disagree.:	89.3%	79.2%	87.5%	81.4%	63.1%	81.7%
	χ_0^2:	11.72**	10.58**				
II. Communication		86.4%	85.4%	85.7%	98.8%	80.1%	95.5%
		6.1%	3.8%	7.1%	0.7%	6.1%	1.2%
		21.83**	163.69**				
IV. Own performance under high stress		48.6%	42.5%	44.1%	38.3%	23.9%	42.7%
		38.4%	34.7%	31.5%	43.1%	53.9%	37.7%
		16.92**	53.73**				
V. Stress management for team member		71.6%	67.4%	68.4%	91.5%	67.7%	82.3%
		18.7%	17.7%	19.9%	3.0%	5.9%	3.7%
		48.16**	294.18**				
VI. Morale & motivation		80.5%	73.9%	71.7%	82.3%	80.3%	65.7%
		11.3%	9.5%	12.6%	7.8%	4.0%	11.8%
		0.002	61.04**				
VIII. Recognition of human error		36.2%	53.3%	39.3%	50.8%	62.9%	51.4%
		47.7%	33.6%	45.2%	36.2%	17.7%	28.5%
		17.11**	0.63				

No Japanese officers are included in the Asian officer group.
Bottom row : Chi square:between doctors/nurses and Japanese ship officers
**: $p<0.01$, *: $p<0.05$

A large part of medial staff has realistic attitudes to and a *realistic* recognition of human error. That is, they recognise that "human error is inevitable," and they do not agree with the question "errors are a sign of incompetence". However, as will be discussed below when we compare the hospital staff data with ship officers' responses, there is a difference in agreement between items comprising this aspect. In contrast to the above two items, there was a large difference in responses to the item "I am encouraged by my leaders and co-workers to report any incidents that I may observe" between the three professional groups. More than 85% of nurses agreed with this question while the percentage agreement of doctors was less than 45%. Regarding attitudes to stress, most medical staff recognise the need for monitoring colleagues' levels of stress and workload. For example, more than 90% of respondents agreed that team members should be

monitored for signs of stress and fatigue during task. In contrast, respondents do exhibit any great awareness of the effects of stress on their own performance. More than half of doctors, and one third of the nurses disagreed with the item "I am more likely to make errors or mistakes in tense or hostile situations". Similarly, only 5% of doctors agreed that their performance is reduced in a stressed or fatigued situation (89 percent disagreement). Percentage disagreement – and a bit lower at 78% for nurse.
Additional results, including comparisons between Danish and Japanese doctors and nurses as well as position and department-based analyses of questionnaire responses, will be reported in subsequent papers.

Doctors' and Nurses' Attitudes to Error Reporting
Doctors' and nurse' responses about error reporting for the two fictitious cases quoted above are depicted in Figure 1. It can be seen from this figure that both doctors and nurses have very positive attitudes to reporting an event to a leader or the doctor in charge of the patient. Similarly, only a few respondents agreed that they would keep the event secret. In particular, nurses' attitudes to error reporting are extremely positive. Their percentage agreements about both submitting the event to the local reporting system and reporting it to their leader was more than 95% for both cases.

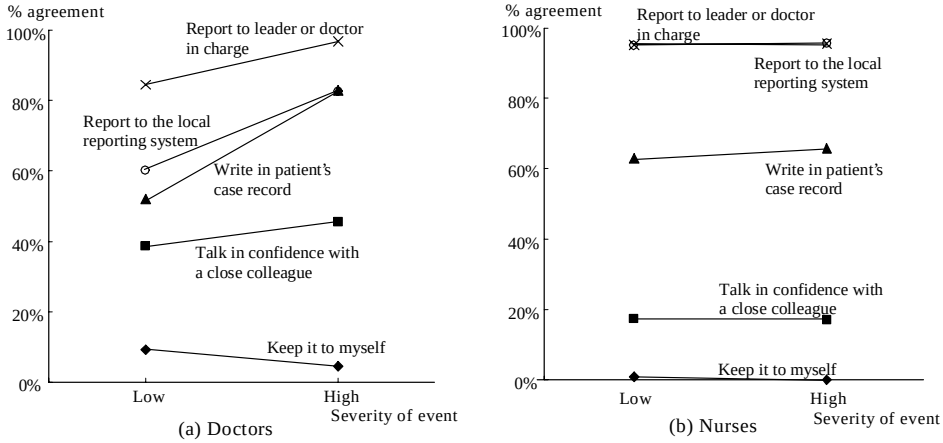


Figure 1 – Willingness to report incident for low and high severity cases

As mentioned above, the attitudes of the nurses to error reporting are significantly more positive than those of the doctors for nearly all items. For the severe outcome case (Case A), there is no significant difference between the groups in their willingness to enter the event into the patient case record (χ^2 =2.64). Only in terms of reporting to one's leader or doctor in charge for the severe case is the doctors' attitude significantly more positive than that of the nurses (χ^2 =10.18) although the absolute difference is small. The reason why doctors are more willing to report the case to their leader or the doctor in charge may have to do with the fact that doctors are responsible for treatment and nurses for care. For the rest of items in the severe case, the nurses' responses were much more positive than those of the doctors. For the milder outcome case (Case B), responses to all the items on incident/error reporting are significantly different between the two professional groups: Nurses had much more positive attitudes to the error reporting than doctors.

Both doctors and nurses agreed that they would take significantly more positive actions for the severe case. Only in terms of the item, "talking in confidence with a close colleague to get support", is there no significant difference between the two cases for both doctors (χ^2 =0.23) and nurses (χ^2 =1.77).

Responses to actions with respect to the patient show a similar trend across the two cases, as shown in Figure 2. For almost all the proposed actions about interaction with the patient, both doctors and nurses have provided more positive responses for the severe case than for the mild outcome case. The more severe the outcome of an error, the more likely it is that the consequence will be explained to the patient, that the patient will be told that the event was caused by the doctor's or nurse's own mistake, that the event will be explained to the patient's family, and that the doctor or nurse will express regrets to the patient about the event. However, a reverse effect was found in nurses' response to informing the patient about the adverse event (χ^2 =10.20), where their willingness to inform was highest in the milder outcome case. No significant

difference was identified between the levels of severity for the doctors' informing the patient (χ^2 =0.48) although a reverse trend was observed but below the level of significance.

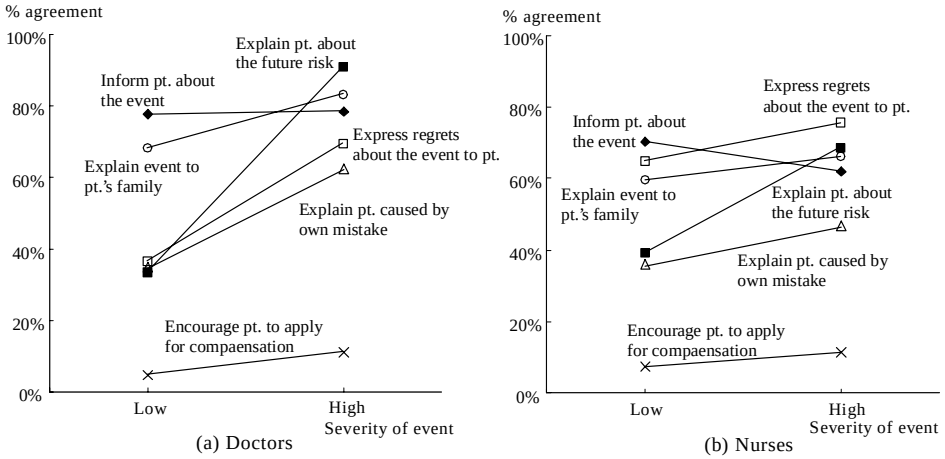


Figure 2 – Actions with respect to patient for low and high severity cases

Nurse responses showed that they are more willing to apologise to the patient about the event than doctors. For the other actions in relation to the patient, doctors responded more positively than nurses, no doubt since doctors have the primary responsibility for carrying out these act vis-à-vis the patient when such events occur.

Table 3 – Ward-based nurse groups' reporting rates of incidents

	Types of incidents					Report rate to system (/person/yr)	Inform. rate to pt.(%)	Rate of annoyance (/person/yr)
Wards	Injection	Oral intake	Fall	Misuse of equipment	Others			
Internal medicine	40.4%	22.9%	12.8%	25.7%	-	1.79	36.7%	0.13
Surgery	35.5%	14.5%	27.6%	22.4%	-	1.27	47.4%	0.00
Outpatient	23.6%	4.2%	4.2%	61.1%	6.9%	0.97	44.4%	0.05
Operating room	10.9%	-	-	48.9%	40.2%	4.60	6.5%	0.20
Mixed ward	31.7%	18.7%	24.5%	25.2%	-	1.62	48.2%	0.08
Total	28.7%	13.3%	14.8%	34.6%	8.6%	1.62	37.9%	0.08

Investigation of Incident Reporting

Incident Statistics: A statistical summary of incident reports submitted from nurses only was obtained from one of the hospitals surveyed in this study. The summary includes the number of incidents reported during the previous year (April 2000 – March 2001) based on incident types as well as the number of cases that have involved some types of "annoyance" or "trouble" to the patient. The "annoyance" or "trouble" cases include not only injuries caused by errors but also minor events undesirable to the patients such as a pain, feeling worse, temporary variation in vital signs or lost belongings, e.g., artificial teeth. In only a single case of the reported incidents did a patient suffer an injury, namely a fracture of the clavicle after falling down from a bed with free railing – a nurse had forgotten to put up the railing. All the other "annoyance" cases were small incidents with no injuries. Table 3 shows the summary of ward-based reporting rates that includes three indices on incident reporting as well as the percentage of each type of incident: (1) reporting rate to the system, i.e., the rate of incidents of the given type submitted to the hospital's reporting system per nurse in a year, (2) the rate of informing the patient, i.e., calculated by dividing the number of acts of informing the patient about the event by the total number of reported cases, and (3) the rate of annoyance cases, i.e., the rate of reported "annoyance" cases per nurse in a given year.

As can be seen in Table 3, the reporting rate to the system varies across the wards. The reporting rate from the operating room was the highest of all the wards in this hospital. Nurses working in the outpatient and

surgery wards submitted incident reports less frequently than those in the other wards. This may suggest that the likelihood of incident occurrence basically depends on the type of medical treatments and nurses' activities. The rank of the ward in this index is correlated with the rate of annoyance (r=0.867). However, the range of the latter index is smaller than the former. The informing rate to the patient seems to have reversed trend compared with the two indices in terms of differences between the wards in which nurses work. The rate of the operating room nurse was much smaller than the others.

Correlation between questionnaire responses and reporting rates: We examined correlations between the nurses' questionnaire response to error reporting and the actual rates of reporting to the system and of informing to the patient. For this purpose, questionnaire responses were rearranged only for nurses in the hospital from which we obtained the summary of incident reports. The percentage agreements and mean scores were calculated based both on the ward and on the position for each safety cultural aspect derived from the safety culture related questions and for each question item under the fictitious adverse event cases. Figure 3 shows the graph plotting all the ward- and the position-based nurse groups in terms of the reporting rate to the system and mean score of questionnaire responses to the item "reporting to the event to the local reporting system". From this figure, no correlation between the questionnaire response and the actual reporting rate appears. Since the questionnaire responses to this item can be a measure of a respondent's willingness to report an adverse event, the identical rate of actual reporting of two groups having different mean scores of questionnaire responses indicates that a higher score group has lower incident risk than the other group. Following this interpretation, the above-mentioned result may suggest that a group having positive attitudes to error reporting has a lower incident risk than – or at worst one which is identical with – a group having more negative attitudes.

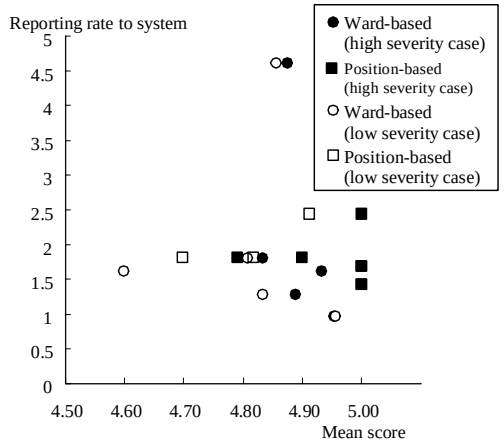


Figure 3 – Relationship in incident reporting between questionnaire responses and its actual rate

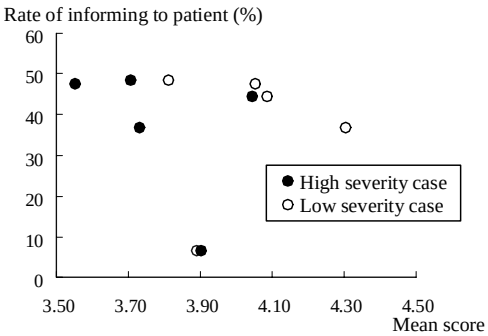


Figure 4 – Relationship in informing to patient between questionnaire responses and its actual rate

Regarding the nurses' attitudes to informing the patient about an adverse event, each ward group – no position-based mapping in this figure – is mapped on the geometric plane of mean scores of questionnaire responses and the actual rate, as shown in Figure 4. As in Figure 3, each ward has two marks having an identical value to the vertical axis for two fictitious cases in this figure. In both cases, there seems to be no correlation between the questionnaire response and the actual rate. However, when we exclude single-ward marks that have an exceptional plot, e.g., the marks having the lowest actual rate (i.e., from the operating room), it can be seen that the nurse's willingness to inform the patient is negatively correlated with its rate of occurrence for both cases. It may be hypothesised that the mismatch between nurses' willingness and the actual rate obtains because the likelihood of the event being conveyed to the patient is determined primarily by other factors rather than the nurse's willingness, for example, by organisational factors such as the policy of the hospital or the department or the organisational climate.

To discuss such effects of safety climate or organisational aspects on the actual reporting behaviour, a correlation of the reporting rate to the system with one of the safety culture aspects, *recognition of human error* is depicted in Figure 5 based on the ward and the position. As can be seen in this figure, the actual rate of incident reporting is negatively correlated with the recognition level of human error (r=-0.944;

p<0.01). This indicates that the more realistic nurses' recognition towards human errors becomes, the less frequently an incident report is brought up. As remarked previously when discussing the relationship between the nurse's willingness to report errors and the actual rate of incident reporting, higher rate of reporting cannot be taken in itself to indicate a higher level of safety. Rather, one may speculate that this index serves to measure accident risk. According to this view, it may be suggested that realistic recognition of human error contributes to a lower risk of adverse events in a hospital.

Figure 6 shows a ward-based correlation between the informing rate to the patient and recognition of human error. As can be seen in this figure, this safety cultural aspect is also correlated with the nurse's willingness to explain an event to the patient (r=0.983, p<0.01). It is natural to interpret this result as indicating that a realistic recognition of human errors facilitates the willingness to inform the patient about the event.

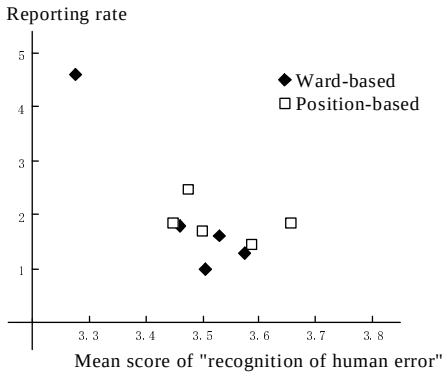


Figure 5 – Correlation between rate of incident reporting and recognition of human error

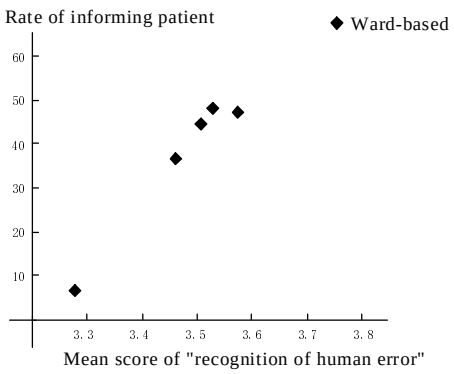


Figure 6 – Correlation between rate of informing the patient and recognition of human error

Another safety cultural aspect, *power distance*, was examined to ascertain whether it influences the rate of reporting and the rate of informing the patient. Thus, relationships of this aspect to the former reporting index are depicted in Figure 7, and to the latter in Figure 8. In terms of the reporting rate to the system, Figure 7 indicates that the power distance seems to have a slightly negative correlation with the reporting index (r=-0.568, p<0.10). Assuming the above-mentioned interpretation of this index about accident risk, this effect of power distance is somewhat unexpected, being in a direction opposite to what common sense would lead us to think, i.e., the larger the power distance, the lower is the accident risk. However, there is a single exceptional data point (an outlier) that deviates from the other four ward groups as in Figures 3 and 5. This plot also comes from the nurse group of the operating room. Excluding this group from the geometric plane, the graph plot indicates that there may be a positive correlation between the reporting rate and the power distance. Thus, it may be suggested that a small power distance (i.e., open communication between team members and leaders and a small psychological distance between leaders and subordinates) contributes to good organisational culture and in turn to patient safety. At the same time, these results may suggest that the actual reporting rate is affected not only by the professional activities, e.g., operating room vs. others, but also by the power distance in the workplace although it is impossible to derive a sound conclusion only from these results.

On the other hand, the actual rate of informing the patient seems to be positively correlated with the level of power distance (r=0.647, p<0.10), as shown in Figure 8. This may suggest that the larger the power distance is in a ward according to nurses' perception, the greater is the likelihood that a given event will be reported to the patient. This result may support the above-mentioned hypothesis on the effect of an organisational factor. However, this effect of power distance on interaction with the patient may also be in an opposite direction to our common sense. It is true that there may be several organisational factors confounded with the type of ward in this type of data, and therefore it is reasonable to consider that such unknown organisational factors contribute to the informing rate to the patient.

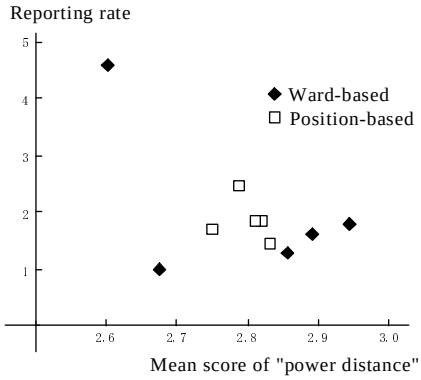


Figure 7 – Correlation between rate of incident reporting and power distance

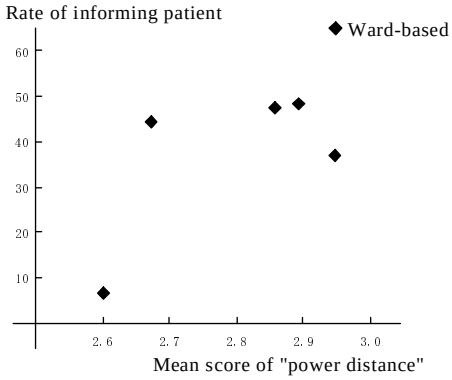


Figure 8 – Correlation between rate of informing the patient and power distance

Conclusion

This paper reported the integrated results from a questionnaire-based survey of safety culture related attitudes among hospital staff and analysis of the rates of incident reporting. The aim of this investigation was to identify safety cultural perceptions and attitudes among medical staff and to elicit their willingness to reporting errors and interacting with the patient who fall victims to adverse events. To elicit characteristics of safety culture, we compared the questionnaire results with the data obtained in our former studies of the maritime domain (Andersen et al., 1999; Itoh & Andersen, 1999). Finally, the actual rate of incidents reported in one of the hospitals surveyed was compared with respondents' attitudes and views as identified in the survey with a view to assessing their connection to patient safety.

A major outcome of the present survey has been to obtain a hypothesis concerning correlations between the actual reporting statistics and some of the safety cultural aspects, e.g., recognition of human fallibility, and power distance, for further investigations. In particular, the survey results seem to indicate that a nurse group who has a relatively larger power distance and unrealistic recognition of human errors will be liable to produce a greater number of incidents. Therefore, in addition to the importance of a realistic recognition of human error potentials, we hypothesise that a relatively small psychological distance between superiors and subordinate members and open communications within an organisation and among team members may be one of the key factors for establishing and maintaining a safe medical organisation.
Finally, we suggest that efforts be devoted to examining statistical correlations of the actual incident rates with the perceptions and views of medical staff about safety cultural aspects. The questionnaire-based method may be a useful supplement to accident/incident data in order to identify high and low risk work units in the medical domain. This is of importance whenever incident reporting is incomplete or when reporting criteria are heterogeneous. Equally, while incident reporting is a retrospective index of safety levels, the survey data may be used prospectively.

Acknowledgements

We are heavily indebted to Robert L. Helmreich, the University of Texas at Austin who provided not only the original "Flight Management Attitudes Questionnaire" and "Operating Team Resource Management Questionnaire" but who also inspired the adaptation of this survey instrument for patient safety. We would like to acknowledge Takako Aoki, Tomioka Municipal Hospital, as well as Naomi Kitazawa, Tokyo Medical and Dental University for their cooperation in this project. We are also grateful to the Danish patient safety survey project group for permission to use their cases and question items.

References

ACSNI (1993). Advisory committee on the safety of nuclear installations: Human Factors Study Group Third Report: Organising for safety. HSE Books, Sheffield.

Andersen, H.B. (2002). Assessing safety culture. Technical Report R-1459, Risø National Laboratory, Roskilde, Denmark.

Andersen, H.B., Garay, G. and Itoh, K. (1999). Survey data on mariners: Attitudes to safety Issues. Technical Report I-1388, Systems Analysis Department, Risø National Laboratory, DK-4000 Roskilde, Denmark.

Andersen, H.B., Madsen, M.D., Hermann, N. *et al*. (2002). Reporting adverse events in hospitals: a survey of the views of doctors and nurses on reporting practices and models of reporting. *Proceedings of the Workshop on the Investigation and Reporting of Incidents and Accidents.* Univ. of Glasgow.

Kohn, L.T., Corrigan, J.M. and M.S. Donaldson, eds. (1999). *To err is human: Building a safer health system*. National Academy Press, Washington DC.

Hee, D.D., Pickrell, B.D., Bea, R.G. *et al.* (1999). Safety management assessment system (SMAS): A process for identification and evaluating human and organization factors in marine system operations with field test results. *Reliability Engineering and System Safety*, 65: 125-140.

Helmreich, R.L. and Merritt, A.C. (1998). *Culture at work in aviation and medicine: National, organizational and professional influences*. Ashgate, Aldershot, UK.

INSAG (1991). International Nuclear Safety Advisory Group, Safety Culture, Safety Series No. 75-INSAG-4. International Atomic Energy Agency, Vienna.

Itoh, K. and Andersen, H.B. (1999). Motivation and morale of night train drivers correlated with accident rates. *Proceedings of the International Conference on Computer-Aided Ergonomics and Safety*. Barcelona, Spain, May (CD ROM).

Itoh, K., Andersen, H.B, Seki, M. and Hoshino, H. (2001). Safety culture of track maintenance organisations and its correlation with accident/incident statistics. *Proceedings of the 20th European Annual Conference on Human Decision Making and Manual Control*. 139-148, Copenhagen, Denmark, June.

Itoh, K. and Andersen, H.B., Tanaka, H. and Seki, M. (2000). Attitudinal factors of night train operators and their correlation with accident/incident statistics. *Proceedings of the 19th European Annual Conference on Human Decision Making and Manual Control*. 87-96, Ispra, Italy, June 2000.

Reason, J. (1993). Managing the management risk: New approaches to organisational safety. In B. Wilpert and T. Qvale (Eds.), *Reliability and safety in hazardous work systems*. Lawrence Erlbaum Associates, Hove.

Reason, J. (1997). *Managing the risk of organizational accidents*. Ashgate, Aldershot, UK.

Spector, P.E., Cooper, C.L. and Sparks, K. (2001). An international study of the psychometric properties of the Hofstede Values Survey Module 1994: A comparison of individual and country/ province level results. *Applied Psychology – An International Review*, 50(2): 269-281.

Learning by Reporting System of Organizational Accidents in Japan

Kenji Tanaka

Graduate School of Information Systems University of Electro-Communications
Chofu, Tokyo 182-8585, JAPAN E-mail: tanaka@is.uec.ac.jp

Abstract : Organizational accidents have occurred repeatedly in Japan over the past several years. Most of these accidents seem to have been occurred within a gray zone between the safety zone and the danger zone rather than within the danger zone. The present paper discusses methods for preventing such accidents that occur within the gray zone. Firstly, it asserts that the gray zone is unavoidable so that measures for the gray zone are indispensable in a design of management system. Especially, safety design is classified into a safety-assurance design and a danger-avoidance design, and the danger-avoidance design method should incorporate an incident reporting system or/and an accident reporting system for preventing accidents within the gray zone. Next, we show three common causes of accidents related to the gray zone and focus on the dynamics of judgement. Lastly, we refer to results of investigation about incident reporting system introduced at a hospital in Japan.

Keywords: incident reporting, accident reporting, gray zone, danger-avoidance, organizational learning

Introduction

In Japan, organizational accidents(Reason,1997) have occurred repeatedly over the past several years. However, we don't learn from the accidents and similar accidents occur again and again. Why do we neglect to learn for preventing accidents? Our research suggests that most accidents occur within a gray zone between the safety zone and the danger zone rather than within the danger zone itself. The gray zone is an uncertainty zone where difficulty is encountered in judging whether the system is safe or dangerous. The present paper insists that incidents reporting system(IRS) and accidents reporting system(ARS) will be effective for preventing such accidents that occur within the gray zone. At many hospitals in Japan, medical incident reporting systems were introduced some years ago, but the systems are not utilized effectively. We show some results of our analyses at a hospital about reasons why IRS has been not utilized effectively. Soft System Methodology (Checkland & Scholes, 1990) was adopted for the analysis.

Accidents Occurred in Japan

Firstly, we introduces major accidents that occurred during recent years in Japan, and shows that they seem to have taken place within a zone of uncertainty that exists between the safety and danger zones rather than within the danger zone itself.

Ex.1 Criticality accident at Tokai-mura: A criticality accident occurred on September 30, 1999 at a uranium processing plant in Tokai-mura, Japan. Two persons died as a result of exposure to radiation. The accident differs from the famous accidents of Three-Mile Island and Chernobyl in that it took place in a uranium pre-processing plant rather than in a power reactor.

Ex.2 Subway derailment accident: A subway derailment accident occurred in Tokyo on March 8, 2000. When the train exited a subway tunnel to tracks above ground, the car left the rail and collided with the cars of a train moving in the opposite direction. Five persons died in the accident and 63 were injured. The derailment took place at a point where the track has a radius of curvature of 160.1 meters. The Japanese Ministry of Transportation has prohibited railway companies from installing track having a radius of curvature less than 160 meters; the accident occurred just at the edge of this limit(Figure 1).

Ex.3 Environmental pollution with dioxin: In March 2000, an official inspection detected that a factory which produces industrial machinery was draining dioxin-containing water into a river. The quantity of contaminants far exceeded the standard. A more detailed study traced the cause to improper connection of piping. Water contaminated in a treatment process for industrial waste drained away through a pipe for rainwater, without removal of contaminants. When the facility was built, a pipe leading to a treatment

facility for excluding contaminants was mistakenly connected with a pipe for rainwater. The facility was inspected during the course of construction, but piping connection was outside the scope of inspection.

Ex.4 Food poisoning by milk: In July 2000, in Osaka, a mass outbreak of food poisoning was caused by residents drinking contaminated milk. In the course of processing, a small amount of staphylococci was intermixed with the milk. A recall of the products was delayed, and resulted in more than 5000 people becoming ill. The staphylococci were detected at a valve in piping for transferring excessive milk to a reserve tank. Staphylococci propagated not in a tank for long-term storage, but at a valve where the milk passed transiently.

Thus, most of resent accidents have taken place within a zone of uncertainty, which exists between a safety zone and a danger zone (Figure 2). So far, various protective measures have been designed and defenses-in-depth have been introduced for processes and sites that have been defined as potentially dangerous. Despite of these efforts, accidents have occurred at the edge where a system may be dangerous under some specific conditions or under improper use□

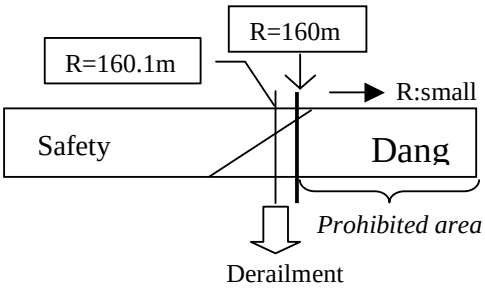


Figure 1 - Subway Derailment Accident

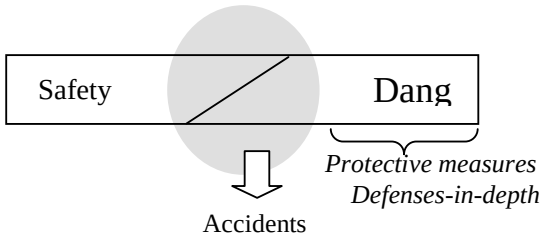


Figure 2 - Gray Zone and Accidents

Role of Incident/Accident Reporting System

In actual large-scale systems, gray zone is unavoidable in design stage, since the following three factors produce a gray zone.

1. Limitation of designer's prediction. There exists a limitation of designer's prediction at the design stage, in relation to changes of situations and conditions of use. Systems are often utilized beyond expectation of their designers. The limitation is an unpredictable factor and therefore is unavoidable in principle.

2. User's misuse. The second factor that generates a gray zone is user's misuse. Among misuses that the designer can predict at the design stage, some can be prevented by several design devices such as fool-proof structure, inspection process, while other can not be prevented by such protective measures. Users' misuses are predictable but uncontrollable.

3. Insufficient information. Gray zone also emerges when sufficient information can not be collected in design stage. However, since efforts to gather information can decrease the influence from this factor, the factor is considered controllable.

System operators or management systems should focus on the fact that IRS/ARS are indispensable for compensating for the first factor of a designer's limitation about prediction. Moreover, also for the second uncontrollable factors at the design stage, IRS/ARS are effective since the reported sheets often include useful information for preventing fatal accidents. Especially, IRS may include the potential information so that it should be utilized to prevent accidents. Thus, both IRS and ARS serve to improve the objective systems toward safe systems in operating stage rather than in design stage. In another words, IRS/ARS give a chance to learn and it seems a unique chance. This also suggests that a designer must construct systems in consideration with a gray zone and that he should design systems to easily improve at operating stage.

Safety-Assurance Design vs. Danger-Avoidance Design for Product

As a gray zone is unavoidable, Tanaka(1996) has already proposed two kinds of design of products: **safety-assurance design** and **danger-avoidance design** (cf. Figure 3). These two design methods provide different interpretations for the gray zone between safety and danger. A safety-assurance design considers the gray zone an unsafe zone that must be avoided, whereas a danger-avoidance design considers the gray zone a non-danger zone for which countermeasures are not to be taken. These two methods of design will be adequate to system design as well as products design as we explain later. A designer needs to understand that a non-dangerous system isn't always a safe system and includes some potential danger in gray zone. Fool proof or fail-safe designs that traditional studies insist as structural or functional safety designs, are examples of safety-assurance design.

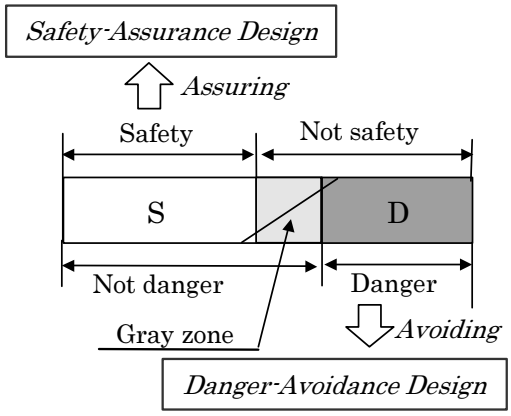


Figure 3 - Safety-Assurance Design & Danger-Avoidance Design

Though such a safety-assurance design is desirable for systems that are required to be safe, practical systems do not always adopt a safety-assurance design since the design generally conflicts with availability or operability. In some systems, activities for conforming to safety are difficult and even become a barrier to adaptive behavior. In contrast, a danger-avoidance design often leads to ease of use and is tolerant of adaptive behavior. Since products or systems subjected to danger-avoidance design may be used in the gray zone under risk, designer should reduce the range of gray zone to lower the risk, or enable all users to perceive the possibility of dangerous situations. We strongly insist that a danger-avoidance design needs to incorporate IRS to prevent accidents at gray zone.

Safety-Assurance Management and Danger-Avoidance Management

Failures concerned with management process are referred as "system failures," and organizational accidents are one type of system failures. For preventing the system failures, management activities should be classified into two types of management; **safety-assurance management** and **danger-avoidance management,** as the same classifications used for product safety designs. For example, recognizing the existence of gray zone is important in determining rules or standards for management.

Example: As an example of two kinds of design in management, we focus on a description of rules. Most management activities are performed according to company rules, industry-wide regulations, or national laws. In consideration of the gray zone, rules can be given in two ways; by positive description and by negative description.

In **positive description,** the rule is expressed by a positive statement such as "Do the following when you want to..." This type of description is often used for safety indications. On the other hand, in **negative description,** the explanation is expressed by negative a statement such as "Avoid doing the following when..." This type of description is used for indicating warning or caution.

In order to assure safety, rules are required to cover not only every predictable dangerous situations, but also situations arising from uncertainty (Figure 4). However, rules are usually expressed by negative description; therefore, they don't include cases for the uncertainty zone; that is, the gray zone. Covering

the gray zone requires too many rules, and this is why accidents occur in the gray zone. One way to avoid drafting too many rules is to express a safety indication restricted to the specific use. Such a safety indication is usually expressed by positive description and allows only instructed modes of operation, accordingly assuring safety with high probability but permitting no flexibility of use (see Table 1). Thus, as positive description has disadvantages, persons who determine rules must consider which type is better, positive description or negative description. For example, positive description is desired for use by a beginner or for a system requiring a high level of safety. In contrast, negative description is useful for use by an expert or for a system requiring a low level of safety.

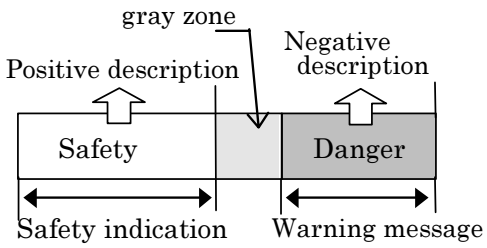


Figure 4 - Positive description and negative description

Table 1 - Comparison of two types of descriptions

	Positive description	Negative description
	-safety indications	-warning messages
advantages	-safety assurance -simple expression	-variety of uses -flexible uses
dis-advantages	-restriction of use -limit to reuse	-lack of warning for unpredictable use, -too many messages

When the person who determines rules selects negative description, he must introduce rules for preventing accidents leading from user's misuse. Moreover, an incident reporting system is expected to be effective for preventing accidents from the viewpoint of limitations of prediction.

Dynamics of Judgement at Gray Zone

By more deeply investigation of the reasons for accidents which occurred in the gray zone, we found that the accidents share three common causes; relaxation of standards, lowered risk perception and over-confidence in Certification. They reveal that judgement at gray zone is dynamic.

(1) Relaxation of standards Within companies that resulted in accidents, procedures or standards had been relaxed before the accidents were occurred. In the example 1 of criticality accident at Tokai-mura, the procedure for uranium processing was changed to use of the bucket and precipitation tank in order to enhance efficiency. Use of the bucket and precipitation tank deviated from the government-authorized dissolution method, but safety management committee recognized such use and authorized it. In the subway company responsible for derailment accident, company's standards for setting safety guard for preventing derailment has been relaxed from 200 meters(1951) to 140 meters(1968). Surprisingly, the current standard of 140 meters is less than the design limit, 160 meters, stipulated by the Ministry of Transportation. A lot of accidents resulted from such an exceeded relaxation of standards.

(2) Lowered risk perception Risk perception of managers or operators seemed to lower after long periods of accident-free operation. Actually, the above relaxations were performed gradually. In the criticality accident at Tokai-mura, the procedure for uranium processing was changed on several previous occasions (Furuta *et. Al.*, 2000).

- In 1986, a method for cross-blending solution was introduced.

- In 1993, the company introduced use of stainless steel bucket for dissolving the uranium.
- In 1994, the bucket was also applied to dissolution carried out in the purification process since the processing time could be shortened.
- In 1995, a buffer column was used to homogenize the uranium solution.
- In 1999, a precipitation tank replaced the buffer column.

These repeated improvement actions resulted from the efforts toward efficiency. However, at the same time, they reveals the lack of risk perception in management. Also, in the subway company responsible for derailment accident, standards for setting safety guard for preventing derailment has been relaxed as follows;

1951 /200m -> 1954 /180m -> 1958 /160m -> 1961 /150m -> 1968 /140m.

The standard continued to be relaxed during accident-free operation. Though long period of accident-free operation doesn't indicate a higher degree of safety, it makes organizations be vulnerable to overconfidence. Paradoxically, absence of accident may lead to an accident. Therefore, minor incidents can be useful in preventing major and sometimes fatal accidents.

On the other hand, we remark that high frequency of accidents leads to loss of trust in user or operator of the systems and that other factors also erode trust. For example, our cognitive experiments show that a specified occurrence pattern of accidents erodes trust even when the frequency of accidents is held constant (Itoh, Abe and Tanaka, 1999). In our laboratory experiments, the influence of successive failure events was compared with the influence of discrete failure events, and we found that the successive failures leads to greater loss of trust than do discrete failures, even at the same frequency of failure. Although the experiments were conducted as computer-controlled simulation of a plant on a personal computer, we believe that most members of an organization would show a similar shift in attitude.

(3) Over-confidence in Certification Thirdly, some of accidents in Japan resulted from over-confidence in certification. Organizations fail to adhere closely to formalities. In the example 4, the company that mistakenly sold poisoned milk was producing milk under the HACCP (Hazard Analysis and Critical Control Point) system. The company, however, neglected to carry out the inspection of the producing process that HACCP instructed the company to do, resulting in its failure to detect the fungi. After the accident, the Ministry of Pubic Welfare in Japan investigated all facilities in Japan that had been certified in their compliance in operating under a HACCP system, and found incomplete records in half the facilities.

Maintaining compliance with a process that has been certified under such as ISO or HACCP, is just as important as obtaining the certification. In Japan, certified organizations aren't checked sometimes whether manufacturing systems are maintained according to instructions. And also, the certification does not always guarantee total safety. IRS/ARS are expected to detect neglected actions and to specify a gray zone.

The three factors discussed above are also found in the background of other accidents in Japan. All companies change the standards by pursuing improvements in work efficiency and performance. As Reason[3] indicated, the problem lied not in the effort to improve standards, but in neglecting to consider the following questions.

1. Would the change in standards increase risk?
2. What new risk would the change generate?

Recognizing the existence of gray zone is important in determining or improving standards for management. IRS/ARS are expected to provide information about their questions. Especially, IRS is desirable to prevent fatal accidents.

Incident Reporting System in a Hospital

In Japan, incident reporting system has been adopted in fields such as aviation and nuclear power. Recently, as instruction by Ministry of Public Welfare, hospitals have started to introduce incident reporting systems for preventing fatal accidents. However, IRS is not utilized effectively in health care field. Our group researched a hospital in Japan and analyzed why IRS has been not utilized in the hospital. Our research group adopted Soft System Methodology(SSM) developed by P. Checkland at Lancaster University[1], and in accordance with the methodology carried out action research with 12 nurses. As a result, we found that three factors impeded the reporting system.

(1) Data of too many events are collected. Hospital policy dictated that all incidents and accidents were to be reported via reporting sheets. As the very considerable number of reporting sheets was collected, gathered sheets were ignored without being analyzed.
Originally, definite danger cases should not be reported, since they must be measured in advance and avoided. However, the risk manager in the hospital forgot that the reporting system should be used to compensate for limitations of prediction, and adopted the misconception that the reporting system is effective for all events. We remark that the incident reporting system should be used only for the gray zone.

(2) Meta-system function is not prepared. Since IRS suggests new failure points that were not predicted in the design stage, some processes or rules are required to change on the basis of IRS. This change belongs to a structure control, so that a meta-functional level of management is required (Figure 5). Main function of meta-layer is to direct to revise rules or standards, or to support the revise activities. For example, risk manager works such an activity. If he refers to change of criterion or value under which rules are determined, the activity belongs to upper function that performs "double-loop learning"(Argris & Schon,1996). Thus, meta-layer makes decision about decision making (Gigch,1991).
Though the hospital we investigated has a clear organizational system including meta-function, almost of hospitals in Japan don't build adequate organizational system for changes of rules. Recently, Japanese hospitals started to adopt risk manager, but the risk manager is not a professional for risk management and is selected among doctors. Hence, risk manager usually performs his/her primary role as a doctor and occasionally analyzes medical incident/accident reports. Risk management committee constitutes of only doctors or nurses except an expert for risk management. In the hospital, as the layer often added rules in response to incidents, not a few nurses complained that the successive addition of rules is overwhelming. Measures should be considered from the viewpoint of the total number of tasks assigned to a nurse. All hospitals should prepare the meta-function by allocating a professional position for risk analysis. A risk manager is required not only for preventing accidents but also for managing the hospital.

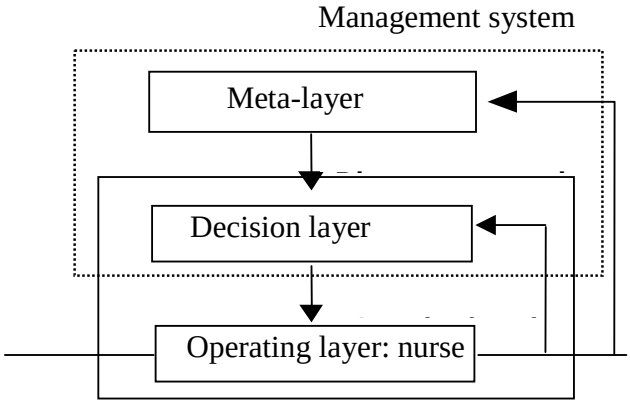


Figure 5 - Meta-system function for structure change

(3) Motivation is not encouraged. Reason(1997) shows three factors that are essential for creating a climate of trust and two factors that are needed for motivating people to file reports.

1. Indemnity against disciplinary proceedings
2. De-identification
3. Separation of department collecting reports from bodies with authority
4. Rapid, useful feedback
5. Ease of making the report

Unfortunately, most of them were not satisfied in the medical IRS in Japan. Risk managers must create an environment in which doctors or nurses have a positive incentive to report the sheet. It seems to be difficult in Japanese culture to adopt indemnity, but it will be easy to separate the department collecting reports from bodies with authority. If the reporting system is established as a duty from the top down, the system will

not be fully performed. The system should be understood as a mechanism for improving activities from the bottom up so as to narrow the gap between rules and real situations.
IRS is the only method to complement the limitation of the designer's prediction. Overcoming the above three factors is a condition for making use of IRS.

Conclusion

The present paper focuses on the gray zone, a factor involved in organizational accidents, which is produced by the limitation of designer's prediction. IRS/ARS are useful to compensate for the limitation. Especially, danger-avoidance management should adopt learning mechanism by IRS. In Japan, now is a first stage to introduce IRS/ARS. The next stage is to establish a management system for utilizing information from the reports and to have safety culture for encouraging.

References

Argyris, C. and D.A. Schon (1996). *Organizational Learning II*, Addison-Wesley, U.S.A.

Checkland, P. and J. Scholes (1990). *Soft Systems Methodology in Action*, John Wiley, Chichester.

Furuta, K., K. Sasou, R. Kubota, H. Ujita, Y. Shuto, and E. Yagi. (2000). Human Factor Analysis of JCO Criticality Accident, *Cognition, Technology & Work*, vol.2, no.4, pp.182-203.

Gigch, J. P. (1991). *System Design Modeling and Metamodeling*, Plenum Press, New York.

Itoh, M., G. Abe, and K. Tanaka (1999). Trust in and Use of Automation: Their Dependence on Occurrence Patterns of Malfunctions," *Proc. of IEEE Int. Conference on Systems, Man, and Cybernetics*, Tokyo, pp.Ⅲ715-Ⅲ720.

Reason, J. (1997). *Managing the Risk of Organizational Accidents*, Ashgate, Aldershot.

Tanaka, K. (1996). Safe Products versus Non-Dangerous Products ―Safety Assurance Design and User-Centered Design―, *Proc. of Int. Conference on Quality '96*, Yokohama, pp.473-476.

Tanaka, K. (2001). Management Learning for Prevention of Organizational Accidents, *Proc. of 4th International Quality Management and Organizational Development Conference*, Sweden, pp.159-166.

A Study of Incident Reporting in Air Traffic Control – Moral Dilemmas and the Prospects of a Reporting Culture Based on Professional Ethics

Marlene Dyrløv Madsen,

Risø National Laboratory, System Analysis Department, P.O. Box 49, DK 4000 Roskilde, Denmark
marlene.dyrloev.madsen@risoe.dk; www.risoe.dk/sys/spm

Abstract: This paper outlines some potentials of developing a profession-based ethics as part of a sound reporting culture. It is proposed that a profession-based ethics guided by the profession itself will foster responsibility towards safety and prevent unsafe behaviour. The idea of a profession-based ethics proposed here derives from results of an interview study of Air Traffic Control (ATC) summarised briefly. The study investigates the moral aspects of the handling of human error in reporting of incidents by comparing differences in reporting practices within two ATC centres (Denmark and Sweden) and analyses the effects of safety culture on air traffic controllers' (ATCOs') willingness to report on incidents. Ethical and organisational dilemmas were identified that arise when seeking to introduce incident reporting into an organisation – in particular dilemmas that turn on the balance between encouraging a high rate of reporting versus seeking to meet the perceived demands for holding operators accountable for errors. The paper briefly outlines and discusses the moral aspects that constitute the ethical dilemma between safety and justice and concludes that safety is the primary factor to consider, but suggest that developing a profession based ethics makes it possible to create a balance between justice and safety.

Keywords: Reporting culture, safety culture, responsibility, regulation, air traffic control, ethics

Introduction

In safety critical domains such as aviation, air traffic control, maritime operations and health care incident reporting plays a primary role in optimising safety. Organisations and companies operating in these domains need to gather and disseminate knowledge about errors, incidents and near-misses in order to prevent reoccurrence and thus prevent potential accidents. To gather this type of information an effective reporting culture alongside a healthy safety culture is essential to the organisation. But establishing an effective reporting culture has many barriers. One of the most significant barriers arises when operators are asked to report on their own errors while, at the same time, they run the risk of being sanctioned by disciplinary action or even penalties imposed by courts if they are found guilty.

Operators have the impression that they may be found guilty merely by their causal relation to an "incident" leaving them with a feeling of injustice since questions of moral responsibility and accountability are not taken into consideration. For instance, it may well be questioned whether it is fair to hold an operator at the "sharp end" responsible for causing an error while trying to meet pressures of production created by decisions made at the "blunt end". The willingness of the personnel at the sharp end to report will therefore often depend on the consequences by which their reporting is met. This leaves the organisation with a dilemma, since the organisation must choose a proper balance between safety (learning from reports) and justice (holding its personnel accountable when they act negligently).
Various proactive approaches - organisational, psychological and anthropological - have been applied to the problems of self-reporting. One popular approach to promoting self-reporting is to argue for the establishment of a non-punitive system in order to foster a "blame-free" culture. Another approach, which stems from the same line of thought, is to promote what is coined a "just culture" (Reason, 1997). A just culture is defined by its capacity to treat its employees just. This entails that management is able to distinguish and draw a line between behaviour that is or is not blameworthy - this being the greatest challenge to establishing and maintaining a "just culture". To draw this line it is necessary to define "responsibility" since responsibility is the condition for holding operators "accountable". To define the nature or extent of responsibility is a classical problem within moral philosophy. One central

question when dealing with human error is whether one should hold operators responsible by the consequences of their act or solely by the intent behind their action.

In this paper the problems of reporting and dealing with human error within safety critical domains will be approached from a moral perspective and certain fundamental questions regarding reporting will be illuminated by reflections derived from moral philosophy. In order to do this the following question is addressed: Which are the moral aspects that should be considered in reporting and dealing with human error in safety critical organisations? Based on the results of an interview study of reporting cultures within Danish and Swedish ATC, this paper outlines the discussions involved in answering this question; the conclusion following; and the perspectives of developing a profession based ethics that support the conclusion.

Summary of results of interview study of Danish- and Swedish Air Traffic Control (ATC)

The reason for choosing Danish and Swedish Air Traffic Control ATC was, on the one hand, that the output of their respective reporting systems differed quite clearly and, on the other, that the two services appeared to be in most respects quite similar (training, safety record, capacity, national cultures etc.). At the time of the study the Swedish ATC centre (Malmø) had a relatively high rate of incident reporting whereas the Danish ATC centre (Kastrup) had a relatively low rate of reporting.[13] The study uncovered the reasons for difference in reporting practices by examining potential differences of organisational, legal, cultural and practical nature that might impact on reporting willingness.[14] In Table 1 below are summarised the main differences and similarities of organisational, legal, cultural and practical nature between the Danish and Swedish ATC (Jensen & Madsen, 2001) in order to prepare the subsequent discussions about moral aspects and professional ethics.

Moral aspects to consider

Based on the results from our empirical study the moral aspects that need to be considered in order to establish and maintain an effective reporting culture are identified and discussed briefly below. First, each moral aspect is introduced by posing questions of relevance to ATC and, next, brief (for reasons of space) answers are proposed to each of the questions, leaving out the complete line of argument and analysis. In the process of analysing moral aspects of reporting it becomes clear that most of them may be understood as part of the dilemma between safety (incidents reporting) and justice (sanctions and punishment). Although these moral aspects, in some respects, may change character within other domains, they give an impression of the complexity of issues that seem to be common to such domains.

Responsibility and accountability: Who is actually responsible for ensuring that the reporting system is working? In Denmark (Kastrup) the reporting system does not work well at all, and in some respects it is not optimal in Sweden (Malmø) either. There is no doubt that the organisations involved - and, in more general terms, the transport ministries and ultimately the politicians - have the overall responsibility for making the reporting system work. But is not the individual ATCO also morally responsible for reporting on dangerous conditions? Even if he or she may risk punishment, negative

[13] After the study was completed, the Danish law regulating the incident reporting in aviation was changed as of August 2001, securing confidentiality of reporting and ensuring indemnity against sanctions. The branches of the Danish CAA responsible for overseeing regulation compliance and the operation of ATC services, respectively, therefore changed their procedures and practices. The ATC service provider has since then succeeded in developing an effective reporting culture based on a non-punitive and mandatory reporting system. The new law dictates that employees who report incidents are secured indemnity, whereas if they choose not to report, they risk sanctions and possibly prosecution; in addition, it has been made a punishable offence to reveal the identity of persons who report on flight safety occurrences.

[14] In the study, a system-oriented approach was applied to investigate safety culture, using qualitative, semi-structured interviews, analysing sample incident reports produced by the two systems and studying the legal and procedural documents defining criteria for reporting, indemnity and sanctioning. Interviews were conducted with personnel on all levels directly or indirectly involved in reporting and handling of incidents: air traffic controllers, middle management representatives, regulators, legal department staff, departments of safety and analysis.

consequences or just disapproval from colleagues? Is it fair and just to require controllers that they incriminate themselves?

Table 1 - Differences and similarities in safety culture in Danish and Swedish ATC

Organisational level:	**Denmark**	**Sweden**
Similarities:	Organisational structure	
Differences:		
Legal system	Simple negligence punishable	Gross but not simple negligence punishable
Human Factors	Not integrated in investigation	Integrated part of investigation
Education	Reporting forms not introduced during education	Reporting forms introduced during education
Briefing	Seldom briefing after incidents	Always briefing after incidents (anonymous by choice)
Safety report	Safety issues not collected in one "place"	Safety issues collected as annual report
Transparency	No	Yes

Effects of legal system:	**Denmark**	**Sweden**
Knowledge of errors and incidents	Limited – because of secrecy	Wide – because of participation & openness
Explanation of error/incidents	Focus on individual human errors – and violations	Focus on interaction between human factors and system
Duration of investigation	Up till 2 years	Usually 3 months
Over a 7 year period	5 sanctioned	1 sanctioned

Individual level: Basic beliefs, values and norms:	**Denmark**	**Sweden**
Similarities:		
Responsibility	Feeling of responsibility toward safety	
Guilt & shame	Feeling of guilt and shame when making error / causing incidents	
Free will	No one makes errors on purpose – "deliberately"	
Punishment	Punishment does not prevent people from making errors	
Justice	Gross negligence should be punished	
Learning	They want to learn from error – wish for better reporting systems	
Support & understanding	General acceptance and support of colleagues who err	
Differences:		
Purpose of investigation	To identify the procedure that was violated, to assign blame	To optimise safety
Trust	Distrust	Trust
Barriers	Distrust and punishment	Lack of motivation and inconvenience
Reporting practise	No report on own error	Reporting – but sometimes only on local reporting form

Within an organisation everybody is responsible for making the reporting system work in so far as everybody has an obligation to support safety measures. On the one hand, the organisation has an obligation to create the right conditions for the employees to report. On the other hand, employees have a duty to report errors and incidents - but only as far as the employee does not incriminate himself. It is the established rights of employees to withhold information that can be used against them.[15] Even though withholding valuable information often made Danish ATC employees feel guilty, knowing they might be jeopardising on safety. So why not report? First of all because the ATCOs fear for punishment, secondly because they fear for unjust punishment. In principal employees should enjoy the right to be treated with fairness, which entails that the organisation has an obligation to seek "just

[15] The right to withhold information that may incriminate oneself is a principal right in criminal laws of democratic societies as well as in the European Human Rights Convention article 6.

treatment" for its employees regardless of the effects hereof. "Justice denies that the loss of freedom for some is made right by a greater good shared by others" (Rawls, 1972).

Free will, intent and negligence: In interviews, ATCOs often pointed out that no one makes an error on purpose. What is the bearing of this observation on the issue of accountability? Is this a way of saying that controllers who make mistakes never do so by deliberate carelessness or, perhaps, never do so out of carelessness at all? If so, is it not possible that some instances of simple negligence – e.g., failure to monitor – are effects of prior *free* choices?

The questions about intentions in actions (and omissions) are connected to the philosophical problems of free will. There are different schools of thought on the implication of the existence of free will. All though the existence of free may not be provable or disprovable, I suggest in this paper the following interpretation: "*just* punishment and *moral* condemnation imply moral guilt and guilt implies moral responsibility and moral responsibility implies freedom and freedom the falsity of determinism" (Strawson, 1974). In this sense, free will is the prerequisite of responsibility and as such not only does one become responsible for one's active choices and intentional actions, one also becomes responsible for one's passive omissions.

The ATCOs' statement that they do not make errors on purpose seems to imply that it is unfair to hold them responsible. But is it? To answer this question we need to make a digression. "To be responsible is to be the one which justly can be held accountable" (Ross, 1970). In other words to be held accountable the following must be true: one must be a rational being with a free will; and one must have had the possibility of acting differently in the given situation. Another way of describing it is to distinguish between prospective and retrospective responsibility where the ATCOs' prospective responsibility (their "safety critical responsibility") is "to do all that one can to promote safety". If the ATCO does not live up to his prospective responsibility he can be held retrospectively responsible, in other words – accountable. But what exactly are the implications of holding a "safety critical responsibility" or "to do all that one can to promote safety"?

The safety critical responsibility involves two steps. First, always being on level with one's competent colleague in terms of knowledge and competencies to do the job. Second, always being attentive of the dangers of the job, and being able to react on these, whether they are caused by oneself, by ones colleagues or by other factors within the environment.

Defining responsibility in this way connects it directly to the *intention* behind a given act and not to the *consequences* of that act. On the one hand, this definition makes it possible to overcome the phenomenon of "moral luck" (Nagel, 1997) since it is the intent with which the act is carried out that is the measure, and not the consequences of the act, which by chance may lead to an accident or no consequences beyond the error itself[16]. On the other hand, it demands a very high level of responsibility in the organisation. In effect the ATCO might *not* be responsible for management decisions, but he *is* responsible for speaking up against these if they reduce safety. Similarly, the individual ATCO is *not* responsible for his work conditions, but he *is* responsible for reacting against them if they jeopardise safety. Furthermore, the ATCO is *not* responsible for his colleague's actions, but he *is* responsible for reacting against "dangerous" colleagues. Finally, answering the question if the ATCO should be held responsible for his errors and mistakes, the ATCO should *not* be held responsible if he *in fact* did not have the possibility to act differently.

By its very nature the definition of error implies a non-intentional act. "Error will be taken as a generic term to encompass all those occasions in which a planned sequence of mental or physical activities fails to achieve its intended outcome, and when these failures cannot be attributed to the intervention of some chance agency" (Reason, 1990). But just because the act in itself might be unintentional does not necessarily imply that the person is free of moral responsibility. For instance if an ATCO is less attentive than his safety critical responsibility demands.

[16] There is no doubt however that in practice to leave all consequential considerations aside especially if an accident has had fatal consequences is difficult.

Guilt and shame: How should we react to the fact that controllers have feelings of guilt when they have made an error leading to a dangerous situation? Is the fact that the controller will feel guilt and shame a relevant argument against punishment or disciplinary actions?

Even though the controllers have strong feelings of guilt and they wear themselves down with self-blame it cannot stand alone as an argument against punishment or disciplinary action. Instead, it may be used as an argument to support and show understanding of the fact that "to err is human". ATCOs who have been involved in incidents tell us that they feel less safe as operators after the incident, partly because they sometimes do not receive much support.

Punishment: Is it justified to punish simple negligence? In Denmark, acts of simple negligence made by controllers (and other public employees) causing dangerous situations were punishable, in Sweden they are not. From a pragmatic point of view, it may be argued that it is futile to punish acts of simple negligence, since it has no preventive effect at all. But can this also be argued morally?
It is true that the preventive effect in regard to punishing for simple negligence is in fact very limited. According to the definition of responsibility endorsed by this paper, one cannot always be held accountable for simple negligence, which in effect may make it morally unjustifiable to punish for simple negligence. In legal terms it could be miscarriage of justice.

'Blame-free': In Sweden controllers are not sanctioned even when they have acted negligently (exempting gross negligence) which, in turn has made it possible for employees to gain trust in a "blame-free" system. Is this practice acceptable or should we punish 'guilty' or negligent controllers regardless of the effects on reporting willingness? Is it morally acceptable to refrain from holding ATCOs accountable just because we wish to promote a reporting culture?

There is, to our knowledge, no study that shows positive effects of punishment on willingness to report. It appears near self-evident that operators will be reluctant to run the risk of reporting if there is chance, albeit small, of being punished. The problem in the case of the Danish ATCOs is that they had the impression that they could not predict which types of errors or incidents that will be met with sanctions; they therefore simply chose not to report. They had no trust that they would be met with fairness, because they had experienced colleagues being punished for minor mistakes.

In principle, it is not morally acceptable to let "the guilty go free", but in the case of promoting reporting as a means to supporting safety it is morally acceptable according to the stand of consequential ethics.

Violations: How are we supposed to react to the alleged fact that controllers routinely violate rules in order to maintain efficiency? Most controllers agree that they bend the rules, but they also insist that they never do so in a way that will jeopardise safety. Should we accept this practise as ordinary professional behaviour, or should we argue that intentional violations of rules be sanctioned? To what extent is management responsible for a working culture that - because of pressure on production - induces its employees to routinely violate rules?

Most studies show that violations are rather common within professional cultures, but the studies also show that there are different reasons for violating rules. These can be categorised into different types of violations (Reason 1990 & 1997): *Routine* violations are very common and are usually motivated by the fact that it is easier to perform a given act by bending a rule. Management, moreover, will nearly always know that these violations do happen as a matter of routine. *Optimising* violations are chosen for personal reasons, "for the thrill of it", i.e. when driving too fast. *Necessary* violations are usually performed in situations of high workload, i.e., pressures of production. Another category close to the last one is the *exceptional* violation that happens very seldom and which is often a reaction to something that has already gone wrong.

From a moral perspective these violations are judged differently in terms of the intent and in terms of the responsibility we can assign. Where optimising violations are morally unacceptable, routine

violations are slightly more complicated since some procedures may in fact be inexpedient. In short when violations become routine, this development - and each single instance - becomes the shared responsibility of the organisation according to the *primary principle of accountability* (PPA). This principle says that a moral agent is responsible for his intentional acts; whereas the *extended principle of accountability* (EPA) says that the organisation as such (management) is responsible for the "second effects involving the actions of other persons" (French, 1988). An organisation that induces employees to violate rules of safety because of inadequate resources or pressures of production is, according to the EPA, responsible for the consequences of such acts.

Violations may never be eliminated in professional cultures but they should however be minimised and controlled so that they do not create unsafe situations. It is in this regard in particular that a professional ethics may be able to provide the means to defeat violations through the establishment of shared and safe norms.

Safety: Many of the questions posed here relate to the basic question: What should we morally accept for the sake of safety? By which moral principles is it possible to put safety considerations over other considerations? To answer this question is also to answer the question that was posed in the beginning. In the following section the answer is formulated in terms of a principle and I shall seek to sum up how the various moral aspects should be considered in relation to each other.

How to consider moral aspects of reporting

When evaluating the moral aspects that should be considered in reporting and dealing with human error in high reliability organisations it is worth stressing the, perhaps obvious, fact that the purpose of dealing with (human) error is to improve safety. In pursuing this purpose one needs to operate within the framework constituted by the safety critical work domain in which employees enjoy rights: the right to avoid self-incrimination as well as the right to be treated fairly. A fair treatment entails that possible error needs to be evaluated in terms of the employee's responsibility, work conditions and his possibility for having acted differently in the given situation. Focus should be on the employee's *intention* in carrying out the act and *not the actual consequences* of the act. In cases of accountability and blameworthiness the measures taken toward the possibly blameworthy operator should consider both safety (including the general will to report) and justice – in such a way that safety in principle weighs the most. Consequently, I would argue that the use of sanctions should be very much confined and punishment only applied in the most serious cases.[17]

Still, (from a deontological perspective), there may be a problem with this conclusion, not in terms of its consistency, but in terms of its moral implications – the requirement of "unjust lenience" – jeopardising justice. As an extension of and perspective to the conclusion I shall therefore suggest that the development of a professional ethics is a feasible way of attaining a proper balance between safety and justice and ensuring a greater involvement and feeling of responsibility by operators.

Professional ethics

The argument of this section is that a professional ethics - based on common norms derived at by the operators themselves through a process of dialogue among themselves and with management - will be able to reclaim justice by reducing unacceptable behaviour. Thus I propose to replace regulation solely by means of external sanctions by regulation guided by an internally motivated responsibility established by the operators themselves. In general a professional ethics will be able to:

- reinforce the profession in its commitment to safety
- develop a working environment where ethical, i.e. responsible, behaviour is the norm
- function as a guide in specific situations (dilemmas)
- function as a tool for education and socialisation of new operators
- indicate to the outside world that this is a profession that is responsible and cares about safety (Olson, 2000)

[17] Serious cases refers to the operators acts, for instance acts of severe [gross] negligence, not the consequences of their acts.

It is argued, further, that the process of developing a professional ethics will be of value to the profession. Developing the ethics will draw focus to general safety behaviour, which is expressed through daily practice and management. In this regard, the ethics should seek to modify employees' behaviour as well as management strategies. For example, employees should, as a professional group, be concerned with bettering work conditions regarding e.g., pressures of production, stress and workload, inappropriate procedures; equally management should be 'open' towards possible change.

The objective of the professional ethics is to develop a single common ethics and common set of standards. It goes without saying that each and every air traffic controller has an ethics and feels an extraordinarily high responsibility towards safety; but what is missing is what they themselves express as "one common ethics"[18]. Although the air traffic controllers are characterised as a relatively small profession with a strong common culture, one still needs to differentiate between them – they are not one single culture. The empirical study revealed that besides differences across teams in ways of "moving traffic" (flights), there also were individual differences i.e. macho-types and cowboy types. Despite general acceptance amongst ATCOs of their differences and individual ways of controlling flights, they do express consideration about colleagues who practice unsafe behaviours. A single common ethics could therefore be a possible way to modify these 'unsafe' behaviours.

In order to fulfil the described objectives a professional ethics for air traffic controllers should seek to create an effective reporting culture and strengthen the organisation's safety culture. The individual steps of such an effort, I propose, consist in:

Regulating employees' behaviour:

- motivate inner responsibility (internalisation of professional responsibility)
- develop a common ethics (taking moral consideration for one's colleagues)
- develop appreciation for the causal relation between errors, incidents and accidents

Regulating management strategies:

- control pressure of production

Bettering work conditions:

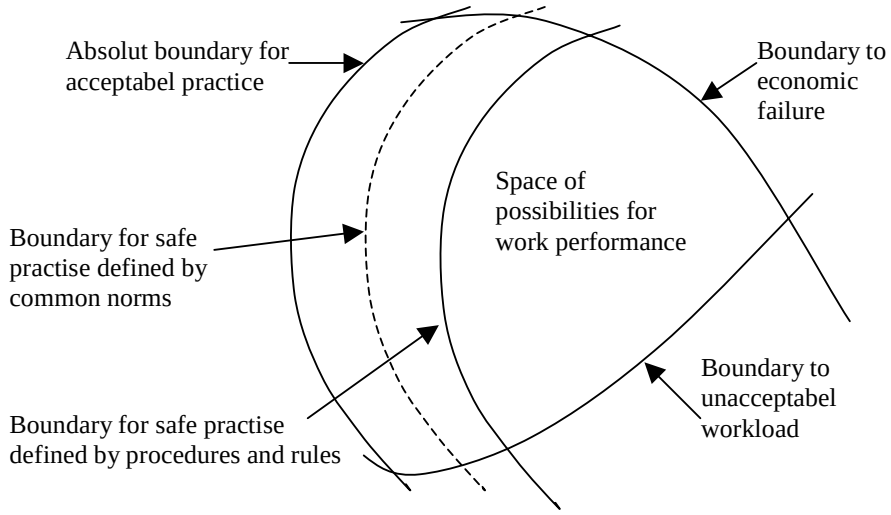
- enhance standards by getting rid of 'bad practice' as well as inexpedient procedures
- strengthen the possibility of objecting against the "system"
- redefine what is "good" versus "bad" practice – developing an ethical environment

To develop a professional ethics is a long and dynamic process. In order to develop a common ethics it is of outmost importance that the professionals themselves through dialogue define the rules and norms. Although it is necessary and indeed crucial that management supports the idea and that middle management and regulators are involved somewhat in the process, it should primarily be in the hands of the professionals themselves – giving them ownership of safety. To develop a professional ethics will also reflect a recognition that it is not possible – with the same degree of success - to control the ATCOs' behaviour through external regulation, and that strategically one should therefore let the ATCOs regulate their own behaviour through a single common ethics. In other words to choose "responsibility" as attitudinal- and behavioural modification strategy.

Process: The process of creating a professional ethics may be described through the following model (suggested by Rasmussen, 2001, for somewhat different purposes):

[18] Interview with Preben Lauridsen, air traffic controller and past IFATCA President.

Figure 1 – Illustration of the boundaries of safety critical work performance



First of all it is necessary to discuss the values and virtues that characterise a good practitioner (Pritchard, 2001). In doing this it will be helpful to take departure in actual cases – to ask what the "good ATCO" would do in this situation. As the model of Figure 1 shows there are many different considerations to make, but most importantly the ATCOs need to make explicit and define the boundaries within which it is possible for them to operate safely. On the one hand, procedures might be so restrictive that it can be almost unsafe to comply with them – "skill and expertise make rules redundant. If people are highly trained and practised at a particular task, rules are no longer required to control their actions" (Lawton, 1998). On the other hand there might be some "macho types" who override rules and procedures even when it is not safe. It is therefore in the interface between procedures and "absolute boundary for acceptable practice" that the ATCOs need to articulate and lay down their common norms. A technique that has been implemented successfully (e.g., Swedish and Danish ATC) in maintaining (and redefining) safe boundaries is to have a team of ATCOs to discuss actual incidents (including violations) with the involved ATCO to evaluate the performance in terms of the chosen norms and decide if it the performance or the norms need adjustment or reformulation.

Conclusion

In order to overcome the dilemma between safety (reporting) and justice (external regulation) it is necessary to practice what I have called unjust lenience. However, as long as the moral implications of this can be dealt with by developing a profession-based ethics, it can be justified morally. Furthermore the implications of defining responsibility in terms of intentionality demands a high degree of shared responsibility towards safety and as such calls for common standards of safety practice. A profession based ethics makes this possible.

The idea of developing a professional ethics as a means of behavioural modification within safety critical domains is only in its beginning. Therefore future work will be concerned to investigate the possibility of developing professional ethics as means to create a healthy safety culture in other safety critical domains e.g. health care. Furthermore I suggest that it is possible to integrate the ideas that lie within a profession based ethics into the theoretical framework of Safety Culture Maturity Models (Flemming, 2002). Many questions are still left unanswered. Is a professional ethics the solution for all types of professions or professional groups working in safety critical domains? Does it require a degree of cohesion that may be lacking in groups such as e.g., professional truck drivers and taxi drivers? Perhaps the establishment of a professional ethics requires certain characteristics of the profession? Is it possible to define criteria in terms of certain "professional" characteristics that have to be met in order that a professional ethics will work at all as a basis for self-regulation of behaviour? Should a professional ethics be founded locally, nationally or

internationally? The type of professional ethics presented in this paper is most suited as a local or national based ethics, still it is possible that international professions would gain by creating a common ethics?

Acknowledgements
I am heavily indebted to Henning Boje Andersen for his critical and constructive review on this paper as well as idea generation and inspiring discussion. I would also like to acknowledge Thomas Ryan Jensen for his work and corporation on the interview study of Danish and Swedish Air Traffic Control from which many of the ideas in this paper derive.

References

Bayles, M.D. (1988). The professions. In Callahan, J.C. (ed.), *Ethical issues in professional life*. Oxford University Press, USA.

Flemming, M. (2001). Safety Culture Maturity Model. www.hse.gov.uk/research/otopdf/2000/oto00049.pdf visited 2002.06.20.

French, P.A (1988). What Is Hamlet to McDonnel-Douglas or McDonnel-Douglas to Hamlet: DC-10. In Callahan, J.C. (ed.), *Ethical issues in professional life.* Oxford University Press, New York, USA.

Helmreich, R.L., and Merritt, AC (1998). Culture at work in aviation and Medicin: National, organisational and professional influences. Ashgate, Aldershot, UK.

Jensen, T.R. & Madsen, M.D. (2001). Filosofi for flyveledere: En undersøgelse af hvilke moralske aspekter man bør tage hensyn til ved behandlingen af menneskelige fejl i sikkerhedskritiske organisationer. [Philosophy for Air Traffic Controllers: An investigation of moral aspects to be considered in the handling of human errors in safety critical organistions]. MA thesis, philosophy and communication studies. Roskilde University, Denmark.

Maurino, D., Reason, D., Johnston, N. & Lee, R. (1995). *Beyond Aviation Human Factors*. Avebury Aviation, Ashgate Publishing, Aldershot, UK.

Nagel, T. (1997). *Spørgsmål om livet og døden*. Samlerens Forlag, København.

Olson, A. (2000). Authoring a code: Observations on process and Organization. Center for Study of Ethics in the Professions, Illinois Institute of Technology, Codes of Ethics Online, http://csep.iit.edu/codes/coe/Introduction.html

Pritchard, M.S. (2001). Responsible Engineering: The importance of Character and Imagination. *Science and Engineering Ethics* 7:391-402

Rasmussen, J. (1997). Risk management in a dynamic society, a modelling problem. Safety Science 27:183-214.

Rawls, J. (1971). *A Theory of Justice*. Oxford University Press, England.

Reason, J. (1997). Managing the Risks of Organizational Accidents. Ashgate, England, 1997

Reason, J. (1990). Human Error, Cambridge University Press, England.

Rowland, D. (1999). Negligence, Professional Competence and Computer Systems. *Journal of Information, Law and Technology*, 2.

William F.M. (1988). Professional Virtue and Self-Regulation. In Callahan, J.C. (ed.), *Ethical issues in professional life*. Oxford University Press, USA.

Forensic Software Engineering and Stories of Failures

Darren Dalcher,

Software Forensics Centre, Middlesex University, London, N14 4YZ, UK.
d.dalcher@mdx.ac.uk

Abstract: Researchers with a keen interest in information systems failures are faced with a double challenge. Not only is it difficult to obtain intimate information about the circumstances surrounding such failures, but there is also a dearth of information about the type of methods and approaches that can be utilised in this context to support information collection and dissemination. The purpose of this paper is to highlight some of the available approaches and to clarify and enhance the methodological underpinning, that is available to researchers interested in investigating and documenting phenomena in context-rich and dynamic environments that characterise many software failures. The focus of the discussion is on approaches to information systems failures that value the situational meanings and knowledge of participants and a naturalistic research perspective, while at the same time advocating a mixture of quantitative and qualitative evidence and analysis.
The paper begins by introducing forensic software engineering and the need to understand failures through the consolidation of a diverse range of subjective accounts offered by participants. Knowledge relating to failure is fragmented, distributed and hidden within the context requiring a naturalistic enquiry process. Moreover, untangling causes is an inherently pervasive problem due to emergent properties and the inability to delineate causes from effects. The solution in the form of case-based methods provides an approach that can capture subjective knowledge and situational meaning, but requires a new perspective offered through detailed and chronological case histories of failures. The paper concludes by proposing the supplementing of case histories with narrative inquiry which extracts fragments of stories that emphasise the multiplicity of views and perceptions and their critical interactions during the lead-up to disaster, thereby capturing the shared knowledge pertaining to failures.

Keywords: software failures, case study, case history, narrative inquiry.

Setting the Scene

Failure information tends to arrive from different sources typically including: anecdotal evidence and journalistic descriptions, reports of investigative committees, official public inquiries, audit reports, public account committee minutes and findings, case studies and empirical surveys. In specific cases this may be augmented by internal documents, interviews, eyewitness accounts, direct observation and archival records (and possibly physical artefacts). In most cases the researcher is exposed to specific information post-hoc, i.e. once the failure is well established and well publicised and the participants have had a chance to rationalise their version of the story. Most of the available sources are therefore already in place and will have been set up by agencies other than the researcher.

Forensic is derived from the Latin 'Forensis', which is to do with making public. Forensic Science is the applied use of a body of knowledge or practice in determining the cause of death. Nowadays extended to include any skilled investigation into how a crime was perpetrated. Forensic systems engineering is the post-mortem analysis and study of project disasters (Dalcher, 1994). The work involves a detailed investigation of a project, its environment, decisions taken, politics, human errors and the relationship between subsystems. The work draws upon a multidisciplinary body of knowledge and assesses the project from several directions and viewpoints. The concept of systems is a central tool for understanding the delicate relationships and their implications in the overall project environment.

The aim of forensic analysis is to improve the understanding of failures, their background and how they come about (Dalcher, 1997). The long-term objectives are improving the state-of-the-practice and generating new insights into methods of managing complex projects. The knowledge generated is then fed-back into the process via a double loop learning system in order to improve the internal (organisational) or external (disciplinary) body of knowledge. Forensic software engineering is thus primarily concerned with documentary analysis and (post-event) interviews in an effort to ascertain responsibility lines, causal links and background information.

The primary mode of dissemination of findings, conclusions and lessons is through the publication of case study reports focusing on specific failures. However, there are limited research methods to explore the

dynamic and fragmented nature of complex failure situations. The armoury of research methods in this domain is often limited to main dissemination mode; case studies. Lyytinen and Hirschheim (1977) noted that more qualitative research methods were needed for IS failure research as well as more extensive case studies that explored problems in more detail and viewed solution arrangements in light of what transpired. The same methods also need to account for group issues and cultural implications. Sadly, 25 years on, the same constraints in terms of methods are still in evidence.

Failure Research — Looking Beyond the Objective

The choice of a research method depends on the type of information that is available to the researcher. The positivist stance, prevalent in the natural sciences, is centred on the notion that all knowledge, in the form of facts, is derived from either observation or experience of real, objective and measurable natural phenomena thereby supporting the notion of quantitative analysis. Facts can thus be viewed as universal truths devoid of personal values and social interactions and independent of time and context. This enables researchers to focus on regularity, repeatability and the verification and validation of causal relationships. The currency of such objective knowledge is the manipulation and metrification of objects and their relationships expressed in the form of numbers, to enable quantitative operations. This stance is difficult to sustain in failure research where the actions, perceptions and rationales of actors are not amenable to quantitative methods. (Note however that the actual findings and the factors leading to accidents can subsequently be modelled using quantitative notations.)

At the other extreme, (interpretivist, constructivist or relativist) knowledge can be viewed as encompassing beliefs, principles, personal values, preferences, social context and historical background which are inevitably dynamic as they change with time (and context). Qualitative research methods originate in the social sciences where researchers are concerned with social and cultural phenomena. Social interaction in human activity systems ensures intersubjectivity as actors are forced to negotiate and agree on certain aspects. The humanistic perspective is outside the conventional positivist norm. The resulting emphasis is on the relevant interpretation of knowledge as held by participants in a social activity. Data sources utilised by researchers include observation, fieldwork, interviews, questionnaires, documents, texts, and the impressions and reactions of the researchers. Such qualitative perspective relies on words (Miles and Huberman, 1994), conveying feelings and perceptions, rather than numbers. Qualitative methods recognise the fact that subjects can express themselves and their feelings and thereby, clarify the social and cultural contexts within which they operate. Meaning therefore needs to be ‘interpreted’ in a process of ‘sense making’. Indeed, Kaplan and Maxwell (1994) argue that the goal of understanding a phenomenon from the point of view of the main participants and their particular social, cultural and institutional context is largely lost when the textual data are quantified.

Making sense of failures retrospectively is difficult. In general, there is very little objective quantitative failure information that can be relied upon. Indeed, a specific feature of failure is the unique interaction between the system, the participants, their perspectives, complexity and technology (Perrow, 1984). Lyytinen and Hirschheim (1977) pointed out that failure is a multifaceted phenomenon of immense complexity with multiple causes and perspectives. Research into failures often ignores the complex and important role of social arrangement embedded in the actual context. More recently, Checkland and Howell (1998) argued that the IS field requires sensemaking to enable a richer concept of information systems. Understanding the interactions that lead to failures likewise requires a humanistic stance outside the conventional positivist norm to capture the real diversity, contention and complexity embedded in real life.

(Note that triangulation, the mixing of quantitative and qualitative methods, offers the opportunity to combine research methods in one study. A good example of such a mix in failure research would entail reliance on qualitative methods to capture the essence, context and webs of interactions in the build-up to failure and complement the presentation by using more formal approaches to model the impact of such interactions.)

Knowledge is Hidden Within the Context

Qualitative research methods are concerned with generating richer knowledge. However, knowledge is not something which exists and grows in the abstract (Boulding, 1956); It is a property of the interaction between agents and the environment and is tied to perspectives, intentions and perceptions. Meaning is therefore not an intrinsic property of a message but depends on the code or set of alternatives from which the message comes (Ashby, 1960; Campbell, 1982; Lissack, 1999). In fact, Nadler (1985) noted that ‘information is not a brick that can be thrown from one person to another with the exact same meaning’.

Knowledge is deeply bound to its original context which enables a 'contextually correct' understanding. It is also strongly coupled to the time frame, and thereby, to the prevailing mindset. Only through the effective capturing of the precise context, can information be evaluated against the rationale, motives and assumptions that applied. Any useful information must therefore be accompanied by additional contextual information that will shed light on its utility, validity and relevance.

The fact that a failure phenomenon is being investigated, suggests that attention has already been drawn to the complexities, breakdowns and messy interactions that such a situation entails. Many such inquiries deal with subjective accounts including impressions, perceptions and memories. The aim of the researcher is to increase in a systemic way the understanding of a situation, yet do so from a position that takes in the complexity of the entire situation and incorporates the different perspectives and perceptions of the stakeholders involved. Phenomenology can thus be described as the study of direct experience taken at face value and may utilise verbal, diagrammatic or descriptive model forms (Remenyi et al., 1998). The focus is on what the subject experiences and its expression in a language and mannerism that is loyal to that experience.

Methods used to research failures need to be systemic and be able to get beneath how people describe experiences to the underlying structure and webs of interactions. Such methods need to:

- offer a holistic view unravelling a systems perspective on the entire topic of study thus enabling researchers to ascend beyond the details – in failure research this enables the investigator the glimpse the 'total system';
- be an inductive approach that enables the construction of meaning in terms of the situation and the development of general patterns that emerge from the cases under study;
- enable researchers to extend the boundaries of the system to capture interactions that may impact the failure; and,
- support naturalistic enquiry enabling phenomena to be understood in their naturally occurring settings

Overall, the purpose of a research method is to enable the researcher to make sense of the complexity of detail and the complexity of interaction and chart the contributory role of different causes and issues in the build up to failure.

Emergence and Simplified Causality Complicate Investigations

Interestingly, many failure investigations try to reduce failure explanations to simple causal pairings (Lyytinen and Hirschheim, 1977) thereby ignoring the role of participants, their knowledge, assumptions and the overall environment. Cause-event relationships do not tend to be objective (Checkland and Holwell, 1998). Interpretivism calls the possibility of uncovering causal links into question because all entities are in a state of 'mutual shaping' so it is impossible to distinguish causes from effects (Thietart, 2001). Moreover, each failure is unique. In many cases complex interactions between actors, systems and failure causes play a part in creating a dynamic (and messy) mix. Therefore it is more important to try to understand the meaning that actors give to reality, as intentions, motivations, expectations, beliefs, perceptions and fears are all grounded in practice. Failure research must proceed by taking into account the sum of all interactions and their dynamic co-linear relationships.

The general phenomena of emergence defies causal analysis forcing greater emphasis on interactions. All systems are composed of inter-parts and the system can only be explained as a whole. Accidents and failures display similar tendencies as unexpected and 'interesting' interactions and properties emerge. When interactions occur in a certain way and order, they give rise to emergent (and often unexpected) patterns of behaviour. The complexity and interconnectedness of interacting components and agents thus gives rise to emergent phenomena. Emergence resulting from such synergies, intra-acting interactions and non-linear dynamics is represented by new properties, capabilities and behaviours of the overall system. All too often emergent properties are neither designed nor planned. Slight changes in input or interaction patterns will thus lead to differences in emergence (i.e. unexpected new behaviours).

Proofs of causality are inevitably tenuous (Lowrance, 1976). Moreover, due to emergence and unexpected interactions, forming a direct link between cause and effect is rather complicated and somewhat misleading (Perrrow, 1984). Cause-effect relationships involve uncertainties in both directions. In principle, separating cause from effect depends on the assumption of stability and minimum change within the environment. In practice, one is often faced with events (or potential events). A more realistic approach is to focus on an event, and trace the range of causes and the effects that have resulted from them:

Cause ← Event → Effect

This approach enables the identification of multiple causes and multiple effects from the same event (as well as the detection of multiple events resulting from the same cause).
One of the major complications in failure investigations is in relating causes to effects (and possibly events) through extended time horizons (Dalcher, 2000). The implications of actions may not be witnessed for years, or even generations. Delays between making a decision and observing the result distort the causal link between the two. As a result, people tend to associate a different level of severity to events occurring following a delay. The perceived severity is thus diminished with the length of the delay further complicating the task of identifying patterns and interactions that contributed to a given failure. Failure researchers are thus required to provide adequate historical accounts of the interactions between actions, perceptions and the passage of time.

Using Case Studies to Describe Reality
Having looked at some of the complications associated with capturing actions, reactions and perspectives, it is now time to turn our attention to the main tool of forensic IT research, the case study. The term "case study" is an umbrella term used in different contexts to mean different things which includes a wide range of evidence capture and analysis procedures. Yin (1993) defines the scope of a case study as follows:
"A case study is an empirical inquiry that:

- investigates a contemporary phenomenon within its real-life context, especially when
- the boundaries between phenomenon and context are not clearly identified"

A case study can be viewed as a way of establishing valid and reliable evidence for the research process as well as presenting findings which result from research (Remenyi, 1998). According to Schramm (1971) the case study tries to illuminate a decision or a set of decisions and in particular emphasise why they were taken, how they were implemented and with what results. A case study is likely to contain a detailed and in-depth analysis of a phenomenon of interest; in our case, the failure scenario.
The general aim of the case study approach is to understand phenomena in terms of issues in the original problem context. A Case study allows the researcher to concentrate on specific instances in their natural setting and thereby attempt to identify the interacting perceptions, issues and processes at work resulting in an in-depth study. Some of these interactions are likely to prove crucial to the success or failure of the organisation/system under scrutiny. Focusing on relationships and processes facilitates a holistic perspective revealing underlying patterns, and possibly some emergent properties. Many of these patterns remain hidden under normal conditions, but can be prised open as a result of the special focus. (Note that case studies may contain rigour and application of careful logic about comparisons in the positivist tradition.)
In the context of failures, exploring a particular case or set of events entails attempting to provide the richest perspective of what transpired through the analysis of multiple subjective accounts of participants, the explanation of phenomena and the retrospective identification of relationships. Case studies provide the mechanism for conducting such an in-depth exploration. They often result from the decision to focus an enquiry around an instance or an incident (Adelman et al., 1977), as they are principally concerned with the interaction of factors and events (Bell, 1999). Indeed, sometimes it is only the practical instances that enable one to obtain a true picture of the interaction (ibid.). The combination of a variety of sources offers a richer perspective which also benefits from the availability of a variety and multiplicity of methods that can be used to obtain new insights about this single instance.
Case studies are more likely to be used retrospectively rather than as an on-going perspective (especially from a failure point-of-view), as researchers are unlikely to know the potential for useful results and interest from the outset. Case studies are useful in providing a multi-dimensional picture of a situation (Remenyi, 1998) in the context of historical description and analysis. The richness of detail can be controlled through the careful placement of systems boundaries and consideration of the wider system environment that is relevant to the phenomenon under study. Case studies can be utilised as source of understanding which is tolerant of ambiguity, paradox and contradiction. A case study is viewed as interpretative when the events in the real world are observed and then an effort takes place to make sense of what was observed, i.e. when one tries to make sense of a failure from the perspectives of participants. They also offer the potential for generating alternative explanations from the different stakeholder perspectives thereby allowing the researcher to highlight contradictions and misunderstandings.

Information collection methods for case studies often use observation, document reading and interviews, but other methods can be selected to suit the particular requirements of a case. Case Study work needs to self-contained, but researchers have the luxury of being able to expand the boundaries to incorporate emerging patterns and perceptions. The data, and indeed the analysis, are grounded in reality.
The main advantages of using case studies include:

- ✓ ability to identify and focus on issues
- ✓ richness of detail
- ✓ multiple perspectives
- ✓ multiple explanations (no absolute truth)
- ✓ cross disciplinary remit
- ✓ ability to minimise inherent complexity
- ✓ ability to show interactions
- ✓ ability to observe emerging patterns
- ✓ conducted in real-life setting
- ✓ encompasses original problem context
- ✓ ability to deal with interpretations
- ✓ can extend the boundaries to include aspects of wider system environment

The main objections to their use include:

- ❖ sometimes viewed as soft data
- ❖ biases inherent in accepting views and perceptions
- ❖ questions about generalisability of findings (especially from a single case), but it is possible to build a library of such cases
- ❖ issues regarding objectivity of approach
- ❖ negotiating access to settings
- ❖ boundaries are difficult to define, but this could also be a strength!
- ❖ mainly retrospective
- ❖ the observer effect
- ❖ reliability of conclusions
- ❖ there is little control over events, but this may also be a strength

In summary, case studies are ideal for exploring interactions between people and their understanding of a situation. The richness of the data obtained by multiple means from multiple perspectives provides a real insight into the main issues at play. The time dimension (sequencing) is critical to understanding interactions and identifying their impacts. Actions (and reactions) can only be understood in context, and case studies create the context for understanding them. Emergence often defies causal analysis forcing a greater emphasis on interactions; however, case studies enable the identification of networks of issues that people are likely to act on. The general use of the case study requires a tighter definition of its meaning in failure research.

From Case Studies to Case Histories

While there may be a tradition of using case studies within the IS community, this is perhaps more often borrowed from the MBA culture than as a result of self conscious effort to adopt them as a research approach (Walsham, 1995; Cornford, 1996). The shift to studying the impact of issues within the organisational context renders case studies particularly useful for investigating failure scenarios. However, the use of the term often leads to some confusion. Case studies have been used to adopt an idiographic (Cornford, 1996), an interpretivist (Walsham, 1993), a constructive (Jankowitz, 2000) or even a positivist (Yin, 1989; Yin, 1993; Benbasat et al., 1987) approach.
After Walsham (1993), we take the view that interpretivist case studies develop deeper understanding of IS phenomena; i.e. software failures. The shift from technical to organisational issues (Benbasat et al., 1987) necessitates a deeper look at how people act on interpretations and perceptions. Generating explanatory models enables expressions of patterns, judgements and values that provide a systemic clue to the unfolding of events. Case studies are typically used to explore issues in the present and the past and comprise of ethnographic studies, single case studies and comparative case studies (Jankowicz 2000). In our experience there is a need to add the failure case study as a special example of a case study focusing primarily on the

background, context, perception, interactions and patterns. We thus propose the use of the label **Case Histories** to refer to the specialised studies focusing on failure incidents.

Case histories are concerned with providing the background and context that are required to endow words and events with additional meaning. Background refers to previous history of the system itself, while context refers to interactions with the environment. As failures are time- and place-dependent, the case history framework enables readers to obtain an understanding of the intimate context surrounding the main event. The primary tool available to the community is the Case Histories of failures (derived from the use of the case study method). These represent a detailed historical description and analysis of actual processes. Their value is in tracing decisions (and recorded rationale) to their eventual outcomes by utilising techniques borrowed from decision analysis and systems engineering. Indeed, the historical description and the presentation of a chronology are based on the recognition that real life is complex, ambiguous and conflicting.

Case histories highlight complexities and trade-offs that are embedded in the acquisition and development processes or in the operation and interaction mode. They also help in the identification, definition and assessment of pervasive problems in a given application domain. Maintaining repositories of forensic case histories is a form of risk management and hopefully, mitigation that can be applied to future undertakings (Dalcher, 2002). Failures are crucial to the development of a mature and responsible discipline that responds to crucial issues that emerge from past failures. Case histories thus aid in understanding the role and significance of failures.

Recommendations: Case histories contain observations, feelings and descriptions. They can be used to construct, share, dispute and confirm meanings, interpretations and scenarios in the context of real events. Such observations must be systematically processed and structured. Their validity depends on the procedures used to obtain the information. Where possible, multiple sources of evidence should be used to support the emerging story. A mix of methods for obtaining the information will also enhance the value of the result. The use of alternative perspectives enables the analyst to consider conflicts and varying perceptions and their role in the unfolding story. Finally, case histories should be composed in an engaging manner to provide convincing reading (Remenyi et al., 1998) with a clear and concise story. However, constructing a convincing narrative of a complex story is a challenge in itself.

Stories are Narrative Inquiry

As we have seen, failures in common with other organisational activities are based on stories. The verbal medium is crucial to understanding behaviour within organisations and systems, and researchers are thus required to collect stories, grounded in practice, about what takes place (Easterby-Smith, 2002). Understanding failures often entails the untangling of complicated webs of actions and events and of emergent interaction patterns.

Historically story telling has been an acceptable form of conveying ideas, experience and knowledge of context. It plays a key role in communicating the cultural or historical context to the listener. Moreover, children are often initiated into culture (and its boundaries) through the medium of story telling.

In practice, the essence of any good case study revolves around the ability to generate an effective storyline. In a large case, a general theme can be obtained from selected excerpts weaved together to illustrate a particular story. Personal stories that form part of a case study can thus be viewed as a valid source of data organised to make sense of a theme or problem. This is particularly useful when the researcher is trying to portray a personal account of a participant, a stakeholder or an observer in an incident, accident or failure. The implication is that the need to address personal aspects of interaction and story (that remain a problem in failure research) is fulfilled by the development of a research-valid narrative. Indeed, Remenyi et al. (1998) contend that a story or a narrative description is valid if the resulting narrative adds some knowledge.

A narrative can be structured to give a voice to the researcher, to the narrator, to the participants, to the stakeholders or to cultural groups, traditions or ideas. In the context of research it is not concerned with the development of a reflective autobiography or life story but rather with the analysis and devolvement of themes (Bell, 1999). Researchers are thus concerned with how information interpreted from a story can be structured in such a way as to produce valid research finding. This form of narration can be particularly useful in uncovering motives and rationales and linking them to the actual consequences and their impact on stakeholder groups. It also suggests an understanding of implied cases as well as emergent interactions.

Failure researchers collect subjective accounts extracted from participants and observers. Developing narratives relies on trust between the researcher and the storyteller. Storytellers reveal personal feelings and motivations which may compromise their position or interests. Sharing the information, and making it public suggests that the storyteller is prepared to release certain details about themselves and their position publicly. This may have ethical research implications. Shared stories imply shared concepts, shared vocabularies and shared perceptions (or as a minimum, the ability to see where the sharing stops).

Narrative inquiry is evolving into an acceptable research approach in its own right in the social sciences and in management research circles (Bell, 1999; Easterby-Smith, 2002; Boje, 2000). The story format provides a powerful way of knowing and linking disparate accounts and perspectives. The main pitfall with this approach revolves around the narrative structure which is developed by the storyteller. If the initial storyteller is not the researcher, care should be taken to eliminate personal biases in terms of outcomes and actions (but these should remain as descriptions of feelings, reactions and motivation). Follow-up questions can thus provide the mechanism for clarifying context, background, rationale or sequence, or more generally for 'objectifying' and 'time-sequencing' the events. When different accounts are combined the story line benefits from the richness of multifaceted insights.

Developing a narrative requires plot as well as coherence (Boje, 2001). In failure stories the plot often emerges a results of the actions and perceptions of participants. Boje (2001) contends that most real life stories are fragmented, non-linear and incoherent. This has already been highlighted as a feature of failure stories. Such stories also tend to be dynamic, polyphonic (multi-voiced) and collectively produced. The stories are not plotted as such and they appear to flow, emerge and network offering complex clustering of events, emergent phenomena, causes and effects. Moreover, the accounts are subjective and often contradictory.

Stories appear to be improperly told, as a story is an 'ante' state of affairs existing previously to a carefully constructed narrative (ibid.). The **antenarrative**, or the 'real' story, is the fragmented, messy and dynamic multi-voice multi-version and complex tale. In the tradition of post-modern inquiry, a real life researcher is often faced with fragments rather than a whole story to tell; and many of the fragments may reflect contrary versions of reality. This is potentially more acute when the accounts attempt to justify roles of participants in the lead-up to disaster. It would also appear from past analysis that there are hierarchies of stories and stories that exist within or interact with other stories. Using the terminology provided by Boje, the purpose of narrative methods is to take a complex situation characterised by collective (yet often conflicting) memory and an **antenarrative** and construct the plot and coherence that can be used to narrate the story of interest.

The reality in failure stories is of multi-stranded stories of experiences and reactions that lack collective consensus. Indeed the discipline of decision making has also recognised that making choices is about forming and selecting interpretations from a mosaic of possibilities (March, 1994; Weick, 1995; March, 1997). Not surprisingly, disasters or traumatic stories are hard to narrate, understand and justify. Stories have three basic properties: time, place and mind (Boje, 2001) which interact and build up as the story evolves. In forensic case histories, these are further clarified through the identification of the background and context which clarify and justify the interpretation in the context of the emerging phenomena.

Boje (2001) argues that the current view is of sequential single voice stories implies excessive reliance on the hypothetical-deductive approach (akin to simplistic cause-effect pairings). The answer is not to develop Harvard case studies but to rewrite stories as polyvocal tapestries enabling different perceptions and interpretations to exist, thereby explaining webs of actions and interactions. What is new in Boje's approach is the **antenarrative** reading. His work thus shows ways in which narrative analysis methods can be supplemented by antenarrative methods, allowing previously fragmented and personal storytelling to be interpreted as a unified whole. The focus introduced by Boje (2001) offers alternative discourse analysis strategies that can be applied where qualitative story analyses can help to assess subjective, yet 'insightful' knowledge in order to obtain true understanding of complex interactions.

Conclusion

With the benefit of hindsight it is possible to re-construct a systematic pattern of events that have led to a failure. The narrated structure provides an explanation as to how and why failures occur.

This paper focused on the qualitative research methods available in the domain of software failures. Failures are often dynamic and confusing; requiring a holistic approach to resolution. Case histories are a special instance of a case study looking at the factors involved in failures in context, and at the dynamic interrelationships between them. Narrative methods (and antenarrative reading) provide an additional facet

for addressing the fragmented nature of failure stories. Combining case histories with narrative descriptions will lead to clearer failure stories that can account for contradictions and misunderstandings. It is hoped that by developing our understanding of methods that help in capturing and structuring histories and in telling stories we will also improve our ability to learn from such experiences. Indeed, the methods discussed in this paper form the front-end required for understanding and capturing knowledge in action (which could be supplemented by more formal methods to model their impact through a process of triangulation).

As for the future, good stories can also benefit from pictures. Once we have mastered the techniques of telling complex, modern stories, we need to focus on composing that information. Even the most gripping story needs to be made attractive and believable. Developing improved techniques for visualising knowledge (such as Net maps) can help in untangling some of the fragmented strands as well as in making the stories more readable and understandable, as well as more appealing.

References

Adelman C, Jenkins D. & Kemmis S. (1977) Rethinking Case Study: Notes from the Second Cambridge Conference, Cambridge Journal of Education, 6, 139-150.

Ashby W R (1960) *Design for a Brain* 2 ed. London: Chapman and Hall.

Bell J (1999) Doing Your Research Project, A Guide for First-time Researchers in Education and Social Science 3 ed. Buckingham, Open University Press.

Benbasat I, Goldstein D K & Mead M (1987) The Case Research Strategy in Studies of Information Systems, MIS Quarterly, 11(3), pp. 369-386.

Boje D M (2001) Narrative Methods for Organisational & Communication Research. London: Sage.

Boulding K E (1956) "General Systems Theory-The Skeleton of Science", *Management Science*, 2, p. 197.

Campbell J. (1982) *Grammatical Man: Information, Entropy, Language, and Life*. New York: Simon & Schuster.

Checkland P & Holwell S (1998) Information, Systems and Information systems – Making Sense of the Field. Chichester, Wiley.

Cornford T. & Smithson S. (1996) Project Research in Information Systems: A Student's Guide. Basingstoke, Macmillan.

Dalcher D (1994) Falling down is part of Growing up; the Study of Failure and the Software Engineering Community, *Proceedings of 7th SEI Education in Software Engineering Conference*, New York: Springer-verlag, pp. 489-496.

Dalcher D (1997) The study of Failure and Software Engineering Research. *Proceeding of the UK Software Engineering Association Easter Workshop,* Imperial College, April 1997, pp. 14-19.

Dalcher D (2000) Feebdack, Planning and Control – A Dynamic Relationship,
FEAST 2000, Imperial College, London, July 2000, pp. 34-38.

Dalcher D (2002) Safety, Risk and Danger: A New Dynamic Perspective. Cutter IT Journal, 2002. **15**(2): p. 23-27.

Easterby-Smith M, Thorpe M & Lowe Andy (2002) Management Research 2 ed. London: Sage.

Jankowicz A D (2000) Business Research Projects. 3 ed. London: Business Press.

Kaplan B. & Maxwell, J.A. (1994) "Qualitative Research Methods for Evaluating Computer Information Systems," in *Evaluating Health Care Information Systems: Methods and Applications*, J.G. Anderson, C.E. Aydin and S.J. Jay (eds.), Thousand Oaks, CA: Sage. pp. 45-68.

Lissack M and Roos J. (1999) *The Next Common Sense: Mastering Corporate Complexity Through Coherence*. London: Nicholas Brealey.

Lowrance W W (1976) *Of Acceptable Risk: Science and the Determionation of Safety*, Los Altos, CA: William Kaufmann.

Lyytinen K & Hirschheim R. (1987) Information Systems Failures: A Survey and Classification of the Empirical Literature, Oxford Surveys in Information Technology, Vol. 4, 257-309.

March J G (1994) *A Primer on Decision Making*. New York: Free Press.

March J G (1997) "Understanding How Decisions Happen in Organisations", in *Organisational Decision Making*, Shapira Z. (Ed.). Cambridge: Cambridge University Press, pp. 9-34.

Miles M B & Huberman A M (1994) Qualitative Data Analysis: An Expanded Sourcebook. Thousand Oaks, Ca: Sage

Nadler G (1985) "Systems Methodology and Design", *IEEE Transactions on Systems, Man, and Cybernetics*, 15(6), Nov./Dec. 1985, pp. 685-697.

Perrow C (1984) Normal Accidents, Living with High-Risk Technologies. New York: Basic Books.

Remenyi et al. (1998) Doing Research in Business and Management: An Introduction to Process and Method. London: Sage

Schramm W (1971) Notes on Case Studies of Instructional Media Projects, Working paper for the Academy for Educational Development. Washington, DC.

Thietart R-A et al (2001) Doing Management Research: A Comprehensive Guide, London: Sage.

Walsham G (1993) Interpreting Information Systems in Organizations. Chichester, Wiley.

Walsham G. (1995) Interpretive Case Studies in IS Research: Nature and Method, European Journal of Infomration Systems, 4(2), pp. 74-81.

Weick K E (1995) *Sensemaking in Organisations*. Thousand Oaks, CA: Sage Publications.

Yin R K (1989) Case Study Research: Design and Methods. Newbury Park, Ca: Sage.

Yin R K (1993) Application of Case Study Research – Design and Methods. Newbury Park, Ca: Sage.

Not reporting successful recoveries from self-made errors ? An empirical study in the chemical process industry

Tjerk van der Schaaf & Lisette Kanse

Eindhoven University of Technology, Safety Management Group,
PO box 513, Pav. U-8, 5600 MB Eindhoven, the Netherlands.
t.w.v.d.schaaf@tm.tue.nl

Introduction

Incident Reporting Schemes have a long history as part of an organization's overall safety management structure, especially in sectors like civil aviation, the chemical process industry, and, more recently, in rail transport and in a few healthcare domains such as anesthesiology, pharmacy, and transfusion medicine.

Their vulnerability in terms of the quantity and quality of the incident reports have led to guidelines for designing and implementing such schemes. Reason (1997, p 197) lists five important factors to "engineering a reporting culture": Indemnity against disciplinary proceedings; confidentiality or de-identification; separating the agency who collects and analyses the reports from the regulatory authority; rapid, useful, accessible and intelligible feedback to the reporting community; and finally, the ease of making the report. Similarly, Lucas (1991) identifies five organisational factors: the nature of the information collected (simply descriptive, or also causal); the use of information in the database (feedback, statistics, and error reduction strategies); analyst aids to collect and analyse the data; the nature of the organisation of the scheme (centralised or local, mandatory or voluntary). She also stresses the importance of the organisation's model of why humans make errors, as part of its overall safety culture.

These are just two examples of the well-documented "organisational design perspective" on reporting schemes. Much less is known about the "individual reporter's perspective":

when and why is one inclined to report a work-related incident to a formal scheme, and if so, what aspects exactly is one able and willing to then contribute? The starting point for the investigation described in this paper was the observation that during a re-analysis of part of a large database of voluntarily reported incidents at a chemical process plant in the Netherlands, we hardly encountered any report of self-made errors (Kanse, van der Schaaf & Rutte (2002)). This was surprising as this particular plant had been highly successful in establishing a reporting culture, where, apart from small damages, and dangerous situations, also large numbers of "near-misses" (i.e. initial errors and their subsequent successful recoveries) were freely reported. Not only the plant employees themselves, but also those temporarily stationed there by contractors, equally contributed to this "Near-Miss System", which had been operational for some 7 years by then, and was regarded to be a "safe" system in terms of guaranteed freedom from punishment.

As a result, this plant could perhaps even be labelled a "High Reliability Organisation" (Roberts & Bea, 2001). Even more puzzling was the fact that these references to self-made errors were also absent in the particular subset of the database we were looking at: successfully recovered (initial) errors and mistakes, which were thus completely inconsequential. Our question therefore was: what are the reasons on the part of the plant operators for not reporting successful recoveries from self-made errors at this plant?

In this paper we will first briefly summarise the (small amount of) literature on possible reasons for failing to report incidents in general, and evaluate it's relevance for our particular research question. After generating a relevant set of possible reasons, we describe a special diary-study where plant operators were asked to report their recovery of self-made errors under strictly confidential conditions, outside of the normal "Near Miss System". We will conclude by discussing the implications of the results, for the plant, but also in more general terms.

Reasons for not reporting

We have grouped the factors influencing incident reporting from the perspective of individual employees, that were mentioned in the literature, into 4 groups:

- FEAR (as a result of a "blame culture") of disciplinary action;
- USELESS (perceived attitudes of management taking no notice, not likely to do anything about it);
- RISK ACCEPTANCE (incidents are part of the job, cannot be prevented; or the "macho" perspective of "it won't happen to me");

- PRACTICAL REASONS (too time-consuming; too difficult).

Adams & Hartwell (1977) mention the blame culture (as does Webb et al., 1989) and the more practical reasons of time and effort (as does Glendon,1991). Beale et al. (1994) concludes that the perceived attitudes of management greatly influence reporting levels (see also Lucas, 1991), and also that certain levels of incidents are accepted as the norm. Similarly, Powell et al. (1971) finds that incidents may be seen as "part of the job" and cannot be prevented. This last point is supported by Cox & Cox (1991), who also put forward a belief in personal immunity ("accidents won't happen to me"; see also the "macho" culture in construction found by Glendon, 1991).

Probably the most comprehensive study so far was undertaken by Sharon Clarke (1998) with train drivers. She asked them to indicate their likelihood to report each of a standard set of 12 realistic incidents (a mix of dangerous situations, equipment failures, and other's errors). Also the drivers were offered a predefined set of 6 possible reasons for not reporting in each case (tell a colleague to report it; part of the job; avoid getting someone else in trouble; nothing would get done about this type of incident; too much paperwork; managers would take no notice).

In what way can the above results be useful in generating a relevant set of possible reasons not to report recoveries from self-made errors? Taking the four groups of reasons reported in the literature (fear; useless; risk acceptance; practical reason) as a starting point, we discussed this with three sections of the chemical plant's employees: management, safety department staff, and operators. Their opinions on possible reasons for non-reporting converged as follows.

The chemical plant operators, being part of an HRO (or at least something close to that) were seen to be highly unlikely to put forward some of the reasons mentioned earlier; the idea of accepting incidents as part of their job, unavoidable, and not happening to them.

Also the concept of the plant's management systematically ignoring reported risks, making the whole idea of coming forward with such information useless, was not considered realistic.

However, in a somewhat softened version most thought it could still be possible that operators would be afraid or ashamed to report their own initial errors and mistakes triggering the necessity for subsequent recovery actions.

Also they could consider it of less importance to report incidents that would be indicative of risks that they considered "no news", as they would be widely known amongst colleagues, minimizing their learning potential. It was further proposed that some types of incidents could be regarded as not applicable for the aims of the reporting scheme. That the fact that they themselves, through their successful recovery, "took care" of it, would make it superfluous to report. The fact that there were no real remaining consequences in the end could possibly turn it into something unimportant. Finally, the time consumption aspect ("always busy") could of course play a role, as could other practical reasons (i.e. not yet fully aware of the system).

Integrating all of the above considerations we proposed the following 6 possible reasons for not reporting recoveries from self-made errors:

- AFRAID/ASHAMED
- NO LEARNING
- NOT APPLICABLE
- RECOVERY
- NO REMAINING CONSEQUENCES
- OTHER

The Diary Study

*Methods:*Following the methods of previous human error researchers who used personal diaries to get reports of everyday errors (Reason & Mycielska, 1982; Reason & Lucas, 1984; and especially Sellen, 1994) we asked a total of 21 operators (all members of one of the five shifts) from the same chemical plant if they would cooperate. For a period of 15 working days (5 afternoon shifts, 5 night shifts, and 5 morning shifts) they would fill out a small form for every case of a recovery after a self-made error, which contained items such as: describe the self-made error(s); what were the potential consequences; who discovered the error(s); what recovery action(s) were then taken; any remaining actual consequences; and finally: "Would you have reported such an incident to the existing Near Miss System, and if not: why?". For this vital last question we did not offer any of the preselected possible reasons as options, as we wanted to leave the operators as free as possible to express themselves in this respect.

Results: In the period of the diary study the 21 operators completed forms relating to 33 recoveries from self-made errors. In only 3 cases they indicated that this incident would also have been reported to the "normal" Near Miss System, while for 5 of the remaining cases no reason(s) for not reporting were listed. Thus 25 cases remained.
The literally transcribed answers of the operators to the last question were then given to two independent coders; one of them being one of the authors, the other one another human factors expert (not being the other author) with ample experience in human error analysis. Firstly, each of the coders identified the separate reasons from the transcripts: one coder identified 32 reasons in the 25 cases, while the other found two additional reasons. They then reached consensus on 32 identifiable reasons. Secondly, the two coders independently were able to classify each of these 32 reasons into one of the 6 categories. They agreed on 28 of the 32 reasons and easily reached consensus on those reasons coded differently. A typical example of each of the statements and the resulting code are shown in Table 1. The overall results are shown below in Figure 1.

Code assigned	Example from transcript
No learning	The unclear/ confusing situation is already known
Not applicable	System is not meant for reporting this kind of event
Recovery	Because I made and recovered the mistake myself
No remaining consequence	Mistake had no consequence
Other	Not reported at the time: too busy then

Table 1: Examples of coded transcripts

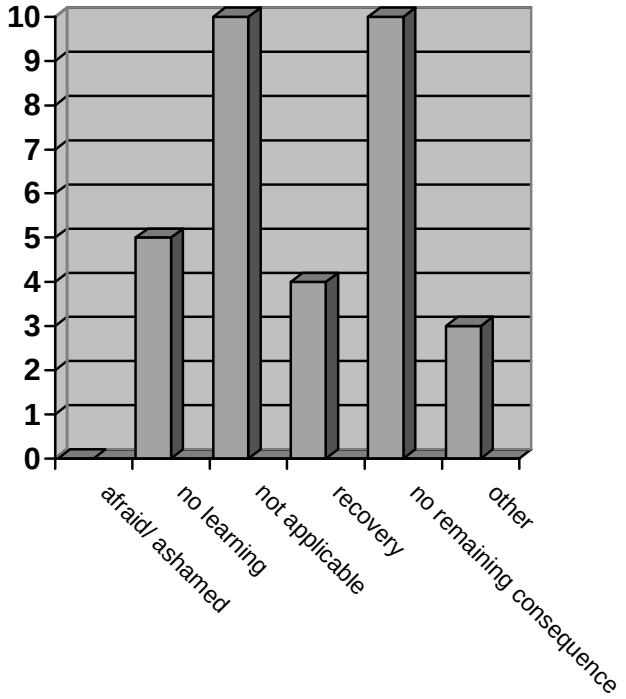


Figure 1: Distribution of 32 reasons given by 21 operators for not reporting 25 'diary incidents' to the existing near-miss reporting system

In addition to the results shown above, it is also worth mentioning that, as one of the items describing the details of the recovery process, operators on average judged the *potential* consequences of the incidents in the diary study (had they not been recovered from) at the same level of seriousness as the incidents normally reported to the existing Near Miss System (Kanse et al.,2002).

Discussion and Conclusions

In terms of the trustworthiness of the results, the diary study was a success in complementing/checking the existing Near Miss database: respondents were very open and frank with the author who collected these data from them, which otherwise they would not have shared with the plant's management and safety staff. They also were quite clear in describing their reasons for not reporting. The fact that the second, totally independent, rater had no problem at all in using this taxonomy of reasons seems to indicate it's potential usefulness in the future.

Looking at the results in Figure 1, the plant's management and safety staff were somewhat surprised: some of them had suspected still some level of fear or shame to report one's own errors, and/or a lower level of perceived *potential* consequences as the major reasons for not reporting successful recoveries. Rather, the results show a genuine difference (between operators and management) in *perceived importance*, as measured by the options of No Learning, Not Applicable, Recovery, and No remaining Consequences. It's up to the plant now to set up a program to clearly communicate their sincere interest in learning of the personal and system factors that make such successful recoveries possible, in stead of an attitude "all is well that ends well", which according to Kanse et al.(2002) is not compatible with an HRO. On the other hand the complete lack of mentioning being afraid or ashamed to report one's own errors may be seen as a very positive indicator of the plant's safety culture.

The success of this limited (in time and resources) diary study would suggest to repeat this procedure sometime after the implementation of a program to convince operators of the importance to report (especially!) successful recoveries: both to measure the impact on operator perception, and to monitor from time to time for other, possibly new, reasons for not reporting.

References

Adams, N.L. and Hartwell, N. M.(1977) Accident reporting systems: a basic problem area in industrial society. Journal of Occupational Psychology, 50, 285-298

Beale, D., Leather, P. & Cox, T. (1994) The role of the reporting of violent incidents in tackling workplace violence. Proceedings of the Fourth Conference on Safety and Well-Being, Loughborough, November (Loughborough) pp 138 – 151

Clarke, S (1998) Safety culture on the UK railway network. Work and Stress Vol 12 No.1, 6 – 16

Cox, S. & Cox, T. (1991) The structure of employee attitudes to safety: a European example. Work and Stress, 5 (2), 93 - 106

Glendon, A. I. (1991) Accident data analysis. Journal of Health and Safety, 7, 5 - 24

Kanse, L., van der Schaaf, T.W & Rutte, C. G. (2002 - submitted) A failure has occurred. Now what?

Lucas, D. A.(1991) Organisational aspects of near miss reporting. *In* Van der Schaaf, T. W., Lucas, D. A. and Hale, A. R. (eds) Near miss reporting as a safety tool. Oxford: Butterworth-Heinemann Ltd

Powell, P. I., Hale, M., Martin, J. & Simon M. (1971) 2000 accidents: A shop floor study of their causes. Report no. 21 (London: National Institute of Industrial Psychology)

Reason, J. (1997) Managing the risk of organisational accidents. Ashgate Publishing Limited, Hampshire, England.

Reason, J. & Lucas, D. (1984) Using cognitive diaries to investigate naturally occurring memory blocks. *In:* J.E. Harris & P. E. Morris (eds) Everyday memory, actions and absent-mindedness. pp. 53 – 70 London: Academic Press

Reason, J. & Mycielska, K. (1982) Absent-minded? The psychology of mental lapses and everyday errors. Englewood Cliffs, NJ: Prentice Hall Inc.

Roberts, K. H. & Bea, R. G. (2001) Must accidents happen? Lessons from high-reliability organisations. Academy of Management Executive, 15 (3), 70 – 77

Sellen, A. J. (1994) Detection of everyday errors. Applied Psychology: An International Review, 43 (4), 475 - 498

Webb, G.R., Redmand, S., Wilkinson, C. and Sanson-Fisher, R.W. (1989) Filtering effects in reporting work injuries. Accident Analysis and Prevention, 21, 115-123.

A framework for re-examining accident reports to support interaction design processes

Anne Bruseberg, Iya Solodilova, Rachid Hourizi, and Peter Johnson,
Department of Computer Science, University of Bath, Bath, BA2 7AY, UK.
http://www.cs.bath.ac.uk/~flightdk.

Abstract: This paper discusses the requirements for informing interaction designers through accident data. It examines the drawbacks of utilising accident reports, as well as formal analysis techniques to identify design opportunities. It suggests a method for re-examining existing accident reports through a systematic, but informal analysis framework that can support design processes more effectively. It argues that, in order to benefit designers, accident re-examination methods should not only provide a deeper understanding of the causal structures and their interpretations, but should also aid both the process of the analysis and the communication of the results. It is essential that they consider the variety of relationships between contributing factors, and take the dynamics of human control activities into account.

Keywords: design process, accident analysis methods, aviation accidents, requirements specification.

Introduction

Information from accident reports is vital to specify design requirements. Producing a deeper understanding of the ways in which humans and systems have failed to collaborate in the past is one of the main sources of insight for designing safer systems. However, designers find it often difficult to access such information due to both unsuitable content and format of the information at hand. Accident reports are not primarily intended to be used directly by designers, or to support design decision-making processes. However, they contain valuable details that designers should consider when approaching their design tasks.

Accident reports aim to establish a comprehensive account of the causes of an accident, and give a summarised list of recommendations, aimed at a range of different organisations. They aim at generating and justifying a set of necessary measures that are required by the regulator. Due to the expectation that accident reports should produce a concise list of causal factors and recommendations, there is a danger of oversimplifying the conclusions. This has the effect of limiting the design potential (Leveson, 2001). The list of recommendations given may restrict designers, since it does not aim at the wealth of potential design opportunities, but a minimum set of requirements (often in a legal context).

Moreover, accident reports are documents that contain a vast amount of information. Design processes are always time-constrained. The structure of accident report documents, and the presentation of the results in a lengthy textual format, not only requires excessive time and effort to study, but can also obscure the line of arguments (Johnson, 1998; Snowdon and Johnson, 1998). Their scope is too wide to be useful for specialists such as interaction designers. Hence, additional analysis methods for re-examining such sources of information are needed to make the information accessible and usable by designers.

It needs to be noted here that the term design can have very different meanings – ranging from organisational design, training design, design of regulations, to interaction design. None of these fields are specifically addressed in accident reports; instead they aim at all of them to an extent. Design opportunities may arise for all of these areas in reaction to an unwanted set of events. Often, different defences interact closely with each other, and have to be planned in conjunction. The method described in this paper mainly aimed to support interaction designers. However, it is not necessarily limited to those. Likewise, interaction designers may be perceived here in a fairly broad sense, ranging from HCI specialists to user interface designers.

This paper aims to clarify the needs for an accident re-examination tool, to be useful for informing design processes. It focuses on the aviation domain. The analysis process is mainly concerned with problems relating to the interaction of human operators with an interface (e.g. communication failure), rather than technical failures as such (e.g. engine fire). It evaluates existing failure analysis techniques regarding these requirements. It then presents a framework for re-examining accident data and demonstrates its application and usefulness using the Cali air accident (Aeronautica Civil, 1996; Kaiser, 1996; Simmon, 1998).

Johnson (1999) showed that CAE (Conclusions, Analysis, and Evidence) diagrams, which track the reasoning behind an accident report in a graphical format, can be linked to QOC (Questions, Options and Criteria) diagrams, which specify and evaluate design alternatives for a given requirement and are a

standard designing technique. Beyond this, the approach presented in this paper aims to identify requirements as part of the accident examination process. Its main objective is to generate a wide list of design opportunities, which may then be evaluated regarding their feasibility and effectiveness.
The technique presented in this paper is an approach that was developed in response to having re-examined the Cali accident in a systematic, but informal way. The process of the analysis, and the experiences of different experts gained in approaching the task, has led to a number of insights regarding the nature and requirements of the analysis.

The requirements for accident re-examination tools
The need to gain a deeper understanding through re-examination: To understand the characteristics of design processes, it is useful to draw on insights from the domain of Industrial Design. Cross (2000) shows that design processes feed off a continuous widening and narrowing of the problem space, thus supporting both inspiration (e.g. through brainstorming, where all critique is suspended) as well as selection of feasible solution paths (e.g. through weighing of design criteria and evaluating intermediate design concepts). Requirements specification tools should therefore not just aim at a minimum set of requirements, but should, at the earliest stage of informing the designing process, enable the generation of a wide range of solution opportunities.
There is a danger in narrowing down design opportunities too early by overly concentrating on the identification of human error as part of the accident causes. Defining a human action as an error bears an element of assigning blame to a person (Leveson, 2001), based on a better, retrospective understanding of the situation in question, and by evaluating an action against an ideal procedure. It is important to look beyond the 'error' made, by placing it in its context, and specifying other factors that led to the choice of a particular erroneous action. Likewise, errors have to be understood in terms of complex error paths, rather than single errors. For example, erroneous actions are embedded into a cognitive activity. Since actions are usually preceded by thinking processes, the (erroneous) beliefs that were held by operators at the time are particularly crucial for understanding the causes of the accident. Thus, a deep understanding beyond the causal chain of events has to be established. In this way, the analyst keeps an open mind to trace 'design errors' as well as 'operator errors'.
A tool for re-examination of accident report data relies on the information given about the accident sequence, as well as many of the interpretations provided by aviation experts. The aim of a re-examination is not to question the results of accident reports, but to

- gain deeper insight into unexplored aspects, particularly those relevant to design opportunities;
- make the reasoning behind the conclusions more effectively traceable;
- double-check the line of reasoning of the report;
- highlight additional contributory factors that may have been missed, and thus opportunities for defences against future accidents.

The need for a communication tool: Designing requires creativity, which is not possible without a deep immersion into the problem space (Cross, 2000). A brief list of requirements is unsuitable to support creative work. Designers benefit from understanding the analysis process behind a design requirement, thus exploring the context of the design task. Requirements can be defined at many different levels of detail, ranging from a fairly broad need (e.g. design a safe system) to a specific request (e.g. ensure consistent colour coding). Often, detailed requirements are sub-sets of more general one's, having examined the problem space, and moving gradually from problems to solutions. Understanding requirements and solving design problems is an iterative and continuous process, where the expertise of different experts overlaps. Hence, if given a requirements specification based on an accident report, designers need to be able to trace the reasoning behind the requests, to be able to interpret the conclusions, and to 'tune' their own thinking process into that of other experts. Thus, by making the reasoning process clearly visible, requirements specifications can open opportunities for design solutions.
Accident reports often make it difficult to trace the underlying reasoning process that lead to the final list of conclusions. The evidence may be distributed across different pages, obscured by additional detail or alternative accounts, or hidden behind the rhetoric of the account given (Johnson, 1998; Snowdon and Johnson, 1998). To make the information more usable, it is necessary to present it in a way that helps effectively trace the evidence for the conclusions. Suitable presentation is valuable, both to track the progress of one's own understanding, and to communicate the results to others. Thus, the format of the presentation is important both during the analysis process, as well as for its final output. Being able to trace

the line of reasoning is crucial for communication between professionals with different sets of expertise (e.g. accident analysts, HCI specialists, interface designers).

The need for supporting the process of the analysis: Understanding accidents often requires piecing together information from various sources. There may be accounts from different experts, as well as different independent accident reports. Analysts require a vehicle to deal with the large amount of (possibly conflicting) insight. A good re-examination method should therefore be able to structure large amounts of data, to support the conceptualisation of the causal sequence and its design implications. The merit of such a tool should mainly lie in the ease of supporting the process of the analysis.

When examining the causes of accidents, complex networks of events have to be understood and interpreted. There are many different *processes* that interact closely during the course of the accident, and have to be understood in conjunction. In aviation accidents, the aspects to be taken into consideration include processes such as

- the physical laws of the plane's actions, in relation to its surroundings (speed, aerodynamics, wind, presence of terrain etc.);
- the goals/priorities of the pilots (e.g. getting to the destination in time, landing the plane safely, following regulations);
- external conditions, and their influence on pilots' activities (e.g. visibility, restricting pilots' perception; plane delay, influencing pilots' goals);
- the thinking processes of the pilots at decision making times, including the understanding of the situation, and plans made;
- the actions taken and their influence on the events that happened subsequently.

Beyond this, there are different levels of abstraction at which to understand causes, and there are many different ways in which causes can relate to each other. It is important to note the different types of *relationships* that may exist between causal factors: Relationships may be determined by time – in terms of the sequence of events, as well as the implications of its restrictions regarding aspects such as workload, mental capacity, or opportunities to communicate. A particular event may not have happened without a particular combination of others, thus specifying AND/OR conditions, as captured, for example, through Fault Tree Analysis. The causal factors recorded in an accident analysis represent an analyst's understanding. The representation produced may have filtered the facts to represent important aspects only. It may represent the mental breakdown of the analyst's thinking at different levels of abstraction, which can be expressed through hierarchical links. For example, to conclude that a situation of high workload has occurred (1), the analyst has realised aspects such as the lack of time (1.1), the number of tasks (1.2), and the complexity of operations (1.3).

When not all events and their effects are known, representations may capture different alternatives. Likewise, when tracing latent causes beyond the accident sequence, connections become speculative since it may be difficult to know whether a particular set of defences would really have had the desired impact on the accident sequence. Thus, such links have, again, a different character.

Analysing accidents involves both (1) establishing a sequence of events (including examination of relationships between processes), and (2) interpreting them at different levels of abstraction. For example, a course of events is evaluated against hypothetical ideal paths of actions and events in order to identify actions as errors. Key actions may change the path of events. These actions can, in retrospect, be identified as 'errors'. Hence, understanding the implications of events to identify defences involves comparing different 'lines of reality' that may have happened if events had taken a different course. Likewise, pilots mentally simulate different possible paths of reality when making predictions and planning a set of actions. It is important to identify these processes to facilitate them through designs. Identifying these different levels of understanding is a complex process that is difficult to capture through a single, formal analysis. It requires different stages of reasoning about the accident.

The need to aid the interpretation process: Livingston et al (2001) provide a comprehensive literature review of techniques available to identify 'root causes' for accidents, and outline the standard stages of the accident/incident analysis process:

Firstly, a basic understanding of what happened during the accident/incident is needed. This may include considering alternatives if the facts are not fully known. Standard techniques to establish the sequence of events and contributing conditions include ECF (events and causal factors) analysis, MES (multilinear events sequencing), and STEP (sequentially timed events plotting). All of these involve constructing sequence diagrams, most of which include a notion of time, and some consider different actors.

Secondly, critical events and conditions need to be established, to reduce the number of issues that need to be considered. By examining what-if scenarios, and through counterfactual reasoning, the most crucial events and conditions that lead to the unwanted course of events can be established. This can be facilitated through Fault Tree Analysis, Petri Nets, Barrier Analysis, and/or Change Analysis, thus highlighting effective defences.

Thirdly, the analysis need to be taken further back in time beyond the accident sequence, to identify management and organisational factors that may prevent future accidents/incidents through suitable barriers or defences. This includes both the evaluation of defensive measures that had been in place but were not activated, and considering new defences. One of the most established techniques for this purpose is MORT (Management Oversight and Risk Tree), which scrutinizes all events for latent errors, using a structured checklist, and drawing partly on Barrier Analysis.

The concept of 'root causes' suggests that there are fairly linear connections between unwanted events and 'latent errors' (i.e. design or management oversights). However, in complex domains such as aviation, such conclusions cannot easily be drawn without a detailed examination of the different types of relationships between the contributing accident factors. For example, most of the above mentioned techniques rely heavily on examining events and conditions only, and fail to consider the complex interactions between human thinking processes and technical systems (e.g. automated interface). Moreover, identifying 'root causes' introduces a focus on narrowing the problem space to the 'most important' contributing factors, rather than supporting brainstorming activities for new design solutions based on a deep understanding of the accident mechanisms. In contrast, the approach presented here does not aim to reduce the problem space, but to structure it better, so that it can be examined in more depth. Since accidents are unlikely to repeat themselves in exactly the same way, the relative importance of different unwanted events is of less significance than understanding why human operators/controllers were not able to cope with problematic situations at the time. Thus, the focus is on identifying design opportunities rather than design necessities.

The most challenging part of accident examination is the interpretation of the accident sequence, particularly for accidents with complex interrelationships between the different factors, since these are very difficult to capture through formal techniques such as such as Cause-Consequence Diagrams (Nielsen, 1975), Petri Nets (Johnson, 1998), Fault Tree Analysis (for example, see Love and Johnson,1997), or Why-Because Analysis (Ladkin and Loer, 1999). Instead, informal techniques are more suited to provide aids in structuring the facts, and examining the context of problems, so that they can be evaluated more easily. Leveson (2001), for example, presents an informal technique to examine accident data using different hierarchical levels (i.e. systemic factors, conditions, events/accident mechanism), leading from the basic causal relationships to an interpretation of the implications. However, the approach does not provide a suitable format to support the communication of the results, since it presents the findings in a lengthy text format only.

When re-examining accident reports, it is vital to comprehend the conclusions made by tracking the line of reasoning. For example, CAE (Conclusion, Analysis, Evidence) diagrams (Johnson, 1999) are a useful tool to examine the reasoning behind the rhetoric of accident reports, and make it graphically visible for further examination. However, a detailed understanding of how the events unfolded over time is also necessary, to be able to follow the conclusions drawn. Moreover, CAE diagrams do not scrutinise the different factors and cognitive activities of the pilots that led to the accident – to re-evaluate and extend the conclusions made.

The drawbacks of Why-Because Analysis as a formal analysis approach: Why-Because Analysis (WBA) was developed specifically for complex, open, heterogeneous systems, such as the aviation domain, where the system behaviour is highly affected by its environment (Ladkin and Loer, 1999). WBA is a formal and rigorous technique that includes formal proof methods to verify the correctness of the causal explanations. However, formal analysis techniques such as WBA require a clear specification of the constructs they produce, since ambiguities leave room for misunderstandings. For example, the level of abstraction at which accident factors are defined depends on the understanding of the analyst, and can be chosen at different levels of generalization. An item such as 'crew used procedural shortcuts' could have been specified at much greater detail – however the analysts (Gerdsmeier, Ladkin, and Loer, 1997) chose not to, possibly for reasons of clarity, or implied understanding, or lack of deeper insight at the time of analysis. This suggests that the specification of the causal factors during a formal analysis is often chosen quite informally. Likewise, having to clearly specify links (e.g. for exact graphical representations) implies a detailed understanding of their nature (e.g. sequential, hierarchical, logical), which can only be developed

gradually, as understanding progresses. Moreover, WBA draws on an intuitive division of factors into states (defined as persisting conditions) and events (seen as synonymous with actions). The classification has implications for the logical proof process of the analysis (Gerdsmeier, Ladkin, and Loer, 1997). In a formal analysis, the need to define a factor as either a state or an event makes the analysis process very complex, since these definitions are not free of ambiguities.

Making such distinctions has mainly the function of providing thinking aids for the analyst (e.g. the definition of a 'state' depends largely on the perception of the analyst). There are many different types of causal factors, such as external events and conditions, pilot's actions, errors (including failures to act), lack of knowledge and understanding, intermediate situations, or resulting problems. In an informal analysis, distinguishing their function can be useful in conceptualising their properties and relationships. Whilst WBA includes statements about the crew's cognitive states, and distinguishes between aspects such as action failures and perceptions failures, it does not support interpreting such findings.

Requirements for helping to structure abstract thoughts: Many formal techniques rely on graphical representations to progress understanding. Producing flow charts places a heavy emphasis on generating visual properties, rather than examining the content. Instead of being a vehicle for the analysis, it can become a hindering aspect. In this context, graphical representations may be more useful for the final presentation of results. Immediate drawbacks occur due to practical aspects (e.g. the delay in creating boxes and links, fitting all factors onto one page, avoiding crossing lines). Whilst drawing tools are useful for grouping different factors visually, thus starting to establish categories of related factors, they do not help to make sense of the mass of information from accident reports. It is very easy to get lost in the detail of specifying links without actually understanding their nature. They cannot deal easily with different levels of abstraction, since they are usually two-dimensional (or become overly complex). The differentiation between different types of data such as (1) events, states, conditions and actions as part of the accident sequence, and (2) latent errors that have happened in the past suggesting design changes, becomes very complex to handle. Hence, there is a tendency to copy the reasoning process of the accident report at a high level of abstraction, without re-considering the facts and interpretations.

Accidents unfold over time, through a number of events, relating to conditions and changes of states and processes. Many techniques, such as WBA and Fault Tree Analysis, use a hierarchy, or tree structure, by drawing on the fact that a (usually single) final unwanted event (e.g. impact with mountain) could be traced back to a variety of causes. However, a tree structure may not be the ideal form of representation for causal relationships, since different types of events, states, conditions and processes can relate to each other in many different ways, creating a complex, network-like, picture. Creating hierarchical representations helps to structure the analyst's understanding. However, they are fairly difficult to conceptualise, since they are often artificial. Instead, the analysis process may benefit from following a network approach, rather than a (linear) branching approach. It is much easier to first create a listing, and then group the aspects according to suitable criteria, before identifying different types of relationships between causal factors.

Developing and applying a new, informal approach to accident re-examination

Through studying the Cali aviation accident, an informal approach was developed, that aimed at gaining a detailed understanding of the accident mechanisms to identify design opportunities. Having re-examined the official accident report (Aeronautica Civil, 1996), it proved to be a very difficult document to gain a sound understanding of the accident causality from. However, other sources of information were available (Kaiser, 1996; Simmon, 1998). All these had to be brought together. The approach presented here includes the following stages, thus gradually increasing the level of interpretation:

1. Understanding the sequence of events;
2. Examination of key events/errors regarding their pre-conditions and implications, through a grouping and categorisation process;

Specification of design opportunities.

The Cali accident is an example of a 'controlled flight into terrain' (CFIT), where the accident cannot be attributed to an aircraft malfunction, but was induced through pilots' actions. These lead to a collision of the American Airlines Boeing 757-223 with a mountain, whilst attempting to approach the Cali Airport (Columbia) in darkness. The pilots had to deal with a serious flight delay, and developed a fixation on making up time. The already high workload of the approach phase was worsened after the acceptance of a late runway change. Miscommunications with Air Traffic Control, in combination with erroneous inputs into the FMS (flight management system), took the aircraft off course into mountainous terrain. The pilots' inability to notice the problem, and their subsequent failure to regain situational awareness and abandon

descent, led to a choice of route where a mountain lay in the way of the flight path. After the sounding of the Ground Proximity Warning System, the escape manoeuvre was inefficient, leading to a collision with the mountain. Of the 163 people on board, 4 passengers survived (Aeronautica Civil, 1996; Kaiser, 1996; Simmon, 1998).

The experience in trying to understand the Cali accident from accident reports has shown that it is impossible to really understand the causality without first developing a good understanding of how the events unfolded over time. Many techniques, including WBA, choose to establish event sequences by tracing the events backwards in time from the critical accident point. Working backwards in time, however, is quite difficult. Constructing a (fairly traditional) timeline proved to be the most straightforward approach.

Table 1 shows an example timeline, based mainly on the voice recorder transcript available. Timelines do not necessarily require a graphical representation of the length of time periods to be useful. A list of events over time represents the course of actions and also establishes insight into timely proximity. Creating a timeline involves a process of selecting the most crucial events and statements, thus starting the process of interpretation. The level of abstraction chosen depends on the needs of the analyst. It might vary between the use of 'raw' information (e.g. the basic course of actions, quotes from the tape transcript) and conclusions summarising facts (e.g. description of pilot's action or judgment). It is useful to distinguish at least three columns to trace the activities of the two pilots and provide a space for other agents (e.g. different Air Traffic Control stations, aircraft's actions). Separating the creation of a timeline from other stages in the analysis makes the process of interpretation less complex.

Table 1: Extract of a timeline generated for the Cali accident.

	Captain – the pilot not flying	**First Officer – the pilot flying**	**Air Traffic Control (ATC)/ other**
21:37:03	brief confusion about whether to head for Tulua, Cali or Rozo next; indirect decision to ignore Tulua and head for ROZO		
21:37:29	request to ATC to go direct to ROZO; acknowledged need to report to Tulua		cleared direct to ROZO by ATC and requested report to Tulua 21miles and 5000 feet
21:37:43			aircraft begins to turn to Romeo (after change in FMS)
21:37:59	pilots agreed (assumed) that lower altitude had been given due to confirmed destination clearance		
21:38:01	announcement of having **inputted ROZO into FMS** (but wrongly selected Romeo)		
21:38:33	"let me put ULQ into here…want to be on raw data with you" *(ULQ is Tulua)*		
21:38:39			requested distance DME
21:38:49		**"uh, where are we"**	

After establishing the sequence of events, the analysis progresses to interpret them, through several levels. The first level of interpretation involves identifying which factors have changed the course of events in an undesirable way – and which were the most crucial one's. The next levels identify the effects of key actions and events, as well as their immediate causes and pre-conditions in terms of what happened – not in terms of what should have happened. The last level examines design opportunities, thus identifying desirable defences.

Table 2: Extract of the factor analysis table generated for the Cali analysis.

Major Factor (error)	**Errors (prior, subsequent, related)**	**Implications (requirements, states of knowledge)**	**Pre-conditions (influential factors)**	**Interaction design issues (any possible 'angles of attack')**
6. Continually proceeding to descent	Wrongly assuming that ATC had cleared further descent Failure to announce descent to ATC Failure to cancel descent (and initiate go-around) after loosing situational awareness	Aircraft was too low to stay clear of the mountain	Priority on 'pressing on' to land quickly – ignoring possibility of danger Human tendency to resolve a problem under stress, rather than cancel the task and alter initial decision False believe in radar cover of ATC Lack of awareness of terrain Belief that Tulua was still ahead and mountains further away Lack of appreciation for the terrain due to not having been familiar with the local conditions sufficiently Inability to appreciate location and proximity of terrain from FMS display	Concentrating on the task of locating the plane for too long made the pilots neglect the priority to stop descending – the interface should have made locating the plane easier Interface and abilities of automation lured the pilots into a false sense of trust (e.g. solid magenta line, presence of waypoints in database to recover direction) Interface should have communicated proximity of danger
7. Failure to understand position and proximity of mountains	Failure to familiarise with terrain information; use of approach chart (not local area chart) as primary reference (terrain shown as altitude dots only), crew did not use all navigational information available (e.g. written terrain information on flight plan, and on maps) Failure to recognise that Tulua had been passed already Failure to detect aircraft's deviation from path early enough Failure to regain situational awareness Failure to interpret signal from ULQ; failure to locate Tulua (belief that there is something wrong with Tulua locator) – fixation Decision to intercept extended centreline to Cali/Rozo without understanding relation to terrain	Lack of awareness of dangerous situation (failure to appreciate terrain information in relation to the flight path) Wrong belief of crew that Tulua had not been passed yet and the mountains where not as close Inability to identify significant turn away from course towards Romeo Crew newer recognised they had passed Tulua	Darkness – no visual terrain cues No availably of ATC radar coverage Frequent radio contacts Very limited time available to perform required tasks Crew was rushed, disorientated and confused Neglect to realise descent into unknown area Belief that interaction with FMS (i.e. locating Tulua, selecting CLO (Cali) to re-gain direction) would be enough – relying on it	Mountain information available on different displays, not clear on main one Accurate magenta line on approach chart may result in overconfidence in automation Change of course was not made obvious enough to busy crew Display did not enable pilots to locate plane under stressed, confused conditions Confusing navigational information displayed – FMS did not clarify the situation Signal from ULQ could not be interpreted having made wrong assumptions about location Automation gave false sense of security through facility to input waypoint (Cali) that changed direction towards desired location without validating terrain proximity

A suitable format for this analysis is a table (see Table 2 for an example). It examines the factors of the accident through a process of identifying related issues, grouping them, generating categories, and concluding design opportunities. This is an iterative process that does not follow easily a step-by-step guide. Major tasks include:

Accident factors such as events, conditions, crucial actions, identified errors, and resulting situations can be listed and related to each other through a process of grouping them according to their closeness to each other – both in terms of time and influence.

The grouping process identifies critical actions or processes that significantly changed the sequence of events to a different, unwanted course. These form major categories that structure the table horizontally, loosely following a timely order, but not primarily. They are listed in column one. In the case of the Cali accident, they were identified as:

(1) Decision to accept runway 19;
(2) Communication failures between ATC and captain;
(3) Deletion of intermediate waypoints in FMS (after inputting *direct* to Cali);
(4) Inadequate awareness of approach characteristics and current location;
(5) Erroneous input of Romeo instead of Rozo;
(6) Continually proceeding to descent;
(7) Failure to understand position and proximity of mountains;
(8) Unsuccessful avoidance manoeuvre.

For each category (the rows in the table), all associated (sub-)errors can be listed to provide details of what went wrong. These appear in column two.

Within each category, the errors (or failures to carry out an expected action) can be examined in terms of their immediate causes and effects, to understand how they relate to their context – by specifying their pre-conditions (e.g. previous actions, external conditions, crew beliefs) and implications (e.g. conditions, requirements for actions, states of knowledge). These are added within columns three and four.

All factors identified within each category (including errors, causes, and implications), and their relationships through context, can then be examined regarding opportunities for defences. This is adding a fifth column. This process introduces a reflection about both the influential factors that had been determined further back in time, as well as possible alternative outcomes, had defences been in place – thus introducing a new dimension to the investigation.

The process of structuring the information in this way helps to gradually narrow down the complexity of the analysis task, by creating smaller units of concern, and specifying the functions of factors within the overall causal structure. The categorisation processes serve as an essential aid in structuring the wealth of information, in supporting the interpretation of the results, and communicating the most important insights. Clustering the information around the most critical factors helps to understand the complexity of causes and implications. The table format makes the information accessible through structuring it, and provides an overview of the insights. It reduces the mass of information significantly without loss of important detail. Although the relationships between different factors are not explicitly specified, they become apparent easily through their positioning in the table, and the context within their category. If needed, graphical link techniques can be added to further scrutinise and visualise different types of interrelations.

Different analysts may have different perspectives on the same problem; therefore it is beneficial to separate the specification of design implications from establishing the facts. The analysis should be carried out with an open mind, since it can identify a range of different design opportunities, ranging from regulatory improvements (e.g. crew rest requirements, company procedures prioritising economics for safety issues) and training design needs (e.g. understanding error types to address skill levels), to interaction design requirements. Although this analysis focused mainly on interaction design issues, it may well serve the identification of other design aspects, since they often closely interrelate.

Having re-examined the analysis results of the Cali accident from WBA (Gerdsmeier, Ladkin, and Loer, 1997), several shortcomings became apparent regarding the links specified, thus suggesting that an informal technique, as outlined in this paper, is able to scrutinise complex problems effectively:

One of the links in the WBA graph suggests that ‘AC flying too low for cleared airspace’ was causal to ‘AC in mountainous terrain’. However, the aircraft flew over mountainous terrain *as well as* being too low, both factors leading to the impact in conjunction.

Another link drawn suggests that ‘lack of external visual reference’ was causal to ‘Crew high workload’. However, the fact that the crew could not see the terrain, lead to the fact that they realised the proximity of the mountain too late. Workload, however, was mainly caused by the lack of time, the pressures of the

landing phase as such, the additional number of tasks due to changing the approach, the higher than usual airspeed, and the number of radio communications.
Similarly, the link between 'FMS erases intermediate waypoints' and 'AC flying too low for cleared airspace' would require clarification, since erasing the waypoints mainly contributed to the confusion over the position of the aircraft in relation to the waypoint Tulua, thus indirectly relating to the neglect to ascent after getting lost, rather than the descent without permission.
The analysis presented here lead not just to a clearer understanding of the accident, and a lucid presentation of the main facts, but also generated additional insight. For example, it is not specifically stated in the accident report that the pilots never actually realised that they had passed Tulua already. Even though the last change of course was orienting the plane back towards the required route, the pilots really tried to achieve something else (i.e. go to the extended centreline of 'Rozo'), thus failing to realise the proximity of the mountains. This understanding is vital to appreciate the seriousness of the fact that the pilots were not able to

- recognise that they were off-course;
- regain situational awareness;
- appreciate terrain information;
- interact suitably with the automated devices.

Such information is essential to examine design options and needs, and to reflect about potential new interface solutions.

Conclusions

This paper demonstrates the value of systematic, but informal approaches to the re-examination of accident reports in order to specify design opportunities. To be thorough, a re-examination method should concentrate on aiding the process of the analysis, both in terms of the understanding of the events, as well as the different levels of possible interpretations. A suitable analysis approach should fulfil the following requirements:
It should enable a thorough re-examination of the conclusions to re-interpret them in order to aid design processes and to gain a deeper understanding of the underlying causal structures. It should also combine means for tracing back other analyst's thoughts and building up one's own understanding.
It has to support the structuring of large amounts of information. Analysis processes involve making choices as to what pieces of information are most relevant, and at which level of abstraction new insights can be handled and presented.
The presentation of results should aid both the process of conceptualisation, and the communication of the results; the format of communicating the results should provide interaction designers with detailed information that does not restrict their creative processes and enables readers to follow the line of reasoning behind conclusions.
The accident analysis method proposed in this paper is an informal framework that aims at structuring complex details from existing accident reports. It aids the understanding and interpretation of the accident by supporting a natural reasoning process. Through organisation of the data into categories reflecting critical changes to the course of events, and by specifying implications and pre-conditions within each category, the complexity of the analysis process can be reduced. Since different analysis techniques suit different researchers according to their reasoning preferences, the framework can be supported and extended flexibly through additional techniques, including more formal approaches, to deal with specific questions.
It has been demonstrated that the method was valuable for facilitating the development of a deep understanding of the Cali aviation accident. The analysis concentrated on identifying the scope of possible design approaches to find suitable defences, not the process of narrowing down design opportunities to a subset too early. The presentation of the findings allows a quick immersion into the facts.
The method has emerged through the process of re-examining the Cali accident. This development process generated a wealth of useful ideas. However, the framework was not available at the outset of the analysis process, and has not been applied to other accidents or incidents as yet. The application to additional analyses processes will enable further refinement of the techniques and concepts presented.

Acknowledgements

This work is being funded by the EPSRC (grant number GR/R40739/01) and supported by QinetiQ and Westland Helicopters.

References

Aeronautica_Civil_of_the_Republic_of_Colombia (1996) AA965 Cali Accident Report, Near Buga, Colombia, Dec 20, 1995: available from: http://www.rvs.uni-bielefeld.de/publications/Incidents/DOCS/ComAndRep/Cali/calirep.html.

Cross, N (2000) *Engineering Design Methods: Strategies for Product Design*. 2nd ed. London: Wiley.

Gerdsmeier, T, Ladkin, P, and Loer, K (1997) Analysing the Cali Accident With a WB-Graph. Proceedings of the *Human Error and Systems Development Workshop*, at Glasgow, March 1997, http://www.rvs.uni-bielefeld.de/publications/Reports/caliWB.html.

Johnson, C W (1998) Representing the Impact of Time on Human Error and Systems Failure. *Interacting with Computers* 11 (September): 53-86.

Johnson, C W (1999) Using CAE Diagrams to Visualise the Arguments in Accident Reports: Department of Computing Science, University of Glasgow, Glasgow, G12 8QQ, available from: http://www.dcs.gla.ac.uk/~johnson/papers/cae_99/.

Kaiser, J (1996) Flight 965, Accident Investigation Summary, APA Flightline – November 1996 – Special Report. available from http://www.alliedpilots.org/pub/flightline/nov-1996/flt-965.html.

Ladkin, P, and Loer, K (1999) Explaining Accidents Causally Using Why-Because Analysis (WBA). Proceedings of the *3rd Workshop on Human Error, Safety, and System Development (HESSD), 7th/8th June 1999*, at Liege, Belgium, available from: http://www-users.cs.york.ac.uk/~loer/.

Leveson, N (2001) Evaluating Accident Models Using Recent Aerospace Accidents: NASA Internal Study, available from http://sunnyday.mit.edu/accidents/.

Livingston, A D, Jackson, G, and Priestley, K (2001) Root causes analysis: Literature review. HSE contract research report 325/2001, available from: http://www.hse.gov.uk/research/crr_pdf/2001/crr01325.pdf.

Love, L, and Johnson, C W (1997) Using Diagrams to Support the Analysis of System 'Failure' and Operator 'Error. In *People and Computers XII: Proceedings of HCI'97*, edited by H. Thimbleby, B. O'Conaill and P. Thomas. London: Springer Verlag: 245-262.

Nielsen, D S (1975) Use of Cause-Consequence Charts in Practical Systems Analysis. *Reliability and Fault Tree Analysis, SIAM*: 849-880.

Reason, J (1990) *Human error*. New York: Cambridge University Press.

Simmon, D A (1998) Boeing 757 CFIT Accident at Cali, Colombia, Becomes Focus of Lessons Learned. *Flight Safety Digest* May-June: 1-31, available from http://www.svt.ntnu.no/psy/Bjarne.Fjeldsenden/Aviation/CaliAccident.pdf.

Snowdon, P, and Johnson, C W (1998) The Impact of Rhetoric on Accounts of Human 'Error' in Accident Reports: Technical report, Department of Computing Science, University of Glasgow, available from http://www.dcs.gla.ac.uk/~johnson/papers/rhetoric/rhetoric.html.

Interactive Evidence: New Ways To Present Accident Investigation Information

Damian Schofield, Jez Noond, Lorna Goodwin and Jack March,

AIMS Research, SChEME, The University of Nottingham, University Park, Nottingham, NG7 2RD, UK.
Damian.Schofield@nottingham.ac.uk, http://www.nottingham.ac.uk/aims

Abstract: In the UK computer-generated presentations are becoming an increasingly important visual aid in courtroom and inquest situations, where complex data relating to a sequence of events is being visualised before a general public who may have little or no understanding of the circumstances in which an accident occurred. This presentation of evidence was recently discussed by Burns (2001) who stated: "The presentation typically takes the form of a report and the scientist must be prepared to explain this report in such a way that a stereotypically science-phobic judge and jury are able to comprehend it. Presentation is everything." This paper will introduce and discuss a spectrum of new technologies that utilise new developments in Computer Graphics (CG) and Virtual Reality (VR) for a range of accident investigation and evidence presentation scenarios.

Keywords: Computer Graphics, Virtual Reality, Evidence, Accident Investigation.

Introduction

Inevitably the future will be digital: digital cameras and videos, electronic document storage, network commerce, internet business, intelligent search agents, computer animations and virtual simulators are already in use. Computer technology in everyday society has altered the fundamental way that certain parts of day-to-day tasks are performed. As computers become more powerful, users redefine their problems to maximise the capabilities of the technology.

This continuing digital revolution will influence the way accident evidence is collected, analysed, interpreted and presented. This paper will introduce computer technology that allows information to be presented in trials, industrial hearings, public enquiries and settlement claims evaluation in an accessible and easily understood, visual manner. These computer generated visualisations can be used to present scenarios that are based on scientific data or to depicting a witness perception of what may have occurred at a given time and location. More meaningfully, the technology can also explore and illustrate "what if" questions and expose the inconsistencies and discrepancies within evidence and expert witness testimony.

It is important to recognise that the use of such computer technology is only the current manifestation of graphical visualisation in a long history of litigation graphics. The use of any new visualisation medium has always had to establish and win precedent in a legal context.

The authors believe that the use of Computer Graphics (CG) and Virtual Reality (VR) technologies is an important development in the history of litigation graphics that is unparalleled in its approach and ability to assimilate and correlate witness testimony, expert data and forensic procedure. A range of accident and incident reconstructions (including vehicle accidents, industrial fatalities and major incidents) will be discussed in this paper.

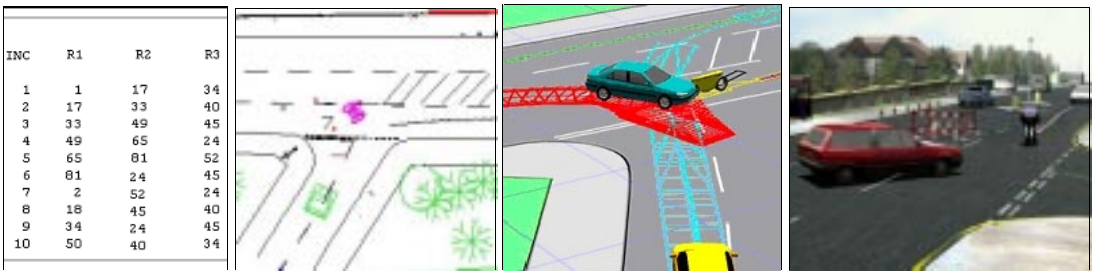
Technology

'Computer Graphics' or 'CG' refers to a suite of computer applications that can be used to produce images and animations. CG utilise numerical three-dimensional (3D) models of real world objects to create artificial environments. Computer technology is employed to build an animation of these environments frame by frame (a series of still images), that, when played back in quick succession, create an experience of space, motion and time. Popular cultural examples of this technique include animated movies such as "Shrek" and "Monsters Inc.".

Based on scene survey data, objects such as equipment, vehicles, human figures, environment details, landscape features and other relevant evidence items can be accurately positioned within the 3D environment. All the scene objects are scaled precisely, and can be texture mapped with relevant images to produce credible lifelike appearances. The facility to visualise and then explore an accident scene enables the viewer to increase their comprehension of the important and underlying issues within that scene.

'Virtual Reality' or 'VR' involves interactive real-time 3D graphical environments that respond to user input and action, such as moving around in the virtual world or operating virtual equipment. An important aspect of VR is its underlying processes, simulations, behaviour and reactions, and the way users interact with objects within the world. A virtual reality user can, for example, sit in a virtual vehicle and drive it. The virtual vehicle will respond to the driver's input and behaviour causing other vehicles in the world to respond to those inputs – such as in the case of a collision (Denby and Schofield, 1999).
Recent and rapid developments in PC technology and the huge potential market for desktop VR have created a climate where many novel applications have emerged. The home computer games market has driven the development of software tools alongside specialist 3D graphics accelerator boards and peripherals. Whilst much of the development is for the leisure industry, there are many real industrial applications being developed for a range of industry sectors. These types of systems can offer advantages over other visualisation media due to interactive nature of the experience they create (Schofield et al, 2001).

Using Virtual Reconstructions
An accident investigation begins with data collection; accuracy is crucial as this data serves as the foundation for the evidence. Traditionally at a scene, an investigator makes field measurements, produces rough field sketches, takes sets of photographs and then later drafts up plans of the scene and collates the information. The evidence from the scene is analysed by experienced and suitably qualified accident investigators. Finally, the investigators present their findings (hopefully, in a clear and precise manner) to a mixed audience of experts and lay people. The evidence must reflect accurately the scientific data available and should augment the testimony of the witnesses. However, to be effective, the evidence must not only tell "the story" but also be understood easily. To that end, accident investigators must strive continuously to develop new and creative ways to present complex evidence (Schofield and Noond, 1999).
The technology used for data collection and measurement ranges from tapes and traditional surveying tools (widely used by private accident investigators) to Electronic Distance Measurement (EDM) technology (widely used by the police) to full 3D laser scanners (such as those used by the UK Transport Research Laboratory). Digital data collection allows the automatic generation of 3D coordinate information that will allow the data to be imported directly into drafting software. These coordinates provide a reliable numerical set for the creation of the geometry that is the foundation of any credible computer model of a scene. This concept is shown in figure 1, which figuratively illustrates the progression from digital data to 3D model.
The environment surrounding the scene may be included within the model. For example, a model may not only show the location of an item of equipment, but also the position of this item in relation to nearby environment features, and place this item within a time frame of scene activity. As seen in the CG used for film and television, the realism in these 'virtual' environments is increasing. As computer-processing power increases and software tools develop it is natural to assume that it will be possible to achieve a similar level of realism within the computer-generated environments used in an accident investigation context.



INC	R1	R2	R3
1	1	17	34
2	17	33	40
3	33	49	45
4	49	65	24
5	65	81	52
6	81	24	45
7	2	52	24
8	18	45	40
9	34	24	45
10	50	40	34

Figure 1: The Progression From Data to Two-Dimensional Plan To Three-Dimensional Reconstruction

The virtual environments created may be used to test hypotheses, such as witness location verification, where lines of sight around obstructions or hazards present in the environment are considered. Other accident data may be embedded within the virtual environment, vehicle and location based accident

statistics may be displayed, calculation and test results visualised and original documents and photographs linked to 3D objects (Schofield and Noond 1999).

Before discussing the use of these virtual environments as evidence in a court or inquest and their subsequent admissibility, it is necessary to clarify the terms used to describe such technology. The standard form of evidence from such virtual environments usually consists of a series of still images and animations. In this context, the term "computer animation" is often misused to describe an animation from a virtual environment that is not based on the laws of physics, but is still represented as "simulating" a given event. The terms 'animation', 'scientific animation' and 'simulation' have specific definitions in the accident reconstruction community (Grimes, 1994).

'Animation' is a general term describing "any presentation which consists of a series of graphical images being sequentially displayed, representing objects in different positions from one image to the next, which implies motion" This term may be used to describe a technical, scientific based presentation or a presentation consisting of artist renditions, sometimes referred to as a "cartoon animation".

The phrase 'Scientific Animation' is consequently used to describe a more technically based presentation, and is defined as "a computer animation that is based on the laws of physics and the appropriate equations of motion". Velocities and positions are time integrals of the acceleration data and objects and environment in a scientific animation are properly and consistently scaled.

In the accident reconstruction community, a 'Simulation' is defined as being based on the laws of physics and containing specific underlying equations. A simulation goes further than a scientific animation and can be further defined as "A model that predicts an outcome. The model may be a physical or a mathematical model, but the significant property is that a simulation predicts a future result".

In summary, an 'animation' may only be illustrative evidence, whereas a 'scientific animation' is more technical and relies upon scientific laws, and thus can be categorised as substantive evidence. The admissibility of a simulation depends on the category of the program used to create it, and thus may either be illustrative or substantive evidence. When scientific evidence is submitted to a court or inquiry, part of the procedural requirements involves the checking of any methodology, including programs and processes used. Depending on the type of software utilised and the method of creating the reconstruction, the evidence may be difficult or relatively straightforward to admit.

Virtual Environment Admissibility

Although computer-generated graphical evidence has only recently started appearing in UK courts, relevant case law from the US and other jurisdictions may be referenced for admissibility requirements. This is particularly true for forensic animations, which by US standards are substantive evidence and thus more complex to admit to trial.

Scientific evidence must pass through a set of criteria before being admitted as evidence in US courts due to potential bias and unfairness, particularly with reference to animations or simulations. Computer-generated evidence is frequently used to explain a broad range of subject matter, in many criminal cases, civil cases (such as personal injury, product liability and patent infringement) as well as public inquiries. The issues in question may be extremely complicated and difficult to explain to the court without some form of graphical representation. A survey by the American Bar Association (ABA) found that jurors are often confused, bored, frustrated and/or overwhelmed by technical issues or complex fact patterns (Krieger, 1992). Other research has indicated that the attention span of the average juror is seven minutes (Lederer, 1994a). This illustrates the need to reduce lengthy explanations that use charts and diagrams alone.

Despite the many benefits of evidence from virtual environments, as with any other form of graphic evidence, these displays may be misused in court, and the consequences of this cannot be underestimated. The persuasive power of 'photo-realistic' rendered displays is also their greatest disadvantage; they leave a strong impression on jurors, they tend to mesmerise, and they relax an individual's natural critical nature. Jurors and triers of fact may be inclined towards a "seeing is believing" attitude, as they do with television (O'Flaherty, 1996).

The use of such technology in the UK is still in its infancy. Reasons quoted for the slow progression of evidence from virtual environments in the UK includes a lack of skills, or simply tradition or that lawyers (who may not be highly trained in computer operation) perceive an air of mystery about the use of computers and await a 'scientific foundation' for courtroom use (Lederer, 1994b and Marriott, 1996).

The first forensic animation shown during a Crown Court criminal case in the UK involved a road traffic accident. This animation was generated by PC Mike Doyle of the West Midlands Police Crash

Investigation and Training Unit. Since then an increasing number of forensic animations have been admitted to UK courtrooms (Fisher, 1998).
Like any other evidence, proving the evidence trail chronologically is necessary; audit trails are needed to ensure that the evidence presented is reproducible from the original data. The audit trail should also note the date and time of creation of models, images and animations. An audit trail also allows the alteration of files to be correlated with digital watermarking systems to ensure the authenticity of any digital evidence. The audit chains for computer-generated evidence can become extremely complex, consisting of numerous links to form the whole. It should be possible to start at any link and move forwards to any item within the finished evidential presentation or backwards to source maintaining the integrity of all data or information throughout (Schofield et al, 2000a).
Digital watermarking software applications are also making their way into the UK legal arena, ensuring that files presented have not been tampered with by an external source. These programs embed information about the author into text documents, video, audio or graphics files. This information, when decoded, can reveal things such as the author's address, reference numbers, terms of use, copyright date, etc. It is important that any such watermarks cannot be deleted or altered. It is also essential to ensure that, while working on producing digital evidence, data files are secure. All models, images and animations must be protected from external (i.e. network) alteration.
As software companies have developed user-friendly graphics packages that utilise the increasingly powerful hardware available, more people have become familiar with 3D graphics systems. CG animation software has also reduced in price and professional quality results can be generated using widely available PC software, including video postproduction and editing. The increasing application of this technology to the presentation of accident data is apparent by the number of new animation companies 'springing up' to offer this service to investigators and the legal profession. It is fair to state, however, that graphics produced by amateur animators and modellers without a forensic\investigation background may not be of sufficient accuracy or quality to be presented professionally in a UK courtroom.
In the second part of this paper the authors will detail case studies taken from a range of applications and each taking a slightly different approach to the presentation of evidence and information.

Case Study 1: Motor Vehicle Accident
The first fatal road accident recorded in Britain (involving the driver and passengers of a motorcar) occurred on 23rd February 1899. While attempting to turn a corner at a speed of over 25 mph the car's wheels collapsed. The occupants were thrown out and the driver and front seat passenger killed. Newspapers of the day hoped that this terrible accident would convince drivers to take greater care and keep their speed down. At the inquest the coroner commented that he hoped this type of accident would never happen again.
Here we are over 100 years after that accident, the cost of road accidents in Britain is now estimated at over £16,000 million per year; this includes hospital costs, damage to property and vehicles, police and insurance costs, lost output, and a notional sum for pain, grief and suffering (RoSPA, 2001).

Figure 2: Images From An Animated Reconstruction Of A Dual Motorcycle Fatality.

The particular case to be discussed here (shown in figure 2) concerns an accident that occurred in the West Midlands, UK, in 1999. The virtual model that were produced from survey data was used extensively in a Walsall Coroner's inquest in helping to establish the events that lead up to the death of two motorcyclists. The animated reconstruction visualised how the two motorcyclists were killed when they collided with a vehicle that pulled across their path at a junction.
A significant and interesting aspect of this particular case was the way in which the animated evidence was subsequently utilised. In the courtroom the animations were used 'interactively', with the reconstructions being stopped at key moments so that witnesses could discuss the speeds and positions of vehicles relative to the scene and to each other. This experience has encouraged the possibility of using a more interactive technology such as VR in the courtroom. This would facilitate the opportunity of allowing the witnesses and investigators involved in a particular case to show interactively their view of what happened.
It was important to safeguard against emotive imagery that could introduce bias into the court. To this end the animations included only the vehicles involved, not victims or any content of a visceral nature. Human figures should be used in animations only when absolutely necessary, for example, in cases of pedestrian collision or where relevant to major crime or incident scenes (Elliot, 1998).

Case Study 2: Industrial Accident
This accident reconstruction is based on investigation information from the West Australian Department of Minerals and Energy (DME WA, 1999). This accident involved the training of a haul truck driver, during the training the truck was driven onto a stockpile, carrying a load of ore. As the truck tipped the ore, the material collapsed and the truck slipped over the edge and overturned. The reconstruction shows the factors that led to this accident and also how the accident could have been avoided.
The Western Australian Department of Minerals and Energy initially planned to use this scenario to train inspectors on how, who and why to prosecute. It was decided to expand this and build a three-part system to help both train and test employees on the circumstances surrounding this particular accident. The core of this system was to be a computer-generated reconstruction of the truck rollover accident described above. The images and animations of the accident reconstruction were to tie in with multi-media based training and interactive VR testing systems to form an integrated training and testing system (Schofield et al, 2000b).
Truck rollovers are not new, the magazine 'Minesafe' which is produced by the Western Australia DME showed a vehicle precariously balance on its rear on the edge of a stockpile. (DME WA, 1990) and its heading was "Dump Precautions". Another issue of the magazine (DME WA, 1994) depicted six photos of trucks that had gone over the edge, the caption this time was "When Will it Stop?" The same magazine in 1997 showed vehicles that went over the edge although targeting the practice of not wearing seat belts, this time the caption read, "Once More We Are Talking Seatbelts! – Be Kind To Yourself" (DME WA, 1997).
Dump trucks are not the only type of vehicle to go over the edge due to bad practise or falls of ground. A front-end loader became bogged on the edge of the stockpile and when the driver reversed his rear wheels went over the edge, no injury was recorded (DME WA, 2000).
The preliminary fatal accident report prepared by the Department of Minerals and Energy detailed the following about the accident reconstructed (DME WA 1999):
"…. It appears from the initial report that at about 1:15pm on Friday 26th May 1999, the deceased was a passenger in a dump truck on the Run-of-Mine (ROM) ore stockpile. The driver reversed the vehicle to the dump point and the truck fell over the edge. It has been suggested that the point where the incident happened was above an area where ore had previously been loaded out from the base of the stockpile."
The accident takes place at the Buzzard Open Pit Mine, near Meekatharra, on the morning of 26th May 1999, Mr. John Spicer was undergoing training as a haul truck driver at the mine. The seat belt on the passenger seat was unusable as one half of the belt had been removed for maintenance two days previously and had not been replaced. At approximately 06.15 am, the truck was driven onto the stockpile to tip the ore. On the previous shift, ore had been removed from the base of the stockpile an the edge of the stockpile was standing near-vertical and was in an unstable condition. As the truck reversed to the edge of the stockpile, the material collapsed under the rear wheels and the truck slipped over the edge and came to rest in upside down with the tray and cab resting on the ground some six metres below. The driver, who was wearing a seat belt, suffered shock and minor injuries; however, Mr. Spicer, who was unable to wear his belt was killed.

Figure 3: A Series of Stills From Virtual Environments Of A Truck Rollover Accident.

A series of stills and animations were created from a number of virtual environments representing different scenes where the incidents leading up too the accident occurred. These begin with the night shift before the accident and show the operation of the front-end loader on the stockpile that caused the instability. The reconstruction then moves to the next day, when the pre-shift truck inspection is carried out, this is when the damaged seatbelt was noticed. The truck then approaches the stockpile, and slides over the edge. A series of stills from the virtual environments created are shown in figure 3.
The accident animation was useful since it could effectively reconstruct the dangerous situations leading up to the accident. The accident situations were viewed from various viewpoints and lines of sight could be examined. Narratives and an atmospheric soundtrack were added to produce a finished item. Although this deviates from the objective ideal of an evidence presentation, it can increase the power of the safety and training message transferred to the workforce (Schofield, 1997).

Figure 4: A Series of Stills From The Multi-Media Training System For A Truck Rollover Accident.

The multi-media system developed for this particular scenario allows the user to investigate the accident by reading about the mine, collecting witness statements (both text and audio versions are available) and investigating the equipment. The user can take the role of an inspector finding out about the causes of the accident. A series example screen from the multi-media system is shown in figure 4. The user can access information about the mine site, mine vehicles and witness statements, relevant video clips from training films as well as computer generated animations and images showing what happened during the accident.
A number of VR environments were built (some of which are shown in figure 5), including one based around a pre-shift truck inspection. Drivers of these large off road haulage vehicles must understand the importance of pre-shift inspections. These inspections are a hazard spotting exercise, ideal for VR training, which allow the driver to increase their safety through the elimination of potentially dangerous scenarios.
The virtual environment created allows the drivers to discover a changing variety of hazards, not only the ones identified in this particular accident scenario. The system also explains the consequences which can occur due to ignoring any particular hazard using the multi-media \ forensic animation system. In this manner the transfer of complex information to the driver is optimised, and their understanding increased.

Figure 5: A Series Of Stills From The Virtual Reality Testing System For The Truck Inspection.

Of particular note is a comment from the State Mining Engineer, Jim Torlach, who emphasises the need to ensure that, not only are the safe working standards of an operation communicated, but also that some positive check is carried out to ensure that the message presented has actually been understood by the employees receiving it (Torlach 1998).

Case Study 3: Post-Mortem Visualisation

This aim of this project is to explore novel techniques of 3D CG visualisation within the field of forensic pathology. Rather than generate the sophisticated medical imagery familiar to professionals within these disciplines, the project aims to assess the plausibility of presenting complex medical evidence and expert opinion in a visual form to a lay audience. The emphasis of this work is to visualise organic systems such as the human body and possible physical circumstances that can be related directly to an autopsy report.

The photographic realism that is possible with modern CG was not an objective in this project. Instead, dimensionally accurate schematic computer models of the human body consisting of case specific external and internal anatomy incorporating inflicted injuries were employed. This mode of representation was deemed more appropriate given the sensitive nature of the data selected, and the overall project aim of demonstrating interaction and process in a broad and general context.

Following discussions with forensic pathologists from the University of Sheffield's Medico-Legal Centre, two stabbing cases were selected for visualisation. The two cases, one suicide and one murder, presented the opportunity to make a series of visual comparisons related to both weapon type and to the type of trauma inflicted upon the body.

The suicide case was selected primarily for the potential of visualising a knife comparison relating to two discrete injuries. The autopsy notes described a non-fatal wrist incision chronologically followed by a fatal stab wound to the chest, both consistent with self-infliction. The transverse linear incision to the wrist severed the flexor tendons of the left arm but failed to penetrate either the ulna or radial arteries. The stab wound to the chest, however, penetrated the chest cavity from the front, cutting through the heart, entering the front of the right ventricle, passing through the left ventricle and emerging from the rear wall of the heart. Two knives were recovered at the scene, a long bladed kitchen knife, consistent with both injuries, and a 'Stanley' type knife, consistent only with the incision to the wrist.

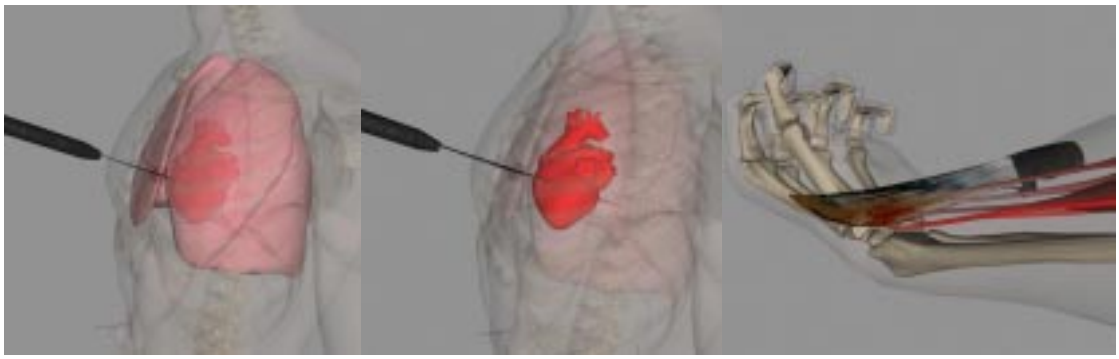


Figure 6: A Series Of Stills From The Suicide Post Mortem Visualisation.

For the wound and knife blade dimensions a high level of modelling accuracy was possible due to the detailed measurements presented in the autopsy report. Relevant internal structures and organs were modelled and positioned relative to 'normal' anatomical measurements and illustrations from medical texts. To act as a test for this form of data visualisation the anatomical model, complete with external injuries, was assembled first in a standard anatomical position. The knife blades were then added, positioned relative to the external wounds and angle and length of wound track. When the kitchen knife was viewed relative to the chest stab wound and internal organs the knife blade produced a match to the wounding as described in the autopsy report. An animation was produced to depict the consistency of the kitchen knife with the fatal chest wound compared to the 'Stanley' type knife, the latter's blade being too short for the depth of the wound. The consistency of both knife blades with the wrist incision was also visualized.
Animations and stills from the model was shown to a group of staff from Sheffield's Medico-Legal Centre, including several forensic pathologists and coroners. One point of discussion was the anatomical dynamics of self-inflicted wrist incisions. How, on movement and tensing of the wrist and hand, flexor tendons involved in this movement can restrict access to the underlying arteries. It was felt by the researchers that just such a dynamic lent itself to the advantages of 3D modelling and animation, and as such was included in a subsequent version of the visualisation. Stills from this animation are shown in figure 6.

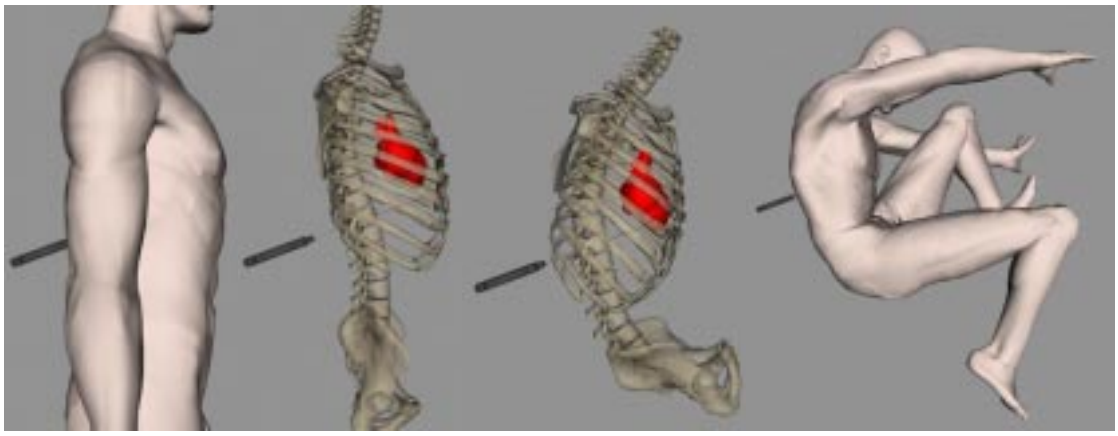


Figure 7: A Series Of Stills From The Murder Post Mortem Visualisation.

This autopsy report for the murder case described several blunt force injuries, including bruising and lacerations to the face and chest of the victim, an incised defence wound to the back of the left hand, and a stab wound to the back between the 11^{th} and 12^{th} ribs. This stab wound passed through the left half of the 11^{th} thoracic vertebrae, continuing through the chest piercing the heart.
The same visualisation methodology was used as in the suicide case. However, unlike the previous visualisation, when the knife blade, external wound, knife track length and angle were lined up the internal documented wounds were not replicated. The knife did not reach the heart.
After discussion with experts from the Medico-Legal Centre, the dynamics of the violent situation were added. The forensic autopsy was conducted with the victim in a prone position, a position which is often different from the situation in which the injury was received. A revised visualisation was consequently produced which included a hypothetical victim dynamic, this crouching position allowed the knife to pierce the heart, as shown in figure 7.
Staff at the Sheffield Medico Legal Centre and researchers at the University of Nottingham feel that these models raise a number of issues that warrant further exploration. These issues include an agreement on the mode of representation of objects involved in the cause of death, the further use of dynamic representation, both of individuals (to include accurate body physics) and their environment, as well as the dynamics of anatomical structures in live situations and under autopsy conditions. With continued development, this type of data visualisation may ultimately be used in UK courtrooms and inquests and may have a potential in hypothesis testing.

Case Study 4: Interactive Evidence

VR offers a unique platform for the collation, interrogation, analysis and presentation of forensic data across a wide spectrum of crime-scene scenarios. This case study explores concepts for real world computer applications in a VR 'scenario' which was constructed using both proprietary software and software developed at the University of Nottingham.

This 'scenario' is designed to explore contexts relating to effective and interactive scene visualisation, including facilitating understanding of chronology and technical data for a jury and also potentially training future investigative and scene management personnel.

In this particular scenario, an inner city underground car park has been the scene of two incidents:

- A vehicle fire involving a Renault 5.
- The suspicious death of a passenger within a Toyota pick-up vehicle.

Figure 8. Images From An Interactive VR Environment And Hypothetical Crime Scene.

The brief relating to this scene was to visualise any possible relationship between the two incidents, especially in relation to smoke toxicity and temperature. The main area of focus within the environment became the visualisation of Computational Fluid Dynamics (CFD) data. Sophisticated transient simulation geometry for smoke within the car park was calculated using CFD software and a CAD model of the environment. 'Snap shot' smoke geometry encompassing the entire duration of the fire was then selected from the data and imported into the VR world, allowing a user to explore the two incidents during the significant time frames of smoke density and extent. Users are given the option of correlating this experience with the recorded evidence of investigators, by selecting objects within the world and accessing relevant information as required. The CFD data is shown in the VR world in figure 8.

A visualisation derived directly from calculated data has many significant implications for jury members. Improved understanding of technical data and a shared visual experience are the most obvious. Issues relating to a visualisation's scientific credibility are important too, as the jury members need to be as sure as possible that what they see is what has been calculated. VR will have an important impact within many cases as the technology and the forensic community develops. Landmark opportunities exist within the legal system, which would benefit enormously from accurate and interactive VR environments that jury members trust. VR will aid investigative personnel during training and later in dealing with the vast array of spatial and technical data associated with public enquiries, major incidents and disasters.

Conclusions

The ultimate question that must be answered is whether or not computer generated visualisations help us to understand what happened in an incident or accident more clearly than can be achieved by existing means. The exponential increase in computational power and the development of sophisticated tools with which to create 3D worlds has led to a massive improvement in the realism and credibility of computer generated images, animations and environments. The ability to represent a range of dynamic, interactive scenarios on a computer screen and view those scenarios from any angle enables forensic investigators, expert witnesses, and lay people, to better understand the underlying issues related to a particular accident.

The technology described in this paper has been successfully applied in a wide range of fields already, from vehicle accident reconstruction to major crime scenes, from industrial accidents to maritime and aviation

disaster visualisations. The rigorous application of guidelines and standards during the generation and presentation of such material will win favour across a professional community striving to visualise complex scenarios. In this respect, it won't be long before legal precedents are won within the UK legal system, enabling CG and VR to become as admissible as other existing forms of litigation graphics such as photography and closed circuit television (CCTV) footage.

References

Borelli, M. (1996). The Computer as Advocate: An Approach to Computer-Generated Displays in the Courtroom, Indiana Law Journal, Volume 71, Number 2.

Burns, D.C. (2001). When Used in the Criminal Legal Process Forensic Science Shows a Bias in Favour of the Prosecution. Science and Justice, Volume 41, Number 4, pp 271 – 277.

Collier, P. and Spaul, B. (1994). A Forensic Methodology for Countering Computer Crime, In: Carr, I. and Williams, K., ed. Computers and Law, Oxford: Intellect.

Denby, B. and Schofield, D. (1999). Role of Virtual Reality in Safety Training, Mining Engineering, October: 59 – 64.

DME WA (1990). Dump Precautions, Minesafe, Volume1, Number 5.

DME WA (1994). When Will It Stop ?, Minesafe, Volume 5, Number 4.

DME WA (1997). Seatbelts !, Minesafe, Volume 8, Number 4.

DME WA (2002). External Information System (EXIS), Department of Minerals and Energy, Internet WWW Page, at URL: http://notesweb.dme.wa.gov.au/exis/Exisonweb.nsf, (Version Current 15th April 2002).

DME WA (2000). MOD For Your Information, Incident Report, Number 448, 18th March.

Elliot, D. W. (1998). Videotape Evidence: The Risk of Over-Persuasion, Criminal Law Review, 59 – 174.

Fisher, P. (1998). Should We Believe The New Realism, The Daily Telegraph, June 18th.

Grimes, W.D., Classifying the Elements in a Scientific Animation, Accident Reconstruction: Technology and Animation 4, Warrendale, USA, Society of Automotive Engineers, pp 397 - 404, 1994.

Krieger, R. (1992) Sophisticated Computer Graphics Come of Age—and Evidence Will Never Be the Same, A.B.A. Journal, December.

Lederer, F. I. (1994a). Technology Comes to the Courtroom, Emory Law Journal, Number 43, 1095 – 1113.

Lederer, F. I. (1994b), Courtroom 21: A Model Courtroom of the 21st Century, Court Technology Bulletin, Volume 6, Number 1.

Marriott, A. (1996). Alexandria Firm Courts Lawyers by Proving Animation Computes, The Washington Times, November 25th 1996.

O'Flaherty, D. (1996). Computer-Generated Displays in the Courtroom: For Better or Worse, Web JCLI, Number 4, Internet WWW Page, at URL: http://webjcli.ncl.ac.uk/1996/issue4/oflah4.html, (Version Current 15th April, 2002).

RoSPA, (2001). The RoSPA Guide to Road Safety Projects, Royal Society for the Prevention of Accidents, Internet WW Page, at URL: http://www.rospa.co.uk/cms/viewarticle.asp?article=2472&scheme=7, (Version Current 15th April 2002).

Schofield, D. (1997). Virtual Reality for Training and Risk Assessment - Practical Improvements, Workshop on Risk Assessment, Safety and Health Commission, European Commission, Luxembourg, 11th-12th November 1997.

Schofield, D. and Noond, J. (199). Accident Reconstruction: Possible Futures, Senior Accident Investigators Conference, Chelmsford, Essex, April 24th – 25th 1999.

Schofield, D., Noond, J., Goodwin, L., Fowle, K. and Doyle, M. (2000a). How Real is Your Reconstruction: New Developments in Computer Graphics and Virtual Reality, Proceedings of Forensic Techniques for the 21st Century Conference, Lloyds of London, 16th – 17th October 2000.

Schofield, D., Noond, J., Goodwin, L. and Fowle, K. (2000b). Recreating Reality - Using Computer Generated Forensic Animations to Reconstruct Accident Scenarios, Proceedings of Minesafe 2000 Conference, Perth, Australia, pp 483-494, 4th – 8th September 2000.

Schofield, D., Denby, B. and Hollands, R. (2001). Mine Safety in the Twenty-First Century: The Application of Computer Graphics and Virtual Reality, In: Karmis, M. ed. Mine Health and Safety Management. Colorado: Society of Mining, Metallurgy, and Exploration, pp 153 – 174.

Torlach, J. M. (1998). DME Bulletin Number 40, WA State Mining Engineer– Safety Bulletin, Number 40.

Genesis of a Feedback System Based on Human Factors for the Prevention of Accidents in General Aviation

Bernard Boudou (1) Olivier Ferrante (2)

(1) REC Coordinator, bernard.boudou@bea-fr.org.

(2) Safety Analysis Division, olivier.ferrante@bea-fr.org.
Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile (BEA),
93352 Le Bourget Cedex, France.

Abstract
This paper is an adaptation of a study that can be found on the website (http://www.bea-fr.org/anglaise/rapports/etudesetstatistiques.htm). It put emphasis on the need of developing a feedback system, the ultimate way of improving safety. It is based on voluntary reporting and aimed to incorporate human factors issues in general aviation. The operational problems and appropriate solutions are highlighted throughout this paper.

Keywords
Safety, accident prevention, confidential reporting system, human factors, feedback, voluntary.

Background
Civil aviation can only survive if all parties involved -- both players and customers -- are convinced that all operations are undertaken in a safe environment. Such safety is often perceived as the tranquillity of mind that results from the certainty that no disaster is about to strike. Civil aviation is commonly divided into two components: public air transport of passengers or freight (these activities are charged for) and general aviation, comprising everything that is not public transport (leisure flying, training, aviation-related activities).
The purpose of this paper is to describe the creation of the French confidential reporting system (REC – Recueil d'Evénements Confidentiel) for improving safety in general aviation.
The study commences with a short preview of safety in general aviation, concentrating in particular on comparisons with public air transport, and argues for the setting-up of an additional feedback system. Next the derived benefits and limits of such a system are analysed. Finally, a description of practical implementation of such a system is employed to highlight the related operational problems and indicate appropriate solutions. The emphasis is placed on the incorporation of human factors, through two scenarii: (1) Man is the main player in the sphere of aviation-related activity, he being the one who carries out the activity; (2) Man, beyond the notion of aviation-related activity, takes a step back to consider his own actions and accounts for them in order to supply a feedback system.

Assessing the safety level

With regard to public air transport, average activity per year in France may be appraised using certain figures: around two million departures representing 100 million passengers and 10 million flight hours. On average, accidents account for approximately ten deaths per year. Thus, the "tolerated" safety coefficient for public air transport is in the order of 10^{-6}, that is to say one death in one million flight hours. The order of scale for general aviation is around two million flights per year, representing approximately one million flight hours, in the course of which accidents result for around a hundred deaths. The safety coefficient can thus be evaluated as 10^{-4}, that is to say, one death per ten thousand flight hours.

The situation regarding feedback: Feedback routes are numerous. With regard to accidents and serious incidents, the Civil Aviation Code stipulates that all such events require an obligatory declaration, and that the technical investigations directed by the BEA shall result in publication of reports which may include recommendations. In the case of incidents, the BEA decides to what extent it should become involved.

In the realm of public air transport, technical investigations into accidents and serious incidents may become extremely involved, since organisations are highly structured, the parties involved are clearly identified, and procedures are standardised. In conformity (notably) with the orientation adopted by JAR-OPS-1 (Civil Aviation Code, 1997), public air transport undertakings have established accident-prevention and flight safety programs. Such programs include all types of information-gathering systems, notably flight analysis through the systematic processing of information gathered by onboard recorders, the appointment of flight safety officers, the issuing of flight safety bulletins, the collation of volunteer reports etc. Thus large investments are devoted to investigating incidents and minor events occurring frequently during operations.
For general aviation, the BEA publishes a monthly information bulletin detailing particularly representative accidents and serious incidents. Since the system lacks intricate structure or wide-ranging standards, investigating the root causes of such events is often impossible. For this reason, reports are generally factual. The BEA may also publish general operational reports covering several general-aviation accidents having related causes. The latter documents contain more detailed analysis and causal research than the information bulletin. Nonetheless, apart from some very rare exceptions, general aviation organisations have not set up their own systems to handle feedback regarding unaccustomed situations or minor events. Moreover, no such statutory constraint exists in this respect.

Event-characterisation: Thanks to the high safety levels attained in public air transport, accidents are very rare. It is therefore difficult to draw parallels between accidents, or classify them into categories. Incident-reporting produces information regarding isolated factors that might lead to accidents, but the combination of factors that could lead to a disaster is unforeseeable. Feedback systems based on minor events give information about the state of the organisation in general, the suitability of personnel for their respective tasks, or changes in the perception of hazards etc (Rasmussen, 1997).

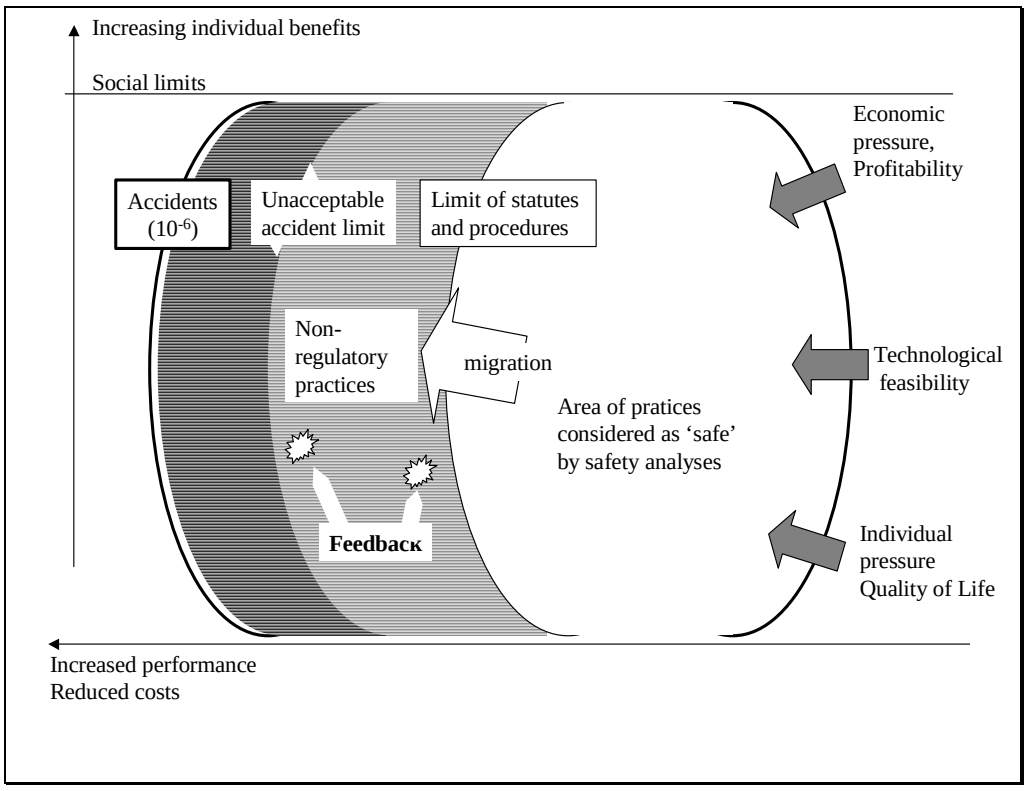


Figure 1 – Migration of Practices in Air Transport

Figure 1 illustrates the shift in the field of operations by air-transport players, under the effects of economic constraints, technological advances and individual human orientations toward least effort. The shifts take

place routinely from an area considered as "safe", towards an area situated beyond regulations and procedures, characterised by blurred boundaries and not addressed by safety analyses. Feedback from this area allows a better understanding of the reality of the activity, and highlights events occurring closer to the unacceptable incident limit. It can be seen that the realm of operations still remains quite remote from the accident boundary.

By allowing an overall view of the operation of a complex system such as public air transport, and identifying failings both in the organisation and on the part of its constituent players, feedback systems provide a means of identifying risks and making adequate corrections. In general aviation, the safety level is lower. Most fatal accidents can be grouped according to origin, into three categories: a strong desire to reach one's destination in poor weather conditions, low-altitude flight, aerobatics.
The general-aviation information bulletin concerning serious accidents and incidents has been published for the past six years. Studies have been provided over the last few years. Nonetheless, the safety level appears to be stationary, with no significant change in the rate of fatal accidents. Would it have increased had the two instruments (bulletins and studies) not existed?

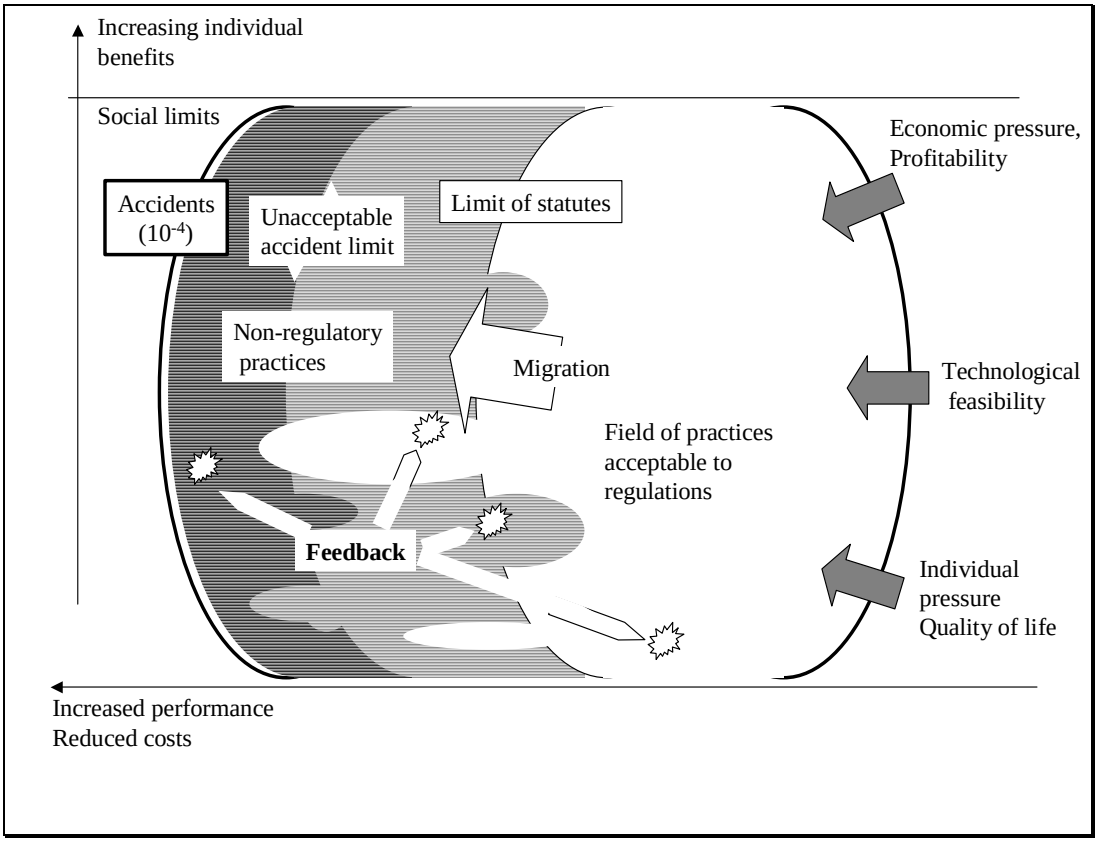


Figure 2 – Positioning and Migration of Practices in General Aviation

Figure 2 is derived from the preceding one. It attempts to explain the situation for general aviation. Conditioned by constraints similar to those of public air transport, the normal practice of activity in general aviation is in a field particularly constrained by existing regulations. Indeed, this activity is characterised by a paucity of defined procedures, the rarity of safety analyses, and the lack of statutory instruments. The latter, extremely precise on certain points, allow a wide latitude of interpretation on others and, in certain cases, a very wide degree of freedom. Cases arise where an accident occurs despite total application of regulations. In the absence of an efficient system for the compilation of events that otherwise left no trace, unacceptable incidents generally pass unnoticed. Therefore feedback could meaningfully address several types of events: incidents that exposed the persons concerned to danger, unaccustomed situations occurring

in a blurred regulatory context, where correct response requires experience on the part of the persons involved, events occurring in the framework of non-regulatory practices, cases where the application of existing regulations is problematic…

The limits of current feedback: Following a public air transport disaster, the material and human means required for performing a technical investigation are set in place. Furthermore, public air transport undertakings have found the necessary resources for setting up internal feedback systems for reporting minor events. The limiting factors in this field are beyond the scope of this paper. For accident investigations in general aviation, clues pertaining to equipment items are generally much easier to detect (breakages, glue failure, fuel contamination etc.) than those relating to human factors. For this reason, investigations often run into problems for several reasons: absence of flight recorders (data and/or voice), often uncontrolled activity leaving no traces, few written documents, since activity is little standardised, pressure exercised by the social environment (where each individual recognises himself...), sometimes heavy media presence, judicial implications influencing the reliability of testimony, disappearance or amnesia of first-line actors (pilots, passengers), lack of eye witnesses, leisure activity in an associative environment secondary to the professional activities of the persons involved. Given such difficulties, aspects associated with the human element are very often difficult to determine. The context is one of mistakes, incomplete knowledge, incorrect judgement etc. Yet such factors are the cause of the vast majority of general aviation accidents. Detailed explanation of such accidents would require a solution to or a workaround for the difficulties mentioned above, in order to deepen investigation in the human-factors field. Although investigators want to detail the context in which the accident occurred, many factors exist to limit their investigative field, and thus the utility of their findings towards improving safety.

The requirement for an additional feedback system for general aviation

For these reasons it was decided to set up a feedback system for reporting minor events, particular situations or unusual circumstances. Certain exploratory studies were undertaken by the Civil Aviation Authorities. The law voted in 1999 (Civil Aviation Code, 1999) facilitated the creation of such a system within the investigative set-up. Thus, in coordination with the DGAC (French CAA), general-aviation user groups and professional bodies, it was decided to create a confidential reporting system (REC) within the BEA (French Accident Investigation Bureau).

The characteristics of the REC are described further in this paper. At this point, a simple overview of the system shows that such a system requires mutual confidence between the specialists responsible for running it and aviation users (the term "user" means "a person acting in the general aviation framework" ; It might be a pilot, instructor, controller, mechanic, ground service personnel etc). The system functions on the basis of the free will of parties submitting reports and must guarantee confidentiality and anonymity for parties involved in events.

Objectives and limits of the REC

Since safety levels in general aviation are evaluated at 10^{-4}, very often the same accident-causes produce the same effects. It is, however, difficult to deepen the technical investigative process, therefore the ability to use information to prevent other accidents is limited. From another angle, many harmless events occur during aviation activity. A brief examination shows that the origins of these events are the same as those of accidents. The idea is therefore to collect information about such events, for formatting, then employ this as feedback, redirected either back to users, or to organisations (administrations, manufacturers etc.).

The events experienced by reporting parties are either situations perceived as anomalous -- generally incidents or events which could have compromised safety or situations perceived as “normal” but out of the ordinary, or special or rarely encountered, which might represent a difficulty for another user or which need to be passed on for the information of the organisations.

Incidents: An "incident" differs in particular from an "accident" by its effects, although the origins are similar. An incident is caused by a reduced series of factors not involving serious consequences. If the incident is of a technical nature or associated directly with the environment, the person will generally disclose in all cases, since his own involvement in it is limited, the occurrence is undoubted, the fault is reproducible or is still visible. This kind of event can be taken into consideration without delay. (Example:

Pilot abandoned takeoff due to appearance of low-voltage warning. Informed airfield control authority immediately, and also the aircraft maintenance workshop.) In many cases, the technical part of the incident is reported, but the state of mind of the first-line actor coping with it too often goes unnoticed: how was the anomaly detected? Which mental processes drove the actor in his decision? What were the consequences of the event regarding continuance of the flight?

This "other aspect" of the incident represents a more interesting source of feedback. (Example: An experienced pilot closed at speed on a controlled airfield just as the radio frequency was very busy. The pilot could not obtain prior authority to join the runway circuit, but he saw that this manoeuvre would not impede other traffic. At the start of the downwind leg, he managed to contact the controller and stated that he was "a few seconds from the circuit". He continued his flight as authorised, to touchdown. At the time of his arrival, the pilot had three options: (1) reduce his speed, (2) extend the flight path and fly a longer downwind leg, (3) join the circuit directly, and "negotiate" with the controller. Faced with this special situation, the pilot relied on experience to reach his decision. We could imagine the same situation occurring with a student pilot on a solo navigation flight. Unable -- at the time -- to determine the first two solutions, this pilot joins the runway circuit, much preoccupied with the fact he is failing to meet a statutory requirement. This additional stress may reduce his awareness and even compromise a safe landing. An accident-prevention measure could be derived from the situation as occurring with the experienced pilot. This could serve as a concrete case study presented to the overall pilot community, indicating a method for evaluating the difficulties and hazards associated with each of the three solutions. Instructors could use the case study as a starting point for useful discussions with their students before solo flights.)

Finally, an incident may consist solely of an error of understanding, mishandling etc. It generally goes unobserved, at the best becoming a conversation-piece in the "squadron bar" between users with little feel for safety. This is the kind of event of interest to the REC approach. Here again, the investigation of this kind of incident avoids some of the problems mentioned in the previous chapter. The first-line witness, or involved party, can speak freely about what happened since (in particular) there are no social, public or judicial implications etc.

The investigation and publication of information concerning incidents is of particular importance in improving safety. In fact the difference between an incident and an accident might depend solely on the presence of a simple aggravating factor. This factor might be a lack of knowledge, a lack of skill, an accident-forming circumstance, defective protection etc. One of the objectives of the REC is to provide users with incident narratives highlighting the full range of causes and designed to stimulate thinking into each factor potentially contributing to the accident.

Unusual situations: Having coped with an unusual situation, an actor may be inspired by dual intent: (1) Actor considers that his experience may be useful to others. His narrative, once formatted, is circulated to all users. (Example: In visual flight, the loss of visual references is the source of many fatal accidents. Normally, nothing is known about possible gaps in the pilot's training, his/her habits etc. However, a good number of pilots have lost their visual references, then recovered them, at the risk of a severe fright... They probably drew interesting conclusions after their experience, but never went on to talk about it: no one ever benefited from their experience.) (2) Actor decides to inform the authorities as to the reality of the event. Details are therefore sent to the organisations in question.

Limits: The systematic processing of event narratives is beset with difficulties, bias and potential pitfalls (Amalberti, Barriquault, 1999).

Risks of spurious results: Unaccustomed situations and minor incidents are very numerous during aviation activity. Reports are submitted according to the desires of the actors involved, and no investigative acts are undertaken on the reported events. Therefore, the quantity and quality of narratives collated by the REC would never perfectly represent the reality of what had actually occurred. It would therefore be illusory to want to establish credible statistics (epidemiological studies) from the REC database. On the contrary, many events are of interest for safety if they are studied and analysed one by one (clinical studies). Furthermore, it is accepted that even figures based on indisputable facts (for example, accidents) must be interpreted carefully. Additionally, events reported though the REC cannot of themselves constitute an

argument for the modification or drafting of statutory instruments. On the other hand they may be used to explain or illustrate the application of such legislation.

Bias in the treatment and understanding of information: One question which frequently arises is whether the description of the facts, and the safety message received by a user after feedback, correspond to the situation experienced by the author of the report. Several types of bias could be mentioned: Did the author fully understand all the facts and circumstances that made up "his" event? Was his perception distorted by preconceived ideas? And certainly, the desire to submit a report testifies to a personal approach whose sole purpose is to defend one's own convictions. Does the design of the form have any influence on the way the event is described? A system of check boxes or a large number of headings to fill out presupposes the a-priori definition of every type of event, and may disconcert certain authors. On the other hand, a blank page may discourage certain others. Does the narrative received by the REC specialist risk being interpreted involuntarily and analysed according to his own convictions? And perhaps insertion of the narrative in a database will be conditioned by the necessary "simplification" of the fields to fill out. And certainly, the way in which the database is employed, the selection of which events are "important" or "representative", or drafting of the text conveying the feedback depends on the specialist's own idea of what constitutes "safety".
In view of the above considerations, surely a confidential reporting system itself runs the risk of addressing only known problems. Finally, should only aircraft pilots have access to the system?

Several answers are proposed to such questions. Any general-aviation actor may submit reports. Receiving several reports from several actors concerning a single event is not ruled out: even if no connection can be established between such reports, they nonetheless offer different points of view that may be pertinent to safety. The report form was designed with very few systematic categories. The author is free to describe the event in just a few lines only, or over several pages. He may add a diagram if necessary. REC is above all interested in the human-factors aspects. Codifying these by means of a large number of categories or check-boxes might disconcert the author of the report, and would still not necessarily reflect the reality of what had occurred. If a problem of a technical nature is mentioned, information such as fault-detection steps, corrective actions, the decisions taken by the person involved, the consequences for continuation of the flight etc. are judged to be a determining factor for REC, together with any information relating to possible dysfunctions at the organisational level and pertinent to the problematic situation. Upon receipt of the report, a telephone call is systematically made to the author of the report. The aim is to validate the narrative, collect additional information, and verify that the safety message understood by the REC specialist effectively corresponds to that intended by the author. Practically the entire text submitted by the author is kept in the REC database. The only items systematically excluded would be identification data of no use in understanding the event. If there is sufficient room, this same text is reproduced in the feedback bulletin or sent to organisations. Since they undertake general aviation activities themselves, REC specialists have a good knowledge of the environment. They undergo training on accident prevention and the difficulties associated with interpreting information received.

Difficulties in measuring system efficacy: Efficacy in the domain of accident prevention is practically impossible to measure. A large improvement in safety levels would only have quantifiable effects after several years, whereas a slight improvement would require several decades to be measurable. Moreover, particularly in the latter case, such measurement would be disrupted by a large number of economic, social or cultural parameters etc. Nevertheless, even a slight improvement could translate into fewer deaths and injuries. This pleads in favour of creating an additional safety system.

Assessment: Both the objectives and the limits of the system are clearly established. Assessment of the objectives indicates that there would be greater risk in doing nothing than in attempting to create the system in question. Nonetheless, such a system would need to be employed with caution, and by exploring the projected field as thoroughly as possible.

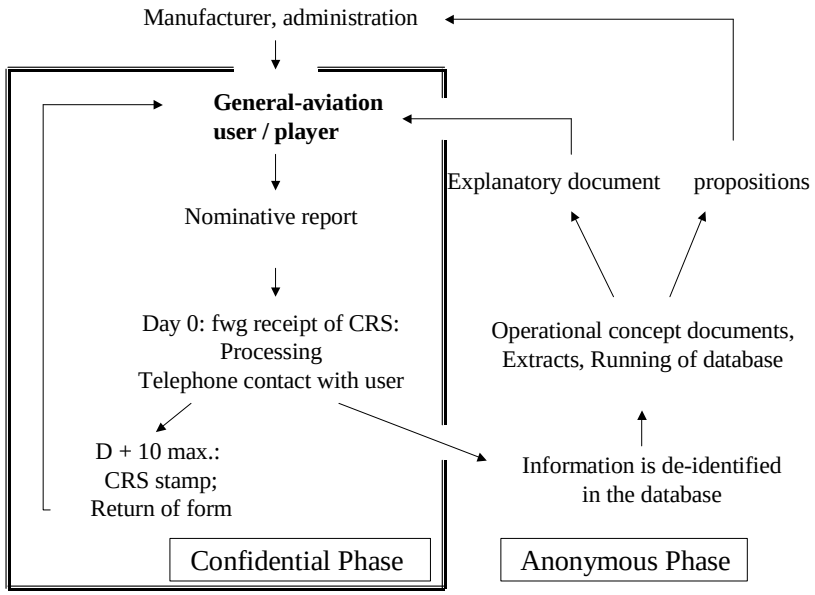
International framework
Incentives for the creation of volunteer report compilation systems: The ICAO General Assembly of October 1997 adopted resolutions 32-15 and 31-10 concerning voluntary reporting systems. In the light of these two resolutions, the Accident Investigation and Prevention Group (AIG, Montreal, 14 to 24

September 1999) examined proposals to amend the existing Chapter 7 of Annex 13 to the Chicago Convention regarding civil-aviation accidents and incidents: voluntary incident reporting systems are non-punitive and guarantee the protection of information sources.
European Directive 94/56/CE dated 21 November 1994, establishing the fundamental principles governing accident investigations and incidents in civil aviation, indicates that activities devolving to the permanent body (in France, the BEA) may be extended to the collection and analysis of information relative to flight safety. The permanent body must be functionally independent of any party whose interests might conflict with its mission.

Operating Principle of Foreign Systems: Certain of the benefits to be drawn in the accident-prevention field, and consistent with ICAO recommendations, several states have set up systems identical to that presented here, with comparable statutory mechanisms. The operating principle is based on voluntary action and confidence, and guaranteed confidentiality and anonymity. Depending on the legislation in force, the author is generally protected from any penalties. In all cases, the report is addressed to the compilation system by secure means, most often by post. A telephone call might be made to the author. The personnel running the system do not keep any nominative information, and the database information do not allow identification of any individual.
Currently, several systems are operating throughout the world (ASRS – Aviation Safety Reporting System in the United States, CHIRP – Confidential Human Factors Incident Reporting Program etc.). Since 1990, the International Confidential Reporting Systems (ICUs) has met periodically to encourage the creation of such systems and to stimulate co-operation.

Study case: EUCARE System: Designed in the 'Nineties, EUCARE (European Confidential Aviation Safety Reporting Network) was designed as a minor-event compilation network open to all parties engaged in civil aviation in the European states, whether in public air transport or general aviation. Conditions for confidentiality and anonymity were to have been met. The Technical University of Berlin was put in charge of producing the system. Unfortunately, the project ceased on 30 June 1999 due to the level of criticism and mistrust it aroused on the part of administrations and user organisations, particularly in Germany.

Practical Implementation of the REC



Confidential phase: During this phase (two weeks maximum), only the user and a REC specialist are in the know as to the details of the event.

Sending the report: An aviation-community actor experiences an incident, or particular situation, and considers that his experience could be of benefit to a large number of individuals. He gets a report form directly from his home organisation, or requests one by phoning (toll-free number 0 810 000 334) the REC, which will mail one straight away. Once the form has been filled out, it can be folded to form a pre-paid envelope, and posted at no cost.

Receipt and initial analysis at REC: Upon receipt, the envelope is locked in the safe of the secure premises employed by the REC. The form is processed in the same premises, as quickly as possible by a REC specialist. After an initial reading, a telephone call is made to the author of the report, in order to: -ensure that the safety message perceived by the specialist corresponds precisely to that which the author wanted to convey, -request any additional information required, -consult as to what information shall be retained in the database, if necessary direct the author towards additional procedures. Within ten days of receipt by the REC, the specialist affixes a date stamp on the form and sends it back to the author. Throughout this phase, no list, numbering system, direct or indirect nominative data or copy is made or kept from the author's report. Only the author retains proof of the information submitted to the REC, in the form of the stamp on the form, showing the date of receipt.

Anonymous phase: entering information in the database: The stored data do not allow any direct or indirect identification of individuals. Thus, not only are names, addresses and telephone numbers removed, but also references to aeronautical titles, aircraft registrations and the precise date and place are eliminated. If it is important to an understanding of the event, the aircraft type may be shown, with the author's agreement. All that is kept is useful information necessary for understanding the event and describing the situation, insofar as it can be easily generalized. (Example: A pilot experienced difficulty in stabilising his aircraft on final approach to XX airfield, which could have compromised the success of the landing.) It is obvious that the aircraft type, the precise date and the name of the airfield are not generally applicable for the prevention of accidents. On the other hand, the wind force and direction, the airfield layout, and presence of obstacles upwind of the flight path, together with information regarding the season, the time, possible pilot fatigue, the sun position, the weight and balance of the aircraft etc. are factors that may be encountered in a multitude of situations, and which must be kept in the database.

Figure 3 – REC Operation

While the prevention of accidents implies incorporation of human factors in order to investigate events occurring in the course of aviation activity, the voluntary submittal of minor-event reports calls upon the human element in a different manner. The aim is to create a climate of confidence, and dissipate a current suspicion with regard to organisations based in Paris. The crux of the matter is how to adapt an information-gathering system to the particular actor in the aviation community.

Initial phase: information-gathering: The objective of the initial phase -- collating reliable information -- depends solely on the free will of actors from the aviation community. Functioning of this initial phase brings to light a certain number of questions. The answers provided for these questions constitute the main features of REC. They are mainly designed to dispel an author's reticence to declare an event.

Declaration procedures: The number of existing organisations; the types of forms to use; identifying the receiving parties; incident-reporting procedures etc: Such factors cause a certain amount of confusion in the minds of users. In many cases, the latter are unaware of the very existence of the organisation, or do not have the proper forms etc. In short, most events encountered are processed by no organisation at all, and are therefore lost. For minor events, REC is seen as a unique system, known to all and for use by all. A telephone number is provided for users wishing to know more about the workings of the system. An informative leaflet is available, including a form; this can be filled out, folded and mailed back at the prepaid-postage rate. If the user needs to know about an event imposing an obligatory declaration to another organisation, the specialist will channel the user's approach along the proper lines.

Statutory constraints regarding declaration of events: In general aviation, there is a very limited statutory obligation to declare an incident. Since there is generally no trace, it is highly likely that rigid arrangements to force actors to report would have any effect; quite the opposite, such an approach could result in distrust

on the part of users, leading to total silence regarding the existence of the said incidents. In short, the effect would be contrary to that sought. Even when an obligation to report does exist, the process is not generally applied, either due to ignorance of the regulations, or unavailability of the proper forms, or confusion as to the particular recipients etc. REC operates on the basis of user free will. The users become involved not in order to satisfy a statutory constraint, but because they are convinced that their approach may contribute to improved safety.

Deciding on what is worthwhile: Many proficient pilots, instructors, mechanics etc. consider that the events they experience are not exceptional, forming part of standard aviation activity, and therefore that there is no point in relating them. For a novice actor, the same events may be exceptional, even putting him/her in a bad position, increasing his workload or leading to an accident. Additionally, accidents also occur with experienced users, often because they were unaware of how to react to an unexpected situation. Reporting parties are assured that their reports will be taken into account. They can be used in three possible ways to support feedback: (1) publication in the bulletin (**REC Info**), for circulation to all users, (2) transmission to various organisations (administrations, manufacturers etc.), (3) included in a corresponding "event group", for the purposes of general safety analyses.

The dilemma of publicising one's errors: Most events of interest to REC are those associated with a human failing. The process is about mistakes: handling errors, lack of understanding, errors of judgement etc. A user may have doubts about associating his name to such types of failings. The reasons are multiple: generally, an individual experiences difficulties in simply recognising his/her mistakes, since he either tries to justify them by criticism, or is ashamed of them (the mistake devalues the image that the individual has of himself); in some cases individuals dread evaluations or value judgements made by others, especially if the value judgement risks being amplified and distorted by rumour without the victim's knowing (the mistake devalues the image that the individual displays to members of his entourage).
To alleviate users' fears, REC specialists offer two strong commitments: Confidentiality of received information. Under no pretext shall information enabling identification of parties be divulged to any person whatsoever. The REC does not retain any copies of documents, nor any lists or numbers. After processing, the nominative report is returned to sender. Only the sender retains any proof that he/she effectively declared an event. Deidentification of information kept or used by the REC. Only those mentions necessary for understanding the event are kept. The language used is as close as possible to the author's text. The following elements are removed: author's name, address and telephone number, the precise date and especially the day of the month (the year, month, time and any "weekend / holiday" context are kept), the exact place of occurrence (features having a bearing on the event are preserved), the aircraft registration or tail number, the aircraft type (in certain cases, aircraft characteristics having a bearing on the event are kept, notably if there is a possibility that human-factors are involved). The statutory and legal dispositions concerning these two undertakings have been carefully studied. It should also be pointed out that the reported events have generated no known consequences.

Proving what has been reported: The user is anxious not to become embroiled in a trying and troublesome procedure demanding a great deal of time and effort. If there is no obligation to declare an event, there is no reason to throw doubt on the author's report. The system is based on mutual confidence. Since no justification is requested, no additional documents need be provided.

The need for perfect understanding: The user submitting a report must have the absolute certainty that his initiative will be taken into account and understood by the recipient. This is why the REC specialist systematically will telephone the author after receiving his report. The purpose of the call is to: (1) Ensure that the author actually exists and that the submittal is not an anonymous letter (a system that relied on anonymous letters would have no credibility), (2) collect any additional information, (3) ascertain that the safety message that the author wanted to convey is properly understood, (3) consider what use should be made of the report, (4) finally, if the report does not lie within the field of interest of REC, to point the author towards the appropriate party. This, for example, would be the case when an accident declaration is sent by mistake to the REC. In no case does the specialist act in place of the author.

Potential penalties: The user fears being on the receiving edge of administrative or disciplinary penalties should infringement of regulations be brought to light. Article L 722.2 of the Civil Aviation Code protects a

person submitting a report in good faith. Lawmakers have foreseen the loopholes that may be left uncovered by such protection. The complete text of the related article, with some commentary, is reproduced in the appendix.

Second phase: processing forms and feeding the database: As stated further, reporting parties are guaranteed confidentiality and anonymity. Forms are nominative and they are treated confidentially. The information kept in the database is de-identified or rendered anonymous.

Third stage: distribution of return information: If during contact with the author or when examining the file in the database, it appears that a particular problem requires immediate correction, the de-identified information will be used to transmit an alert message to the concerned authority. In two cases concerning serious events, upon completion of report-processing, information kept in the REC was immediately forwarded to the administration responsible. The events in question concerned the detection of carbon monoxyde in a cabin and the use of transponder code 7000 for control organisations. Periodic database lookup is performed in order to publish reports of events seen as pertinent for safety and to produce subjects of reflection or study topics for task forces, as likely inputs for operational concept documents, recommendations or suggestions.
Every three months, relevant cards are extracted from the database. These are forwarded to the appropriate organisations. Events are grouped into three categories: (1) "for information", if it seems relevant to inform the administration as to the precise events, (2) "reported", if it seems likely that a measure ought to be considered, or that the event could be linked with others identified through other channels, (3) "input requested", if the REC wishes to receive comments from the administration.

Feedback to users: A bulletin - "**REC Info**" - constitutes the main feedback tool. This takes the form of a double A4 page, and contains some accounts of events (five or six) considered relevant to the field of accident prevention. If it were any larger, it would probably not be read in full. The yearly publishing level is ten issues. Circulation is 1200 copies, sent to all user groups, schools, aviation-related contractors, unions, newspapers etc. It is also available on the BEA website (see publications on http://www.bea-fr.org/rec). Narratives are very factual. The aim is not to give lessons in safety, because this might be taken poorly, seen as out of place or non-applicable. On the other hand, reader-stimulation may be generated by means of a commentary in the margin if a certain aspect appears complex upon reading the narrative. Receiving "REC Info", and user-interest in reading it, are the best means of ensuring the lasting reputation of the system, confirming its goals and working principles, and strengthening user-confidence. The bulletin constitutes the best advertisement for the system, and generates input in the form of reports.

Conclusion and Outlook

The characteristics of the REC correspond to the specifications proposed by international bodies. The system operates according to principles tried and tested in Anglo-Saxon countries employing related systems. In a climate of confidence, a confidential reporting system allows knowledge-enrichment for every actor in the aviation community, based on the experience of a few. It can also provide information for organisations responsible for supervising activities. The REC field of action above all comprises events associated with human factors that are not often reported by other means.

A useful system for general aviation: In a few months, the system has become known and appreciated. It is operating as forecast. The reports received are of great interest due to the quality of the lessons learned. REC forms one element of the feedback chain, one tool towards preventing accidents. Assessing its contribution in terms of safety improvements will be very difficult, even impossible to determine. It is nevertheless important that the system exist for general aviation. In effect, it forms a link between users and organisations, administrations and so on. This link operates not in the event of complaints or accidents, but by reporting unaccustomed situations or frequent minor events occurring in the course of aviation activity. The system is still fragile, and must be consolidated. Once the database reaches the required size, general operational reports on related events will be considered.

Adaptability of the system to public air transport: Public air transport is characterised by features quite different from general aviation: a safety coefficient around a hundred times higher, standards, regulations, procedures, heavy financial resources, an homogeneous population consisting exclusively of professionals,

voluntary and confidential event-compilation, often integral with airline internal feedback systems. Given such considerations, a system for the public air transport sector would differ appreciably from the REC system. Notwithstanding, experience acquired in the general aviation sector would remain of great use.

References

Amalberti, Barriquault, 1999. Foundations and limits of feedback. Annales des Ponts.

Civil Aviation Code, 1997. JAR-OPS-1. Joint Aviation Requirements - Commercial Public Air Transportation - Aircraft. Decree of 12 May 1997 concerning the technical conditions for aircraft operation by a public air transport company.

Civil Aviation Code, 1999. Law No. 99-243 of 29 March 1999 concerning the technical investigation of accidents and incidents in civil aviation. Book VII of the civil aviation code.

Rasmussen, 1997. Risk management in a dynamic society: a modelling problem. Safety Science 27 (2-3), 183-214.

"The simpler it seems, the more you have forgotten…" New Challenges in Investigation and Safety Management

Graham Braithwaite,

Dept. of Aviation, University of New South Wales, Sydney, 2052 Australia.
G.Braithwaite@unsw.edu.au http://www.aviaton.unsw.edu.au

Abstract: In aviation and other complex socio-technical systems, accident and incident investigation has advanced considerably in recent years. This is primarily a function of a broader acceptance of human factors issues and a general shift towards investigating higher-frequency, lower-consequence events such as incidents or 'normal accidents'. However, with what appears to be an asymptotically stable accident rate, the question remains whether such advances are translating into greater levels of overall safety?
Whilst concepts such as Reason's Organisational Accident Model (1990, 1997) have done much to explain why accidents occur, has the clarity with which safety case studies are now presented actually led to an oversimplified attitude towards prevention? Alternatively, are our efforts actually leading to the expenditure of 'excess safety' in the name of improved efficiency through the process theorised by Wilde (1994) as 'risk homeostasis'?
Investigators and researchers are challenged to question the effect of their work at a systemic level. To understand how and why accidents occur is of limited value unless they can be translated into effective prevention strategies that can be employed by the population at large.

Keywords: investigation, reporting, organizational accidents, risk homeostasis, education.

Introduction

Accident and incident investigation, whether classified as a science or an art form has advanced considerably in recent years, especially within the aviation industry. The ability of investigators to piece together the chain of events leading up to an accident is almost beyond belief. The loss of Pan Am flight 103 over Lockerbie in 1989 provides a clear example. From a wreckage trail over 180 miles long, the Air Accidents Investigation Branch (AAIB) were able to deduce not only the break-up sequence of the aircraft, but also the bomb's location on the aircraft, the case and container it was placed within, the clothes it was surrounded by, and even the type of radio cassette player it was concealed in.
However, the uncovering of such facts is of limited value unless used effectively for accident prevention. The assembly of experts at IRIA 2002 provides an opportunity to challenge some of the ways that investigators and researchers communicate their findings. The cutting edge of investigation and research becomes somewhat blunted if the findings are not interpreted and implemented in the way in which they were intended.

Common sense and brilliance

Reason's Organisational Accident Model (1990, 1997) has had a profound effect upon the investigation of incidents and accidents, particularly within the aviation industry. Adopted by the Australian Bureau of Air Safety Investigation (BASI) and its successor, the Australian Transport Safety Bureau (ATSB) as the core of its investigation methodology, the model has also been adopted by the International Civil Aviation Organisation (ICAO) and adapted into investigative analysis tools such as *BHP Billiton's* ICAM system. One of the strengths of the model is its clarity and the cultural shift it has helped facilitate in the way the industry considers accident causation.
However, whilst the model's value is unquestioned, its misapplication is becoming a concern, particularly for those involved in safety education. Indeed, it is the model's brilliance as a 'common sense' explanation that uncovers its weaknesses.
The organizational accident provides a conceptual framework that fits the majority of accidents involving complex, sociotechnical systems such as aviation. Accidents are demonstrated to be the consequence of a mixture of latent conditions, active failures and local conditions. The natural tendency to focus on a single

'primary' cause, and the desire for blame that is associated with it, are diminished by demonstrating the multiple defences that need to be breached in an accident.

As an investigation tool, the model provides some guidance as to the type of contributory factors that may lead to an accident. However, attempts to 'investigate by numbers' using the categories suggested by the model can have negative consequences. The investigation of the Ansett Australia B 747 that landed with its nose wheel retracted at Sydney in 1994 demonstrated that an overly rigid application of the model as an investigative framework becomes problematic. The model is, after all, just that; a model and not a self-contained accident investigation methodology.

In an increasingly litigious society, the (mis)use of Reason's model to shift blame higher up an organization is both dangerous and unwarranted. With the spectre of corporate manslaughter looming large since the 1987 loss of the *Herald of Free Enterprise* at Zeebrugge, the role of senior management in major accidents has become increasingly the focus of Coroners, lawyers and investigators alike. The delivery of a simplified version of the model on training programs such as CRM (crew resource management) can have the negative consequence of convincing subordinates that the model can shift responsibility up the chain and away from them. Similarly, senior managers are becoming increasingly wary of their own investigative processes incriminating them. Whilst the accident investigation, Coronial and legal processes are supposed to be separated, this is sometimes more a theoretical distinction than reality.

Applying the lessons

In teaching safety management and investigation at both undergraduate and postgraduate levels, the author has anecdotally observed a number of undesirable by-products of the model's clarity. These include a form of fatalism whereby individuals believe that the system will conspire to fail around them and there is very little they can do to change things; a form of complacency where the 'blame' has been shifted up to the CEO and Board level; and an oversimplification of the event that degrades the student's ability to recognize similar events in the future before they turn into incidents and accidents. It is the latter concept that is the focus of research currently being undertaken by a postgraduate student at the University of New South Wales. As the safety research community continues to find new ways to explain accident causation, it must be careful not to contribute to hindsight bias or oversimplification on the part of those it is trying to assist.

This does not suggest that the model is deficient. Rather it places a greater responsibility on academics, safety professionals and trainers to ensure that the model's original intent is preserved. In training operators to recognize failed defences prior to an accident occurring, it is important for them to know how subtle the warning signs may actually be. Those involved in major accidents rarely know what is about to happen until it is too late. For example, for the junior crewmembers on the flight deck of the *KLM* B 747 at Tenerife in 1977, there is no evidence that they did not believe that the *Pan Am* aircraft had actually cleared the runway. They did not *choose to die* rather than speak up. It is more likely that they reconciled in their own minds that their senior crewmember had heard a clearance that they had missed. In the vast majority of similar situations for junior crew, they would probably have been correct. As Byrne (2001) suggests, whilst accident causation models can shine like a torch to illuminate the breached defences in retrospect, looking forward can be nothing short of blinding.

Standing still whilst running

The aviation industry has suffered what appears to be an asymptotically stable accident rate since the early 1970s. In spite of major advances in technology (e.g. GPWS (Ground Proximity Warning System), ACAS (Airborne Collision Avoidance System)) and in training (e.g. LOFT (Line Oriented Flight Training), CRM (Crew Resource Management)), the industry has found itself at an accident rate that is predicted by Boeing to equate to one wide-body hull-loss per week by the year 2015, assuming continued traffic growth. The question for safety professionals to consider is whether the industry has reached the limits of acceptable risk. In other words, has aviation decided that it is 'safe enough'?

Wilde's theory of risk homeostasis (1994 etc.) is as difficult to test properly, as it is controversial. However, safety professionals ignore its potential at their peril. In simple terms, the theory suggests that an individual is prepared to accept a certain level of risk: Where increased safety measures lower the risk exposure, Wilde suggests that this utility may be spent on other things such as completing a task more quickly. For example, as a motorist becomes more familiar with a particular route, they may start to drive it more quickly – as their experience level increases, so their additional utility may be spent in the form of speed and hence their risk exposure is maintained.

The theory is aimed at individual risk taking, but could it also be applied at a systems level? In other words, have advances in safety led to trade offs in utility that return the overall system safety level to close to its original levels – a form of *systemic* homeostasis?

Consider two separate changes proposed under Australia's aviation regulatory review process (Braithwaite, 2001). The first involved the reduction of Aviation Rescue and Firefighting (ARFF) cover at several regional airports and the other involved the increase of cabin crew to passenger ratios from 1:36 to 1:50. Both changes were proposed independently.

The proposal to reduce ARFF cover drew upon research conducted at Cranfield University, which had found that assertive cabin crewmembers played a significant role in the efficiency of aircraft evacuations. Arguing that a lack of aircraft fires within Australia suggested a low risk and that the presence of cabin crew would expedite evacuations, the regulator proposed the removal of ARFF from a number of airports that served aircraft as large as Boeing 737. Meanwhile, another review proposed that the number of cabin crew required on RPT (Regular Public Transport) aircraft could be reduced so that there was only 1 per 50 passengers rather than 1 to 36. The justification was 'harmonisation' and, once again, an apparent lack of accidents. The overall effect was that with ARFF cover reduced, the responsibility on cabin crew was increased at a time where their number was being decreased. The irony was that the accident that led to the cabin safety research at Cranfield was the result of the 1985 *British Airtours* accident at Manchester. A Boeing 737 aircraft caught fire on take off at an airport with ARFF cover far exceeding the requirement for the aircraft type caused the deaths of 55 souls. Without ARFF, which commenced firefighting within 12 seconds of the aircraft coming to halt, none of the exits would have been cleared for evacuation by the cabin crew, assertive or otherwise.

The above example is specific, but not an isolated one. At a more general level, advances in aircraft avionics have equated to reductions in flight crew complements; the fitment of ACAS has led to the introduction of reduced vertical separation minima (RVSM) and the proposal for reduced climb separation in air traffic control; increases in engine reliability have led to extended twin engine operations (ETOPS) over water, and so on.

The challenge?

As the research community discovers more about the complexities of accident causation and safety management, there is a significant risk that the focus is on 'the same for less' rather than a concerted effort to reduce the accident rate. With aircraft as large as the Airbus-A380 only a few years away from production, the ability of the industry to cope with the loss of even a single 650+ seat aircraft is questionable. To lose aircraft of this size at the same rate as, for example, Boeing 707s or Douglas DC8s is unlikely to be acceptable. Titanic was, after all, only one hull loss for the shipping industry…

This paper neither proves the existence of systemic homeostasis nor does it prove that misapplication of Reason's Organisational Accident model has led to negative safety outcomes. Its purpose is simply to stimulate discussion on new theories. Wood and Sweginnis (1995) of the Southern California Safety Institute suggest that there are three basic attributes that describe all good investigators.

1. They are not afraid to be wrong. They will accept facts that are contrary to their present theory;
2. They readily admit that they don't know everything. When they need help, they seek help;
3. They listen to other investigators. They don't necessarily believe them, but they do listen to them.

Hopefully this workshop will demonstrate those principles.

References

Braithwaite, G. R. (2001) Aviation Rescue and Firefighting in Australia – Is it protecting the Customer? Journal of Air Transport Management 7 (2001) 111-118.

Byrne, J (2001) The Use of the Reason Model in Accident Investigation. Unpublished research thesis. UNSW, Sydney.

Reason, J (1990) Human Error. Cambridge University Press, Cambridge.

Reason, J (1997) Managing the Risks of Organizational Accidents. Ashgate Publishing, Aldershot.

Wilde, G. S. (1994) Target Risk. PDE Publications, Toronto.

Wood, R. H. and Sweginnis, R. W. (1995) Aircraft Accident Investigation. Endeavor Books, Casper.

The Critical Incident Analysis Tool: Facilitating to Find Underlying Causes of Critical Incidents in Anaesthesiology for Novices in Human Error

Marcus Rall, Haible T, Dieckmann P, Zieger J, Schaedle B

Center for Patient Safety and Simulation
Dept. of Anaesthesiology (Chairman: K. Unertl), University Hospital Tuebingen, Germany
marcus.rall@med.uni-tuebingen.de

Keywords: human error**,** patient safety, critical incident analysis, human factor, safety culture

Short abstract:
Errors in medicine are among the leading causes of death and therefore play a major role in patient safety(Kohn et al., 1999). Experts in the field of patient safety and human factors can identify most of the underlying problems leading to patient harm(Cooper and Gaba, 1989). The same applies for the analysis of incidents and accidents(Reason, 1994; Maurino et al., 1995; Reason, 1997). Experts are able to analyse most of the factors contributing to a medical disaster if they are provided with reliable and deep insight into the accident situation. They also need time to question involved people in order to finish the puzzle of how and why it happened. But the problem we face in medicine is a very firmly established culture of blame. Right now experts will not be allowed to analyse incidents, to ask questions or access material. They will also not hear the truth.
The new Critical-Incident-Analysis (CIA) Tool wants to bridge this gap in anaesthesia. The prototype CIA-Tool is designed to be used by individuals for their own purpose. The prototype version is paper-based but is intended to become an interactive computer-based application. Through the Tool normal anaesthesiologists will be asked questions they would not have thought of, because they have no special training in human error, human factors or the systematic analysis of incidents. The Tool tries to address topics from well known human error and safety textbooks, papers and resources.
It is hoped that by using the CIA-Tool anaesthesiologists discover causes and dangerous hot spots in their medical system which would have gone undetected without the Tool. By using the Tool on a regular basis it is expected that they like it more and more, learn more about accident chains and start to foster an open atmosphere to talk about mishaps and error. This would be the foundation stone for a proactive safety culture(Gaba, 2000).

Background:
Since the “To err is human”-report of the Institute of Medicine at the latest, it is known that errors in medicine are a leading cause of death in hospitalized patients. The reasons for most of these errors are human factors, often combined in an accident chain. A thorough knowledge of accident evolution and human factors seems to be necessary for a meaningful analysis of critical medical incidents. Due to the culture of blame in medicine, the analysis of incidents has no tradition. There are almost no official reporting systems. A general medical doctor can therefore not be expected to have expertise in this area. This means that doctors who try to do an incident analysis by themselves might only be able to find out the errors which harmed the patient, but not the many latent errors and underlying conditions which lead to the error. The “Critical Incident Analysis Tool” is designed to help doctors without special training in accident causation to elucidate human errors and to track the accident chain up to the latent conditions and organizational factors. Existing reporting systems for medical devices have a different focus. The Swiss CIRS (www.cirsmedical.org) aims for a national or even international reporting system with statistical analyses and is not as detailed as the CIA-Tool. Up to now, no such instrument as the CIA-Tool is available.

Method:
We developed a questionnaire-based tool for medical personnel to analyse an incident on their own. It is based on theoretical human error and human factors knowledge and on experiences from our realistic

patient simulator environment. It was also tried to incorporate published causes of medical incidents. We present a pre-beta-version on paper. The final version is intended to be computer-based and interactive. The CIA-Tool at this stage is neither intended to be used by groups or for conferences nor is it designed to be entered in a database. It is meant to be a personal critical incident analysis help and guide. Of course this does not preclude that it may be used in group discussions at institutions with a good positive error culture.

Results:
The "C.I.A.-Tool" consists of some introductory pages, followed by a timeline diagram, facilitating the representation of the crisis evolution and acting as an overview for the analysis. The user is asked to mark some milestones on this timeline and take the sheet out of the Tool for a better orientation during the following analysis. There is also a separate page with an evaluation matrix, where the user can mark important or relevant findings. The main part is a questionnaire covering most of the human factor aspects of incidents known to the authors (Vincent, 2001). There are sections on general aspects of the work conditions and structure of the organisation. There are sections dealing with the actual measures and countermeasures. These are divided in areas like human factors (individual, team, task, organisational)(Reason, 1997; Reason, 2000) and technical man-machine issues. There are also sections on the recovery of the incident and on what was done exceptionally well. Especially the last items seem very important to the authors.
After finishing the Tool the user is asked to look at the evaluation matrix and think about connections and chains of the entered items. This is thought to facilitate the recognition of interacting causes. Then the user may rate the importance of the different findings using his professional knowledge and his in depth view into his organisation. He may then choose to speak with the chairman, his colleagues, nurses or whoever he thinks is appropriate. Of course he may also choose to keep silence.

Discussion:
Ideally errors in medicine should be investigated by an interdisciplinary team of professional incident specialists including all involved participants. This is well practiced in other domains like commercial aviation or nuclear power industry (Helmreich, 2000). In medicine we are faced with a culture of blame and almost no systematic error analysis instruments. In this context the C.I.A.-Tool is our attempt to enhance the analysis of critical incidents in anaesthesia until more sophisticated methods are implemented. We are aware of the fact that it will not be possible with our tool to nearly find out all contributing factors. Depending on the role during the incident a lot of information will not be available for the lonesome user of the Tool. Some information will not only be incomplete but also partly erroneous. Therefore most probably only a minor part of the factors leading to the incident will be elucidated. But the Tool may be helpful to find at least some systematic and human factor causes. It is hoped that even these little drops will steadily contribute to enhance patient safety, prevent some patients from harm and lead to a gradual promotion of safety culture in an organisation. Then the way would be paved for the establishment of large scale incident reporting systems with the active support of the users and stakeholders.

Conclusion:
The CIA-Tool is a pre-beta-version of a personal human factor and critical incident analysis tool. The authors hope for the feedback of the scientific community to refine the tool. Experience from first studies evaluating the CIA-Tool will reveal whether the expected advantages can be realised and whether it is worthwhile to develop a more easy-to-use computer-based version. From a theoretical point of view the innovative CIA-Tool seems promising to slowly enhance the safety culture and rapidly safe lifes.

Literatur:
1. Cooper JB, Gaba DM (1989), A strategy for preventing anesthesia accidents. *Int Anesthesiol Clin* 27:148-152
2. Gaba DM (2000), Anaesthesiology as a model for patient safety in health care. *BMJ* 320:785-788
3. Helmreich RL (2000), On error management: lessons from aviation. *BMJ* 320:781-785
4. Kohn LT, Corrigan JM, Donaldson MS (1999), *To Err is Human - Building a Safer Health System.* Washington: National Academy Press
5. Maurino de, Reason J, Johnston N, Lee rb (1995), *Beyond Aviation Human Factors.* Aldershot: Ashgate
6. Reason J (1994), *Human error*. Cambridge:

7. Reason J (1997), *Managing the Risks of Organizational Accidents*. Aldershot: Ashgate
8. Reason J (2000), Human error: models and management. *BMJ* 320:768-770
9. Vincent C (2001), *Clinical Risk Management*. London: BMJ Books

Using Web Site Synthesis in an Experiment on Causal Perception of Aviation Accidents

Siu-wai Leung[1], Dave Robertson[1], John Lee[1], Chris Johnson[2]
[1] Informatics, University of Edinburgh
[2] Computer Science, University of Glasgow

Abstract

An obvious way of presenting incident and accident information to the public is via Web sites. There is, however, little understanding of how best to present this sort of material. We do not even know how sensitive people might be to differences in the way we construct our Web sites. A key issue is the degree to which those viewing an air accident Web site believe that certain events cause accidents. There may be certain styles of presentation or navigation within and between web-based accident reports that can either hinder or support the readers' ability to interpret evidence about previous failures. This, in turn, can undermine the argument that investigators present to support the causes that they distinguish in their report. The psychological literature offers various models of causal perception. If a model could be found which (even roughly) predicted strength of causal perception for incident/accident reporting Web sites then this would help us design sites which are more likely to give the perceptions of causality which we intend. The problem is that the available predictive models do not give similar predictions. We describe an experiment investigating the predictive power of two models of causal perception applied to accident reporting Web sites. Our experiment is novel in its use of automated synthesis to construct experimental Web sites which are guaranteed to have the same content but which vary in Web site structure according to a number of key parameters. This allows us quickly to construct experiments involving large but closely comparable Web sites.

Introduction

The main part of this paper describes an experiment on causal perception based on Web site browsing. Our aim is to identify the extent to which existing psychological theories (distilled into simple predictive models) can account for the perceptions reported by people browsing accident reports on web sites. Section 3 describes how the experiment was conducted. Section 4 then summarises our initial results and Section 5 discusses their significance. These sections follow a familiar pattern of empirical experimentation. The way in which we set up the experiment, however is unusual because it involved automated synthesis to produce the Web sites describing accident information. Section 2 describes this approach to Web site experimentation.

Web Site Synthesis

One of the problems in conducting experiments on Web sites is in controlling the structure of the sites. To be realistic in domains such as accident analysis, each site must be large and complex. If our aim is to perform a comparative analysis between different Web sites which present the same content material then we must be confident that the sites we build do all present the same content (not less, not more). We also may want the differences in presentation between different sites to be uniform. Ideally, we should be able to describe how the entire site design relates to the content we are presenting so that the differences in decisions made during the design of each site are explicit. If our experiment needs numerous such Web sites then there is the additional practical issue of constructing large numbers of complex Web sites quickly and reliably. We use automated synthesis to tackle this problem. Our basic means of synthesising Web sites, and its broader applicability, is described in [1,2,6]. In what follows we summarise the features pertinent to our experiment.

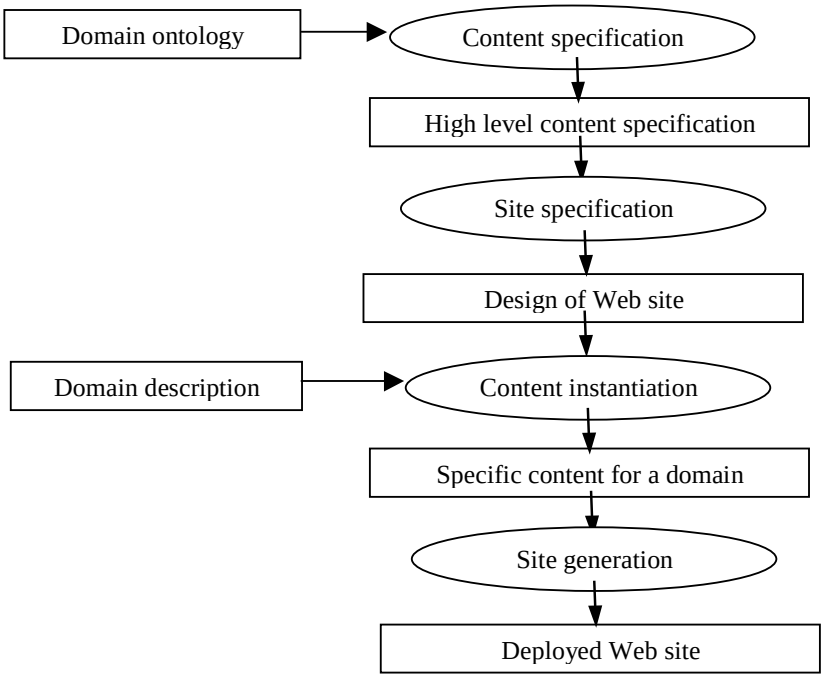


Figure 1: Overview of synthesis method

Figure 1 describes the stages involved in our form of automated synthesis. Two kinds of information must be supplied from the domain to which synthesis is being applied. First we must give the ontology used in describing the domain and the description of the domain itself. For the purpose of example let us assume that this consists of simple type declarations for the terms used to describe the domain, although in practice our ontology is more complex than this. An example from the accident reporting domain is:

position	:	$aircraft \times time \rightarrow altitude \circ distance$
aircraft	:	*nominal*
altitude	:	*ratio*
distance	:	*ratio*
time	:	*interval*

which says that the position of an aircraft gives its altitude and distance from the runway, with aircraft being a nominal (categorical) term while altitude and distance from runway are ratio terms. The types of time and position are not limited to the exemplified. For example, the time can be real time used in the ATM and recorders or a complex relative model of time used in maritime incidents. The values of position can either be longitude, latitude and altitude or navigational waypoints and flight levels.
This typing information can be used to derive a high level specification of the content which the Web site should present (depicted by the process at the top of Figure 1). This is described by composing the functions introduced in the ontology. For example, we might want our Web site to describe the positions of aircraft over time, in which case our high level content specification would contain the expression:

$\lambda T{:}time.position(aircraft,T)$

Given a high level content specification, the next step is to specify a Web site at a high level by committing to particular visualisations of the content specification. For instance, one way of viewing the position of aircraft over time is as a table (with rows as times and columns as altitudes for positions); another is as a graph (with X-axis as time, Y-axis as altitude position, and Z-axis as the distance from runway). We know these are options because of the typing of the expressions (graphs and tables are ways of describing the variation of ordinal or ratio variables and we know time and altitude are of this type). The choice of high level site specification is not uniquely defined, in general, but we can constrain it either through interaction (someone chooses the "appropriate" ones during synthesis) or by constraining the synthesiser (someone thinks ahead about what the appropriate choices are). For our running example we might choose a table view of aircraft position:

table(λT:time.position(aircraft,T))

Given a high level site specification, the next step is to instantiate this by supplying the actual content for our domain. In our running example, this would include the particular positions for particular aircraft in our table. This information can either be supplied directly or it can itself be synthesised from other descriptions. As we shall see later, one option is that it can be synthesised from generic accident descriptions - thus allowing us to generate content for experiments on causal perception. Supposing we have only three time points at which altitude and distance from runway were known then the pairs of values for the instantiated table providing position information in terms of time, mean sea level, and nautical miles might be:

table(λT:time.position(aircraft,T),[(10.15,1023,13.5),(10.42,876,9.25),(10.49,833,6.25)])

The last synthesis step is to construct the "physical" site from the instantiated specification. This is done by passing the specification through a "compiler" which enforces a particular style of description for the visual terms of the specification (such as tables) and produces as output the HTML pages for the site. Our choice of HTML is because it is standard but a similar solution applies for other Web-oriented languages such as XML.

Experiment on Causal Perception

We have used the form of synthesis described in the previous section to provide the raw material for a study of causal perception in accident reporting. The basic idea is to use typical patterns of aviation accident to instantiate the domain description of Figure 1. We perform the instantiation by selecting from our set of patterns and generating a permutation of it (substituting particular aircraft names and accident events, plus allowing a reordering of certain event types). Thus a single pattern can generate many possible accident incidents. We keep constant the other stages of synthesis, restricting them to be fully automatic, so that we can generate many different Web sites from the same accident patterns. This is how we generate the experimental material: Web sites, each generated from the same basic accident patterns but each site different and each containing 100 accident cases. The rest of this section describes an experiment based on this material.

Hypotheses/Key Findings

We are, in general, interested in studying how the presentation of accident reporting information on Web sites might influence people's perception of the causes of accidents. More specifically, we would like to see if any of the two well known general models of causal perception (contingency theory and PowerPC theory) is a better fit to observed perceptions of accident causes in those browsing our synthesised Web sites. If so, we would like to assess the sensitivity of this goodness of fit to the form of "rhetoric" used in the Web site. We shall consider two different forms of rhetoric: a strongly causal rhetoric in which the causal links between events (known from the accident patterns used to generate the sites) are emphasised as hyperlinks between events; and a temporal rhetoric in which the sequence of events for each accident is shown and causal links are mentioned but not emphasised. To reduce the effect of subjects using their knowledge of aviation accidents we generate sites for some of the experiments using event descriptions

meaningful to aviation (such as “speedbrake extended”) and for other experiments we generate the same causal description but with cryptic names unrecognisable in the aviation domain.
The two theories of causal perception under scrutiny are, first, the contingency theory (see for example [3]) which predicts a measure, ΔP, of the extent to which a candidate cause, c, and an effect, e, are perceived to covary according to the equation:

$$\Delta P = P(e|c) - P(e|\neg c) \qquad (1)$$

where $P(e|c)$ is the probability of e given that c occurs and $P(e|\neg c)$ is the probability of e given that c does not occur. ΔP is often called contingency or contrast.

Our second contender is a probabilistic contrast (PC) model - the PowerPC model (compared to the contingency theory in [4,5]) - then is defined as follows:

$$Power = \Delta P / [1 - P(e|\neg c)] \qquad (2)$$

where *Power* is the generative power of c with respect to e. ΔP is as defined in equation (1). These two models give different predictions of the causal strength associated with a cause and an event. For instance, if the probability of an event given that a cause occurs ($P(e|c)$) is 0.9 and the probability of the event given that the cause does not occur ($P(e|\neg c)$) is 0.8 then contingency theory predicts the strength of causation to be 0.1 while PowerPC predicts 0.5. By setting up Web sites describing sets of accidents with known $P(e|c)$ and $P(e|\neg c)$ we can assess the predictive power of these two models by comparing their predictions to those observed from human subjects who have browsed those sites. The hypotheses we explore in the experiments described below are:

1. Causal perception of aviation accidents follows the PowerPC theory of causal perception, not contingency theory (ΔP).

2. Causal perception of accidents gives similar patterns even when the information is presented in different styles.

3. Normal causal ratings are different from counterfactual causal ratings.

Methods

Participants
The participants were 47 undergraduate students of the University of Edinburgh, taking the Artificial Intelligence 1 course as one of the three constituents of their university year. The experiment was run during the experimental methodology module of the course. Participation in the experiment was not compulsory. The course attracts students from a broad variety of backgrounds (it may be taken by those enrolled in science, engineering, psychology and linguistics degrees) but training earlier in the year ensures that they all have basic computer skills and are familiar with Web browsing.

Tasks
Each participant filled in a pre-experiment questionnaire which collected background information such as their familiarity with the Web and with aviation operations. Then the participant browsed a synthesised Web site which was randomly assigned to the participant. Each Web site presented 100 cases of aviation accidents and incidents. After browsing the Web site, the participant filled in a post-experiment questionnaire in which he or she was asked to give causal ratings of some candidate causes to be the contributing causes of the aviation accidents.

Dependent Variables
The dependent variables were normal causal rating and counterfactual causal rating. The values of normal causal rating and counterfactual causal rating were expressed on a scale between 0 and 100. The wording of

the question to probe normal causal rating was as follows, where *AccidentCategory* describes a type of accident and *EventName* describes a type of event:

"Out of each 100 flights without an accident [*AccidentCategory*], how many might have had an accident [*AccidentCategory*] if [*EventName*] had occurred?"

The wording of the question to probe the counterfactual causal rating was as follows:

"Out of each 100 flights with an accident [*AccidentCategory*], how many might not have had an accident [*AccidentCategory*] if [*EventName*] did not occur?"

The participants were also asked to rate their confidence in giving their normal causal rating and counterfactual causal rating. These confidence ratings were acquired using a 7-point scale.

Independent Variables
The independent variables are positive covariation frequency $P(ac|ca)$ and negative covariation frequency $P(ac|\neg ca)$, where *ac* is an accident and *ca* is each candidate cause. The corresponding versions of the ΔP and *PowerPC* models (see earlier equations 1 and 2) are as below:

$$\Delta P = P(ac|ca) - P(ac|\neg ca) \quad (3)$$

$$Power = \Delta P / [1 - P(ac|\neg ca)] \quad (4)$$

Other independent variables are rhetoric and terminology. The only two possible choices of rhetoric are *causal* and *temporal*. Using causal rhetoric ensures that the events in each aviation incident described by a site are arranged using explicit causal links, *i.e.* an event is caused by another event and may lead to zero or more other events (Figure 2). Using temporal rhetoric ensures that the events in each aviation incident described by a site are arranged simply according to their temporal sequence, *i.e.* two adjacent events on display may not have causal relationship (Figure 2). The two possible choices of terminology are *sensible* and *cryptic*. Choosing sensible terminology ensures that each event is described using an English phrase which normally used to describe such an event when reporting an incident (Figure 2). Choosing cryptic terminology ensures that the events are described using invented names which have no meaning as far as aviation incidents are concerned (Figure 3). The cryptic terms were used as a check that our human subjects were not using pre-conceptions of terminology to influence their causal judgement.

Experiment Design
Each synthesised Web site presented the aviation accident information in one of the following styles:

- Multiple pages per case with hyperlinks indicating causal relations of the events (causal rhetoric) and description using common terminology (sensible terminology).

- Multiple pages per case with hyperlinks indicating causal relations of the events (causal rhetoric) and description using cryptic terminology (cryptic terminology).

- Single page per case with the events appearing in temporal sequence (temporal rhetoric) and description using common terminology (sensible terminology).

- Single page per case with the events appearing in temporal sequence (temporal rhetoric) and description using cryptic terminology (cryptic terminology).

Each synthesised Web site presented five patterns of aviation accidents with five different combinations of covariations between candidate causes (ca) and accidents (ac). For each of these, we can calculate the values of $P(ac|ca)$ and $P(ac|\neg ca)$, thus allowing us also to calculate the predictions of the ΔP and *PowerPC* models for each combination. These calculations are shown in the table below. We shall return to this in the analysis of results in Section 4.2.

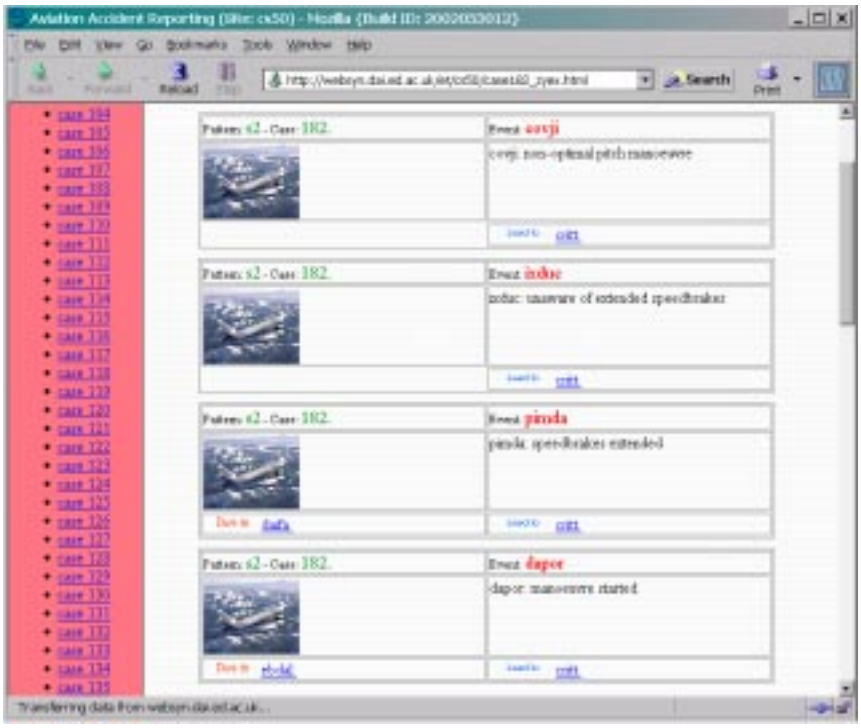


Figure 2: A Web site with causal rhetoric and sensible terminology

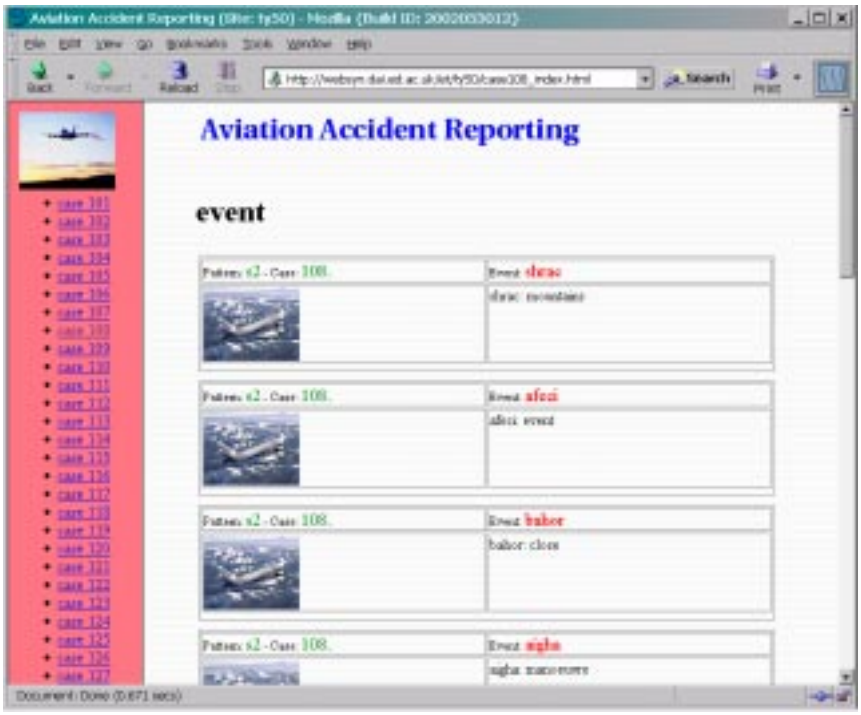


Figure 3: A Web site with temporal rhetoric and cryptic terminology

Combination	$P(ac\|ca)$	$P(ac\|\neg ca)$	ΔP	*PowerPC*
A	1.0	0.5	0.5	1.0
B	0.9	0.8	0.1	0.5
C	0.8	0.5	0.3	0.6
D	0.7	0.5	0.2	0.4
E	0.6	0.2	0.4	0.5

Web Site Synthesis
The Web sites in this experiments were synthesised via the automated process described in Section 2. The synthesiser could be parameterised by the rhetorical and terminological style (see Section 3.2.4) and could also be set to generate a sample of accident cases with the combinations of covariations shown in the table of the previous section. These samples were generated from standard patterns of aviation accident culled from the literature.
Five patterns of aviation accident information content with the same causal tree structure were randomly assigned to the five combinations of covariation of candidate causes and accidents. This avoided the influence of specific accident information content on the causal perception test. The diagram of Figure 4 shows an example pattern of aviation accident information.
This aviation accident information was used to synthesise the experimental Web sites as described in Section 2. As an additional measure, the hyperlinks of each synthesised Web site were checked for errors such as dead links, raising our confidence (although of course not guaranteeing) that our synthesiser had worked reliably. Every Web site was unique in terms of the settings of rhetoric, terminology, and the order of presenting cases.

Procedure
All participants began by filling out a Web pre-experiment questionnaire collecting their personal information such as year of birth, programme of study, levels of computer and Web skills, current usage of computers and the Web, and familiarity of aviation operations. Subsequently, each participant login to the Web site which was randomly assigned to him/her to browse. Each participant was also given instructions concerning how to browse the synthesised Web sites, particularly how to interpret the provided navigation aids. The browsing task took 45 minutes. On the completion of the browsing tasks, participants filled in a post-experiment questionnaire on the Web to give their normal causal ratings and counterfactual causal ratings.

Data Analysis
Independent variables are (1) the combinations of covariation of candidate causes and accidents (within subjects), (2) rhetoric (causal or temporal), and (3) terminology (sensible or cryptic). Dependent variables are (1) normal causal rating, (2) confidence of giving the causal rating, (3) counterfactual causal rating, and (4) confidence of giving the counterfactual causal rating. The data were analysed using analysis of variance (ANOVA) and linear regression by R statistical software.

Results
We now summarise the preliminary results obtained from the experiment described above, prior to a discussion of them in Section 5.

Pre-experiment Questionnaire
Our pre-experiment questionnaire was intended to detect bias which might have come from some imbalance in the social or technical background of the subjects. We saw no obvious difference in any category of the background of the participants randomly assigned to browse the Web sites with different rhetorics and terminology.

Post-experiment Questionnaire

Our post-experiment questionnaire asked for estimates of normal causal rating and counterfactual causal rating, via standard questions for each site (see Section 3.2.3). Our preliminary analysis of the answers to these questions is as follows:

- There was a highly significant difference (P=1.574×10^{-9}) in the normal causal ratings among five covariation combinations of the aviation accidents regardless of the rhetoric and terminology.

- There was a moderately significant difference (P=0.029) in the counterfactual causal ratings among five covariation combinations of the aviation accidents regardless of the rhetoric and terminology.

- There was no significant difference in normal and counterfactual causal ratings among five convariation combinations of the aviation accidents in four different groups with different rhetoric and terminology. (Sample size was too small to draw any conclusion that rhetoric and terminology have no/little effect on causal perception. Probably the effect of rhetoric and terminology were less obvious than the covariation of candidate causes and accidents.)

- The medians of normal causal ratings were well predicted by the PowerPC theory of causal perception. The table below shows, in the first two columns, the causal power predicted by the ΔP and *PowerPC* theories for each of the five covariation cases described in the table of Section 3.2.5. Compare these to the median and mean values for these cases observed in our experiments, shown in columns three and four of the table below.

ΔP	*PowerPC*	Median	Mean	SD
0.5	1.0	0.9	0.72	0.29
0.1	0.5	0.5	0.56	0.28
0.3	0.6	0.6	0.57	0.20
0.2	0.4	0.4	0.48	0.21
0.4	0.5	0.5	0.48	0.16

- There was a significant difference between the root-mean-square distance (RMSD) of the normal causal ratings to the theoretical ΔP and the RMSD of normal causal ratings to the theoretical *PowerPC*. This was also true in terms of the actual values and ranks of the normal causal ratings. This indicates that the normal causal ratings should be closer to either ΔP or *PowerPC*.

- The mean RMSD between the normal causal ratings and theoretical *PowerPC* was less than the mean RMSD between the normal causal ratings and theoretical ΔP. This was true in terms of both the actual values and ranks of normal causal ratings. This indicates that the normal causal ratings are closer to *PowerPC* than ΔP.

- There was a significant difference between the normal causal ratings and counterfactual causal ratings. The normal causal ratings and counterfactual causal ratings were not correlated well (R = 0.26) with each other in linear regression analysis. This indicates that the normal causal ratings and counterfactual causal ratings are not equivalent in causal power in this experiment.

- Although the median counterfactual causal ratings look closer to ΔP, the significance was not high enough to support this. In addition, the ΔP did not give a more accurate prediction of the ranking of counterfactual causal ratings than *PowerPC* (P=0.263).

Discussion

The results reported above are from the first experiment we have conducted of this kind and our conclusions are thus preliminary. In particular, the number of subjects (47) is lower than we need for convincing experimentation, and without further experiments on similarly constructed Web sites it is

impossible to feel confident that our results generalise. Nevertheless, the preliminary analysis yields interesting and perhaps surprising results:

The normal causal ratings in this experiment were more accurately predicted by PowerPC theory than the contingency theory (ΔP) of causal perception. The result of this experiment supports the PowerPC theory more than the contingency theory.

The covariation of occurrence frequencies between candidate causes and accidents had greater influence on the causal perception of the participants than the causal/temporal rhetoric and sensible/cryptic terminology.

The counterfactual causal ratings were significantly lower than the normal causal ratings. This indicates that there is a difference between counterfactual causal ratings. If the questions could accurately probe the counterfactual causal ratings in this experiment, it would be interesting to investigate the human perception of counterfactuals and to evaluate the effectiveness of counterfactuals in causal reasoning in other situations.

Neither the contingency theory nor the PowerPC theory accurately predicted the counterfactual causal ratings or their ranking. As PowerPC theory predicted the normal causal ratings well, its failure in predicting the counterfactual causal ratings supports that the counterfactual causal perception in this experiment is different from normal causal perception.

With the small sample size of this experiment, we cannot rule out any effect of the causal/temporal rhetoric and sensible/cryptic terminology on the causal perception even though no significance was found. We believe that the rhetoric and terminology can provide useful information to influence causal perception or reasoning. As seen in this experiment, their effects are less obvious than the covariation frequencies. Whether the causal/temporal rhetoric and sensible/cryptic terminology play roles in causal perception is to be confirmed in a larger scale experiment.

Further Work

The roles of causal/temporal rhetoric and sensible/cryptic terminology will be further tested in a larger scale experiment. If they can affect causal perception, they should be considered in generating accident reports on the Web.

The Web sites in this experiment did not include all events described in full accident reports. Once some factors affecting the causal perception have been found, further experiments presenting greater numbers of events may be performed. The causal perception of simulation video of the accident events should also be tested.

The relationship between normal causal ratings and counterfactual causal ratings should be delineated. In case the counterfactual causal ratings are a separate concept from the normal causal ratings in causal perception, further cognitive research to formulate a more appropriate theory will be valuable because counterfactuals are widely used in causal reasoning.

Acknowledgements

This work has been supported by EPSRC grant GR/M98302 for research on Communicating Knowledge about Accidents from Synthesized Web Sites.

References

[1] J. Cavalcanti and D. Robertson. Synthesis of Web sites from higher level descriptions. *3rd Workshop on Web Engineering*, May 2000, Amsterdam, The Netherlands.

[2] J. Cavalcanti and D. Robertson. Web site properties using computational logic. In *Information Modeling for Internet Applications*. Idea Group Publishing, 2002.

[3] P. W. Cheng and L. R. Novick. A probabilistic contrast model of causal induction. *Journal of Personality and Social Psychology*, 58: 545-567, 1990.

[4] C. N. Glymour. *The mind's arrows: Bayes nets and graphical models in psychology*. MIT Press, 2001. ISBN 0262072203.

[5] Y. Lien and P. W. Cheng. Distinguishing genuine from spurious causes: A coherence hypothesis. *Cognitive Psychology*, 40: 87-137, 2000.
[6] D. Robertson and J. Agusti. *Software Blueprints: Lightweight Uses of Logic in Conceptual Modelling*. Addison Wesley/ACM Press, 1999. ISBN 0201398192.

Figure 4: Example aviation accident pattern