ELSEVIER

Applied Ergonomics 33 (2002) 319–336

APPLIED ERGONOMICS

www.elsevier.com/locate/apergo

Development and application of a human error identification tool for air traffic control

Steven T. Shorrock[a,*,1], Barry Kirwan[b,1]

[a] Det Norske Veritas (DNV), Highbank House, Exchange Street, Stockport, Cheshire SK3 0ET, UK
[b] EUROCONTROL Experimental Centre, BP15, F91222, Bretigny Sur Orge, France

Received 10 January 2000; accepted 11 February 2002

Abstract

This paper outlines a human error identification (HEI) technique called TRACEr—technique for the retrospective and predictive analysis of cognitive errors in air traffic control (ATC). The paper firstly considers the need for an HEI tool in ATC, and key requirements for the technique are noted. The technique, which comprises a number of inter-related taxonomies, based around a simple cognitive framework, is then described. A study concerning a real-world application of TRACEr is outlined—the evaluation of several options for reduced separation minima in unregulated UK airspace. In this study, TRACEr was used predictively and retrospectively, looking forward to pre-empt potential problems and looking back to learn from experience. The paper concludes that TRACEr is a valuable aid to design, development and operations in UK ATC, and has indeed been used as a basis for further applications in ATC both in Europe and the USA.

Keywords: Human error; Air traffic control; Human error identification; Incident analysis

1. Why air traffic control needs a human error identification tool

Air transport is seeing increasing growth year by year, with passenger air transport, in particular, becoming more affordable and feasible for both short journeys and long haul flights. This has resulted in a growth in air traffic movements by 6–7% per year in the UK, which is ultimately reflected in an increase in workload for air traffic controllers. Thankfully, there have been few mid-air collisions in controlled airspace in the world's aviation history. Indeed, Richard Profit, Group Director Safety Regulation, UK Civil Aviation Authority, has stated that the fatal accident record for UK public transport operations generally is four times better than the world average, with a flight safety record among the best in the world (Profit, 2001).

Nevertheless, many more near-misses have occurred wherein aircraft have come closer than their required separation distances (e.g. 5 nautical miles laterally and 1000 ft vertically in UK en route airspace), sometimes by as little as 100 ft. A near-miss occurred at Heathrow Airport on 29 April 2000 where a British Airways Boeing 747-436 was instructed to go-around at a late stage of its approach, descending to 118 ft radio height above the runway, whilst a British Midland Airbus A321 was still on the runway for departure, with a tail fin height of 38 ft 7 in (Air Accidents Investigations Branch, 2001). This was characterised by air traffic control (ATC) errors in planning and decision-making, as well as problems with on-the-job training arrangements.

Despite the presence of automated safety nets, ATC is heavily dependent upon the capabilities of the human, and some ATC-relevant accidents were characterised by 'human errors', with underlying failures in safety management. Tragic examples include the 1977 Tenerife runway collision of the Pan AM Boeing 747 and the KLM Boeing 747, which killed 583 people, and the 1996 mid-air collision involving a Saudia 747-100 and Kazakstan Airlines IL76 over Dadri, India, with the

*Corresponding author. Tel.: +44-161-477-3818; fax: +44-161-477-3819.
E-mail address: steven.shorrock@dnv.com (S.T. Shorrock).
[1] Formerly with National Air Traffic Services Ltd., ATMDC, Bournemouth Airport, Christchurch, Dorset BH23 6DF, UK.

0003-6870/02/$ - see front matter
PII: S0003-6870(02)00010-8

loss of 349 lives. With air traffic density increasing, it is vital that the ATC community learns both from these catastrophes, as well as from the many more occurrences at the 'bottom of the accident triangle'; the near misses and unreported errors. The investigation, analysis and classification of human error offers perhaps one of the best ways forward for learning from such near misses so that accidents remain rare events.

As in all industries, human errors in ATC occur in several different forms, as evidenced in the causal factors of UK Airprox (Aircraft Proximity) incident reports (UK Airprox Board, 2000). Some examples are shown below.

(i) Airprox 221/99 (14 December 1999). The TC BIG SC [sector controller] did not detect an instruction by his trainee, which put both aircraft at the same level without standard separation.
(ii) Airprox 200/99 (1 November 1999). The LATCC NS SC did not take the subject B767 into account when he descended the B737.
(iii) Airprox 164/99 (8 September 1999). The Pennine Radar controller descended the BAe146 into conflict with the Tornado F3.
(iv) Airprox 152/99 (30 July 1999). Following a distracting telephone call, the Luton APR did not ensure standard separation between the subject aircraft.

These causal factors illustrate a variety of controller errors, involving perception, memory, decision-making, communication and team resource management (TRM). Classifying errors in a meaningful way is therefore essential to record such data in a way amenable to the detection of trends in incident occurrence, or in identifying different ways in which the system could fail. Put simply, error analysis is an essential component of safety management.

Approaches for error classification, typically termed human error identification (HEI), have been developed for the past 20 years, primarily in the process industries. These include SHERPA (Embrey, 1986), GEMS (Reason, 1990), CREAM (Hollnagel, 1998), and HEIST (Kirwan, 1994). Many HEI techniques have been influenced heavily by Rasmussen et al. 's (1981) Skill-, Rule-, and Knowledge-based (SRK) behaviour framework and Reason's (1990) classification of slips, lapses, mistakes and violations (or a combination of both). Whilst great headway has been made in this area, the available techniques have, in fact, had considerably less real use than might be expected considering the volume of work involved in their development (see Lucas, 2001). Indeed, Johnson (1999) asserted that human reliability approaches have had little impact upon many industries, largely due to the failure of human factors research seriously to consider the problems of systems development. According to Johnson, until practical problems are addressed, increasingly esoteric models of cognitive and organisational failure will be of little practical benefit. Such 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. Whilst Johnson's main assertion is debatable, the fact remains that the transfer of this technology to the design and operation of safety-critical, interactive systems has encountered serious problems.

Nonetheless, HEI has been applied to some new industrial sectors, such as manufacturing (Paz Barroso and Wilson, 2000), rail (Vanderhaegen, 2001), consumer products (Baber and Stanton, 1994), public technology (Baber and Stanton, 1996), and medicine (Nyssen, 2000; Taylor-Adams and Vincent, 2000). Following an earlier paper (Shorrock and Kirwan, 1999), this paper describes the development of a tool that is currently being used in ATC, and the application of this tool to the evaluation of several options for reduced separation minima in unregulated airspace. This technique is called TRACEr—the technique for the retrospective and predictive analysis of cognitive errors in ATC.

The need for TRACEr was originally prompted by a feasibility study for the use of HRA (including HEI) techniques in ATC (Evans et al., 1998). This study used SHERPA, and the authors concluded that the method was developed for use in the nuclear industry, and it would be of greater benefit to ATC safety for a classification system to be constructed specifically for use in HRA in ATC.

Other available techniques were, therefore, rated by the present authors against the following criteria, developed from an original set of evaluation criteria for HEI techniques proposed by Kirwan (1992a). Table 1 summarises these comparative evaluation ratings.

- *Comprehensiveness*—the ability to discriminate and classify a comprehensive range of errors.
- *Structure and consistency*—the degree to which the technique is structured, leading to more consistent analyses between different users and with the same user over time.
- *Life cycle stage applicability*—the degree to which the technique can be used throughout the formative and summative phases of system design lifecycle.
- *Predictive accuracy*—the degree to which the technique is able to predict potential errors.
- *Theoretical validity*—whether the technique is based on a framework describing human performance, with a theoretically plausible internal structure.
- *Contextual validity*—the degree to which the technique adequately captures the circumstances in which an event occurs.

Table 1
Comparative evaluation of other HEI techniques

	SHERPA	CREAM	GEMS	TAFEI	HEIST	PHEA
Comprehensiveness	M	H	H	L	H	L
Structure	H	M	L	H	M-H	H
Life cycle stage applicability	H	H	L	L	M-H	M
Inter-rater reliability	H	M	L	H	N/K	M
Predictive accuracy	M	M	L	M-H	N/K	N/K
Theoretical validity	H	H	H	H	H	M
Contextual validity	L	M-H	L	L	L	L
Flexibility	M	M	L	L	M	M
Usefulness	M	H	M	L-M	M-H	M-H
Resource efficiency (Training)	M	L	L	M	M	L-M
Resource efficiency (Time)	M	L	L	L-M	M	L
Resource efficiency (Experts)	M	L-M	L	M	M	L
Usability	M	L-M	L	M	M	M
Auditability	M-H	M-H	M	H	M	M-H

L = Low; M = Medium; H = High; N/K = Not Known.

- *Flexibility*—whether the technique enables different levels of analysis according to the project requirements, known information or expertise of the user.
- *Usefulness*—whether the technique suggests, or can generate, effective error reduction or mitigation measures.
- *Training requirement*—the time taken to become proficient in the use of the technique. (TRACEr was initially aimed at those working in the field of human factors/ergonomics.)
- *Resource usage*—the amount of time required to collect supporting information and conduct the analysis.
- *Usability*—the ease of use of the technique.
- *Auditability*—the degree to which the technique lends itself to auditable documentation.

This evaluation took account of validation evidence (Kirwan, 1992b, 1998a, b) and other papers that have reported on the use of the various techniques. These existing techniques were not considered adequate to address the needs of ATC. The main problems were considered to be low usability (often due to lack of structure, excessive jargon or excessive 'resolution', i.e. distinctions which were not possible to make reliably), low contextual validity for ATC (particularly important for performance shaping factors—PSFs), and limited applicability (e.g. to skill- and rule-based performance only, or to small-scale systems or applications only). The criteria above were therefore considered throughout the development of TRACEr. There is a balance to be achieved in meeting the requirements, and no technique will fully satisfy all of them. For instance, a technique that is highly comprehensive with subtle distinctions between many categories, will often lead to higher resource usage and lower consistency of use when compared to a gross list of broad categories.

2. The 'Janus' perspective

An important observation on the evolution of HEI approaches relates to their being based on the real operational context of the domain in which they are being applied, and the scope of their use within the domain. Traditional methods of addressing this have tended to focus exclusively on different stages of the system development lifecycle, including prospective methods (e.g. predictive HEI) at the design stages, and retrospective approaches (e.g. incident analysis) during operation. Furthermore, in the development stages, prototyping and real-time simulation have been used in many industries to provide evidence of safety. All approaches share a need to analyse human error, and yet incident/accident analysis and performance prediction have been pursued as two largely separate activities, by psychological and engineering communities (Hollnagel, 2000). The best way to maintain a proper account of context, and hence the best way to ensure the accuracy and insightfulness of a HEI tool, is for it to be used both predictively and retrospectively. This is termed 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.

A tool that is both retrospective and predictive will be continually tested and refined via incident analysis, and will evolve along with the technology and work environment. It is also possible that such a tool will help to bring together the (largely) separate communities responsible for incident investigation/analysis and performance prediction, e.g. for new system design, training, etc. Therefore, this dualistic role of the HEI technique described in this paper is fundamental, and should enhance its utility and added value to safety.

3. Theoretical architecture and practical framework

TRACEr was developed in an iterative fashion with inputs from a variety of activities, including an experimental study, a literature review (covering over 70 sources), the analysis of ATC incidents from 1996 to 1999, interviews of approximately 30 controllers on human error, several large-scale real-time simulations, the use of knowledge elicitation methods, and controller reviews of TRACEr taxonomies.

TRACEr has a modular structure, comprising eight taxonomies or classification schemes. There are three main types of taxonomy: those describing the context within which the error occurred—essential in an HEI technique (see Dougherty, 1993; Hollnagel, 1993); those addressing the production of the error; and those describing the recovery of the error. The modular structure shares some similarities with the multifactorial taxonomy of Rasmussen et al. (1981), and has several benefits. First, it allows the analyst to describe the error at a level for which there is supporting evidence. For example, if the cognitive origins of the error are unknown, the analyst can still describe the external manifestations of the error. This increases the flexibility of the analysis. Second, it allows users to select only those taxonomies that are purposeful in the context of the analysis, thus increasing the efficiency of resource usage. Third, it explicitly maps the relationships between the various classifications, as opposed to a 'pick list' approach, which could confuse fundamentally different types of classifications. Fourth, when combined, the various classifications from each taxonomy form a rich picture of the event. Fig. 1 depicts the taxonomies within TRACEr and their relationships. Fig. 2 depicts the process of using the TRACEr taxonomies for retrospective or predictive analysis. Each taxonomy is further described in the following text.

4. Context

The *task error* taxonomy provides 13 categories describing controller errors in terms of the task that was not performed satisfactorily, and is used for retrospective analysis. Task error categories include, for example, 'radar monitoring error', 'co-ordination error', and 'flight progress strip use error'. These categories provide a high-level view of error that controllers and investigators can easily relate to, and an organising structure that may be required for periodic reports of error trends.

The *information* taxonomy describes the subject matter or topic of the error, and the terms within the taxonomy relate specifically to the internal error modes (IEMs) described later. For instance, what information did the controller misperceive, forget, or misjudge, or miscommunicate? 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, or alternatively what is being misperceived or misjudged. However, few such taxonomies exist in other HEI tools. This is probably because of the difficulty in capturing the

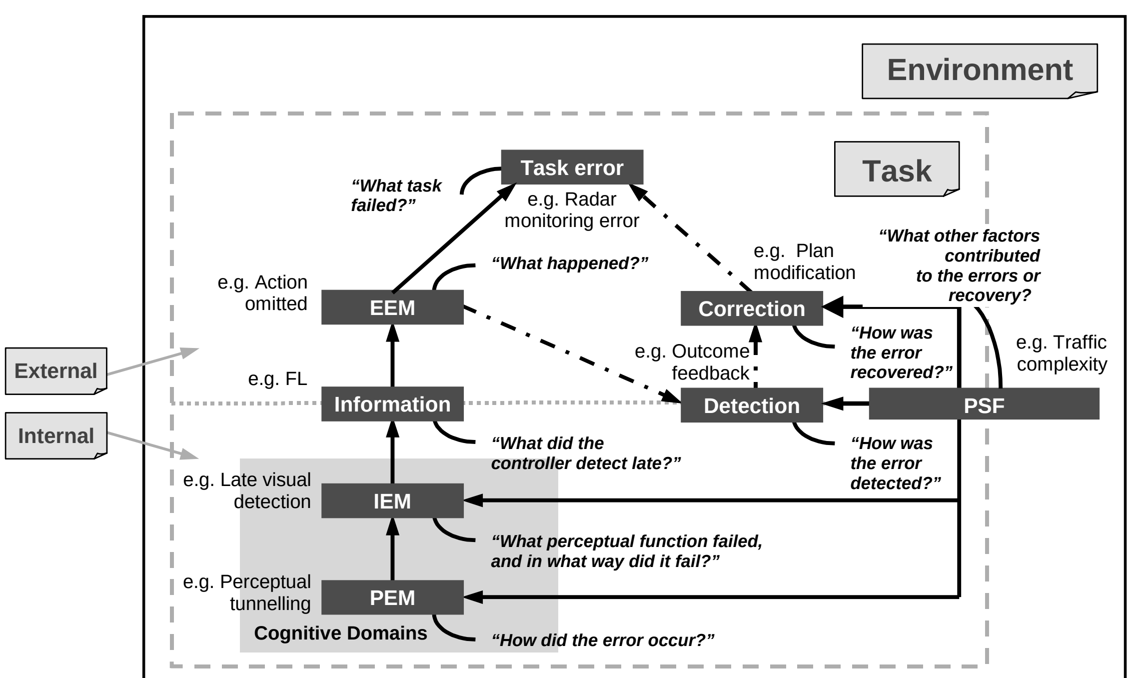

Fig. 1. Relationship between TRACEr classification systems.

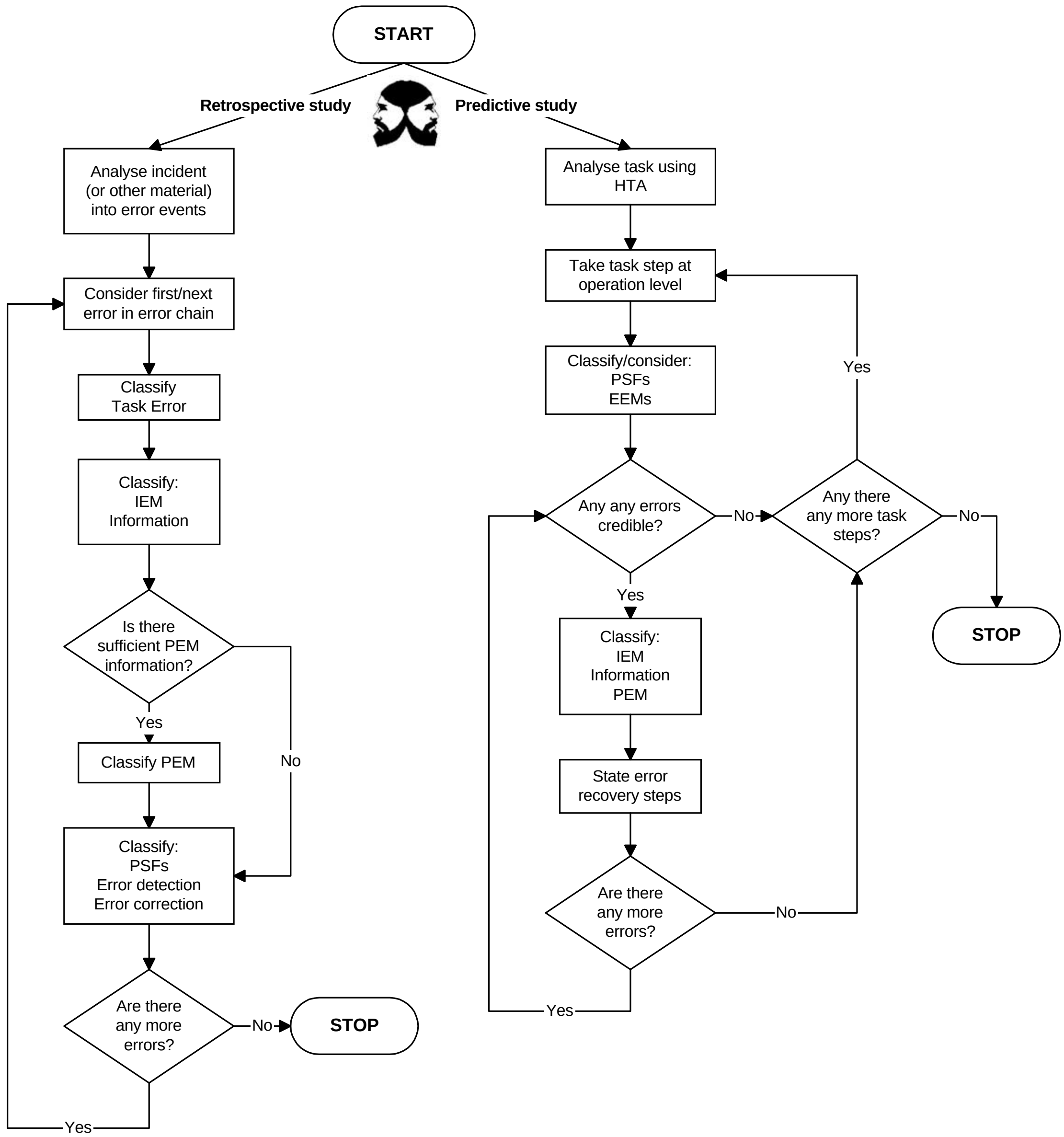

Fig. 2. Process of using the TRACEr taxonomies for retrospective and predictive analysis.

relevant contextual information factors in changing operational environments.

A *performance shaping factors* (PSF) taxonomy classifies factors that have influenced or could influence the controller's performance, aggravating the occurrence of errors, or perhaps assisting error recovery. Extracts from these lists are shown in Table 2.

5. Error production

Error production can be classified at a number of hierarchical 'levels' within TRACEr. For instance, a controller might fail to respond to a visual alert (an 'omission'). This omission could have occurred for a number of reasons. For instance, the controller might not have seen the alert; a failure of perception ('no detection'). Alternatively, the controller might have seen the alert but decided not to act on it. This could be described as a failure of decision making ('poor decision'). It is possible to analyse the error further by exploring why the controller did not see the alert or why the controller decided not to act on the alert. For instance, the controller might have been visually fixated on another part of the display ('perceptual tunnelling'), or assumed that the alert was a false alarm ('false assumption'). This example illustrates how a simple classification of 'omission' is insufficient and potentially misleading. To capture these various layers in error production, TRACEr classifies errors in three ways. Each of these will now be introduced in more detail.

Table 2
Extracts from TRACEr's task, information and PSF taxonomies

Task error	Information category and keyword	PSF category and keyword
Separation error	Controller materials	Traffic and airspace
Controller-pilot communications error	e.g. Flight progress strip (fps)	e.g. Traffic complexity
Radar monitoring error	Controller activities	Pilot/controller communications
Aircraft observation/recognition error	e.g. Transfer	e.g. RT workload
Co-ordination error	Variable aircraft information	Procedures
Control room communication error	e.g. Callsign	e.g. Accuracy
Aircraft transfer error	Time and location	Training and experience
Hand-over/Take-over error	e.g. Airspace type	e.g. Task familiarity
Flight progress strip use error	Airport	Workplace design, HMI and equipment factors
Operational materials checking error	e.g. Runway	e.g. Radar display
Training, supervision, or examining error	Other	Ambient environment
Human-machine interaction error		e.g. Noise
Other task error		Personal factors
		e.g. Alertness/fatigue
		Social and team factors
		e.g. Handover/takeover
		Organisational factors
		e.g. Conditions of work

Table 3
TRACEr's external error mode taxonomy

Selection and quality	Timing and sequence	Communication
Omission	Action too long	Unclear information transmitted
Action too much	Action too short	Unclear information recorded
Action too little	Action too early	Information not sought/obtained
Action in wrong direction	Action too late	Information not transmitted
Wrong action on right object	Action repeated	Information not recorded
Right action on wrong object	Mis-ordering	Incomplete information transmitted
Wrong action on wrong object		Incomplete information recorded
Extraneous act		Incorrect information transmitted
		Incorrect information recorded

5.1. External error modes

External error modes (EEMs) classify the external and observable manifestation of the actual or potential error, based on logical outcomes of erroneous actions, in terms of timing, sequence, selection, quality, and so on. EEMs are context-free and independent of cognitive processes (e.g. intention). TRACEr's EEM classification (see Table 3) is adapted from an influential taxonomy (Swain and Guttmann, 1983), which distinguished between three main categories of errors: errors of omission, errors of commission (e.g. selection, sequence, timing and quality) and extraneous errors. EEMs are generally only used as prompts for error prediction, since they have little descriptive meaning.

5.2. Cognitive framework

Several authors have advocated the use of an underlying model of human performance for human error classification. Rouse and Rouse (1983) assert that the 'internal consistency of a classification scheme is likely to be enhanced if the scheme is based on a model of the process within which errors occur' [p. 540]. Such a model, they argue, can help to identify categories within the classification scheme and illustrate the relationships among categories. A theoretically plausible model or framework is particularly important for error reduction purposes (Kirwan, 1992a).

A number of cognitive frameworks and models of task performance and human error were considered (note that many 'models' in the literature are better described as frameworks). Those considered included the following (and their derivatives): Bagnara et al. (1989), Berliner et al. (1964), Fleishman and Quaintance (1984), Hollnagel (1993, 1998), Jones and Endsley (1996), Norman (1986), Rasmussen (1982), Reason (1979, 1987a, b, 1990), Rouse and Rouse (1983), Wickens (1992) and Zapf et al. (1994).

No widely accepted models of controller performance were identified, and such a model may not be desirable, since it would be subject to considerable change in the advent of new technology. Hence, it was considered that the most suitable cognitive framework for TRACEr was one broadly based on Wicken's (1992) framework, and Hollnagel and Cacciabue's (1991) 'simple model of cognition' (SMoC).

The concepts within the cognitive framework were termed 'cognitive domains', a term borrowed from Reason (1987a). This helps to overcome the outdated notions of serial processing that characterised the early information-processing tradition. Wickens (1992) notes that 'information flow need not start with the stimulus…sometimes our decisions or responses are internally triggered by 'thoughts' in working memory' [p. 20]. Wickens also notes that information flow need not progress through the perception-cognition-action stages. Various studies have shown, for example, that memory and visual imagery can affect perception (both interference and facilitation effects). Hollnagel and Marsden (1996) also emphasise the cyclical nature of cognition in the SMoC. The cognitive domains within TRACEr comprise the following:

(i) *Perception*: errors in visual detection and visual search, and errors in listening.
(ii) *Memory*: forgetting (or misrecalling) temporary or longer-term information, forgetting previous actions, and forgetting planned actions.
(iii) *Judgement*, *planning and decision-making*: errors in judging aircraft trajectories, errors in making decisions, and errors in planning.
(iv) *Action execution*: actions or speech performed not-as-planned.

The cognitive framework above is deliberately simple, with only four major categories. Hollnagel and Marsden (1996) note that there is general agreement about the functions and functional characteristics of human cognition, and in particular performance limitations. The impact of the cognitive domains above in ATC has been well documented in numerous studies (see Roske-Hofstrand and Murphy, 1998). However, there is less agreement about the details: 'It seems that the more detailed a model of cognition is, the less likely it is to be correct' [p. 41]. Also, Roske-Hofstrand and Murphy (1998) assert that 'there are a variety of controller positions, which make quite different cognitive demands on the controller' and that 'the cognitive task requirements…result in qualitatively different cognitive work experiences' [p. 69]. The framework selected is also widely known, and employs concepts (e.g. perception, memory) that are familiar to those with no formal training in human factors, such as air traffic controllers. It is the authors' experience that other process-based frameworks (e.g. Rasmussen, 1982) are less acceptable to air traffic controllers, who find them too complex, too process-oriented, or too difficult to relate to their own experience. Furthermore, newer concepts such as situation awareness and mental models are steeped in controversy (e.g. Flach, 1995), and one might question the usefulness of attempts to use them in a classification scheme (e.g. Jones and Endsley, 1996).

In order to create a taxonomy of psychological errors, a comprehensive search was made for error types documented within three sources. First, error types were identified from previous psychology and human factors research (over 70 reference sources). Second, existing error classification techniques were surveyed (a selection are shown in Table 4). Third, a number of errors were identified from ATC aircraft proximity (Airprox) incidents and data from real-time ATC simulations.

Many documented error types were recorded, which had to be filtered to a coherent and manageable set. The error types were therefore checked to ensure that, as far as possible, they were mutually exclusive and applicable to ATC. This latter check also involved demonstrating TRACEr to air traffic controllers, who described situations where the errors could occur. It became apparent that the error 'database' described psychological errors at two or more 'levels'. For instance, an error may be described as 'misidentification', but going a level 'deeper', one might find that this was due to expectations, i.e. seeing what you expect to see—'expectation bias'. Such findings led to the creation and differentiation of 'internal error mode' (IEM) and 'psychological error mechanism' (PEM) taxonomies.

The cognitive framework was used to organise IEMs and PEMs directly, according to their inferred location within the cognitive framework. However, recognising that errors could be associated with more than one cognitive domain, one guiding principle in determining this was in examining the mapping used by other authors. Table 4 shows a comparison of the cognitive domains utilised in the present work and some comparable stages of information-processing or cognitive domains from a selection of other human error classification schemes. Those error types that were not included in previous frameworks could be located within a cognitive domain by considering the research context of the error type.

5.3. Internal error modes

IEMs are linked specifically to the *functions* of the cognitive domains, and describe what cognitive function failed or could fail, and in what way. For instance, the cognitive domain 'perception' was divided into 'visual' and 'auditory', as well as 'detection' and 'identification', and 'recognition'. The cognitive functions within each cognitive domain were then combined with a *keyword.* Example keywords include late, none, incorrect, etc. IEMs therefore describe the internal manifestation of

Table 4
Comparison of 'cognitive domains' and comparable stages of information processing from other human error classification systems

Cognitive domain	Developer	Comparable stage of information processing/cognitive domain
Perception	Payne and Altman (1962)	Input errors
	Andersch et al. (1969)	Hears and reconstructs
	Pew et al. (1981)	Activation/detection of system-state signal, observation and data collection, identification of system state
	Rouse and Rouse (1983)	Observation of system state
	Norman (1986)	Perception, interpretation
	Reason (1987a)	Recognition failures, attentional failures
	Hollnagel (1993)	Perception/observation, Interpretation
	Kirwan (1994) based on Rasmussen (1986)	Activation/detection, observation and data collection
Memory	Payne and Altman (1962)	Mediation errors
	Reason (1979)	Storage failures
	Norman (1981)	Slips during the formation of an intention
	Reason (1987a)	Memory lapses, inaccurate and blocked recall
	Reason (1990)	Skill-based lapses
	Hollnagel (1993)	Memory
Judgement, planning and decision making	Payne and Altman (1962)	Mediation errors
	Andersch et al. (1969)	Structures, evaluates
	Pew et al. (1981)	Identification of system state, interpretation of situation, evaluation of alternative strategies, definition of objectives, procedure selection
	Rasmussen (1982)	Knowledge-based errors
	Rouse and Rouse (1983)	Choice of hypothesis, testing of hypothesis, choice of goal, choice of procedure
	Rasmussen (1986)	Interpret, evaluate, define task, formulate procedure
	Norman (1986)	Evaluation, goals, intention
	Reason (1987a)	Errors of judgement, reasoning errors
	Reason (1990)	Knowledge-based mistakes, violations
	Hollnagel (1993)	Interpretation, planning/choice
	Kirwan (1994) based on Rasmussen (1986)	Identification of system state, interpretation, evaluation, goal selection and task definition, procedure selection
Action execution	Payne and Altman (1962)	Output errors
	Andersch et al. (1969)	Reacts, transmits
	Reason (1979)	Discrimination failures, program assembly failures, test failures, and sub-routine failures
	Norman (1981)	Slips that result from faulty activation of schemas, slips that result from faulty triggering of schemas
	Pew et al. (1981)	Procedure execution
	Rouse and Rouse (1983)	Execution of procedure
	Rasmussen (1986)	Execute
	Norman (1986)	Action specification, execution
	Reason (1987a)	Unintended words and actions
	Reason (1990)	Skill-based slips
	Hollnagel (1993)	Action execution
	Kirwan (1994) based on Rasmussen (1986)	Procedure execution

the error within each cognitive domain (e.g. 'late detection', 'misidentification', 'hearback error').

IEMs provide an interface between EEMs, PEMs, and the cognitive framework, and thus give an intermediate level of detail. IEMs are usually obtainable from incident reports, and form a very useful part of the analysis. This classification scheme can be compared with Rasmussen et al.'s (1981) concept of the 'internal mode of malfunction', which sits between the 'mechanism of human malfunction' and the 'external mode of malfunction' in his framework. Table 5 shows how the TRACEr IEMs were generated for each cognitive domain.

5.4. Psychological error mechanisms

Psychological error mechanisms (PEMs) describe the psychological nature of the IEMs within each cognitive domain; the cognitive biases that are known to affect performance. PEMs within 'perception' include

Table 5
Generation of IEMs within TRACEr

Cognitive Domain	Cognitive Function		Relevant Keywords	Example IEM
Perception	Vision	Detection	None, late, incorrect	Late detection
		Identification	None, late, incorrect	Misidentification
	Hearing	Recognition/ Comparison	None, late, incorrect	Hearback error
Memory	Recall perceptual information		None, incorrect	Forget temporary information
	Previous actions		None, incorrect	Forget previous actions
	Immediate/current action		None, incorrect	Forget to perform action
	Prospective memory		None, incorrect	Prospective memory failure
	Stored information (procedural and declarative knowledge)		None, incorrect	Misrecall stored information
Judgement, Planning and Decision Making	Judgement		Incorrect	Misprojection
	Planning		None, too little, incorrect	Underplan
	Decision Making		None, late, incorrect	Incorrect decision
Action Execution	Timing		Early, late, long, short	Action too early
	Positioning		Too much, too little, incorrect, wrong direction	Positioning error: overshoot
	Selection		Incorrect	Typing error
	Communication		None, unclear, incorrect	Unclear information transmitted

'expectation bias' (i.e. seeing or hearing what you expect to hear), 'perceptual confusion' (i.e. confusing two things that look or sound alike), and 'distraction/preoccupation'. Examples of how the same 'source PEMs' could affect different cognitive domains are shown in Table 6.

PEMs provide a fine level of detail, which is useful for error reduction and mitigation. However, they may require significant understanding of psychological aspects of an error, which may not always be obtainable from incident reports. Table 7 shows the IEMs and PEMs within TRACEr.

5.5. Error detection and correction

Attention to error recovery in theoretical and applied work is a newer development, but recording such

Table 6
Examples of the effect of 'source PEMs' in different cognitive domains

Example 'Source PEMs'	Example cognitive domain	Example PEMs
Complexity, understanding	Memory	Insufficient learning
	Judgement, planning and decision making	Integration failure
Expectation, assumption	Perception and vigilance	Expectation bias
	Judgement, planning and decision making	False assumption
Association, confusion, interference, habit	Perception and vigilance	Perceptual confusion
	Memory	Negative transfer, similarity interference
	Action execution	Habit intrusion
Tunnelling, fixation	Perception and vigilance	Perceptual tunnelling
	Memory	Memory block
	Judgement, planning and decision making	Cognitive fixation
Overload, underload	Perception and vigilance	Vigilance failure
	Memory	Memory capacity overload
	Judgement, planning and decision making	Decision freeze
Internal distraction, preoccupation	Perception and vigilance	Distraction/preoccupation
	Memory	Distraction/preoccupation
	Action execution	Environmental intrusion

information can bear fruits for future error reduction. For example, if it is known that team members are frequently pointing out conflicts to other team members in one ATC watch, then this may be an example of good team resource management (TRM) the ATC version of crew—or cockpit—resource management (CRM, see Wiener et al., 1996). Alternatively, if many errors go undetected, only to be detected late into the development of the incident, this may signal the need for better TRM. With these considerations in mind, a list of error detection keywords was developed, influenced by the work of others such as Sellen (1994), Rizzo et al. (1995), Wioland and Amalberti (1998) and Kontogiannis (1997). Four questions prompt the selection of keywords:

1. How did the controller become aware of the error? E.g. action feedback, inner feedback, outcome feedback.
2. What was the feedback medium? E.g. radio, telephone, radar display, flight progress strips.
3. Did any factors, internal or external to the controller, improve or degrade the detection of the error? (Refer to PSFs.)
4. What was the separation status at the time of error detection? E.g. separation lost, separation maintained.

Once error detection has been classified, it is clearly useful to identify if and how the error was corrected or recovered. The following questions prompt classification.

1. What did the controller do to correct the error? E.g. reversal or direct correction, automated correction, plan modification.
2. How did the controller correct the error? (In operational terms, refer to information keywords.) E.g. turn or climb.
3. Did any factors, internal or external to the controller, improve or degrade the detection of the error? (Refer to PSFs.)
4. What was the separation status at time of error correction? E.g. separation lost, separation maintained.

6. Representation and use of TRACEr

TRACEr is represented as a set of colour-coded decision-flow diagrams and tables. Separate decision-flow diagrams have been developed for predictive and retrospective use. Such diagrams were selected because they increase the usability of the technique, assist training and familiarisation, increase inter-analyst agreement, and help to specify the taxonomic relationships between errors. Decision-flow diagrams have been used previously in HEI techniques (e.g. Embrey, 1986; Pew et al., 1981). An example diagram is shown in Fig. 3.

7. TRACEr applications

TRACEr has been applied to a variety of ATC projects over 5 years for three types of work by several human factors specialists.

1. Retrospective analysis

(a) Analysis of UK aircraft proximity (Airprox) incidents (a mandatory reporting system) occurring

Table 7
Internal error modes and psychological error mechanisms within TRACEr

Cognitive Domains	
IEMs	PEMs
Perception	
No detection (visual)	Expectation bias
Late detection (visual)	Spatial confusion
Misread	Perceptual confusion
Visual misperception	Perceptual discrimination failure
Misidentification	Perceptual tunnelling
No identification	Stimulus overload
Late identification (visual)	Vigilance failure
No detection (auditory)	Distraction/preoccupation
Hearback error	
Mishear	
Late auditory recognition	
Memory	
Forget to monitor	Similarity interference
Prospective memory failure	Memory capacity overload
Forget previous actions	Negative transfer
Forget temporary information	Mislearning
Misrecall temporary information	Insufficient learning
Forget stored information	Infrequency bias
Misrecall stored information	Memory block
	Distraction/Preoccupation
Judgement, planning and decision making	
Misprojection	Incorrect knowledge
Poor decision	Lack of knowledge
Late decision	Failure to consider side- or long-term effects
No decision	Integration failure
Poor plan	Misunderstanding
No plan	Cognitive fixation
Under-plan	False assumption
	Prioritisation failure
	Risk negation or tolerance
	Risk recognition failure
	Decision freeze
Action execution	
Selection error	Manual variability
Positioning error	Spatial confusion
Timing error	Habit intrusion
Unclear information transmitted	Perceptual confusion
Unclear information recorded	Functional confusion
Incorrect information transmitted	Dysfluency
Incorrect information recorded	Misarticulation
Information not transmitted	Inappropriate intonation
Information not recorded	Thoughts leading to actions
	Environmental intrusion
	Other slip
	Distraction/preoccupation

within both controlled and unregulated airspace (this latter work discussed later).

(b) Analysis of confidential incident/error reports (voluntary reporting system) from the confidential human factors incident reporting programme (CHIRP).

(c) Analysis of controller interviews regarding unreported human errors.

2. ‘Real-time’ analysis

(a) Analysis of errors occurring in large-scale real-time simulations as part of the New Scottish Centre (NSC) programme (Shorrock et al., 2001).
(b) Analysis of errors occurring in small-scale military simulations of reduced separation standards in unregulated airspace (Shorrock et al., 2000; also discussed later).

3. Predictive analysis

(a) Human error prediction for the final approach spacing tool (FAST) and NSC (Evans et al., 1999; Shorrock et al., 2001).
(b) Human error prediction to support an analysis of reduced separation standards in unregulated airspace (Shorrock et al., 2000; also discussed later).

One project above was approached using both the retrospective, ‘real-time’ and predictive modes of TRACEr; the analysis of reduced separation options in unregulated airspace. The remainder of this paper describes this study, which had implications for safety, efficiency and capacity.

8. The study: reduced separation in unregulated airspace

8.1. Background

This study involved the evaluation of several options for reduced separation minima in unregulated airspace (Class F and G) (Shorrock et al., 2000). Within such airspace, pilots may be offered a radar information service (RIS) or a radar advisory service (RAS), or receive no ATC service whatsoever. Pilots are not required to follow ATC instructions with these services; they may elect to maintain their own separation.

Under a RAS, the controller will pass information on nearby traffic to pilots, and will provide advice to prevent a loss of separation. Within UK airspace, separation standards define the minimum distance (laterally and vertically) that must be maintained between aircraft receiving an ATC service. The advice provided by the controller under a RAS for the prevention of losses of separation may include climb, descent or turn instructions, or any combination of these.

At the time of the study, under a RAS in unregulated airspace, the prescribed separation minima were 5000 ft Mode-C vertically and 5 nautical miles (NM) horizontally when the RAS traffic was separated from unknown traffic. These minima were in place for over 20 years,

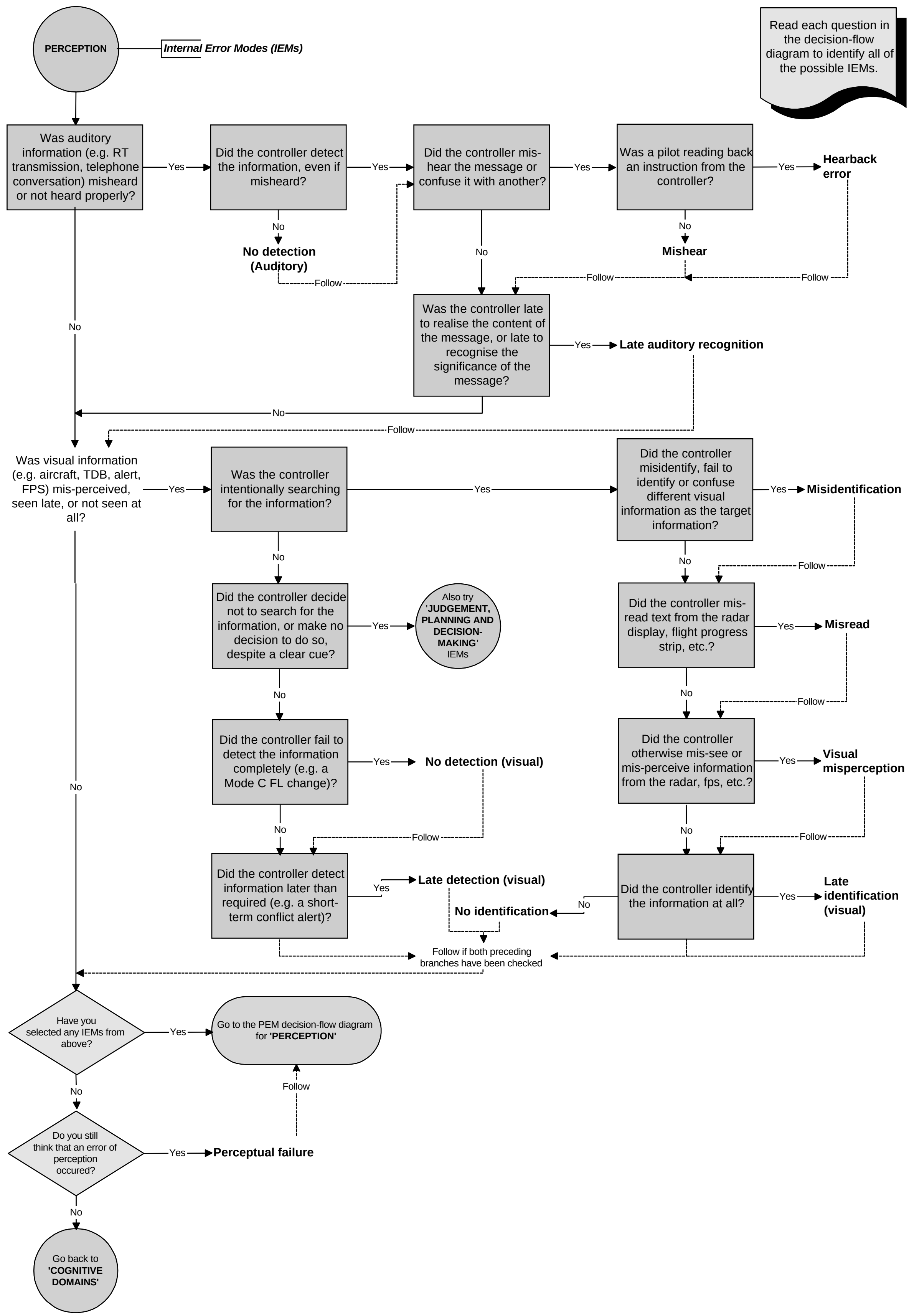

Fig. 3. Extract of the internal error mode diagram for the perception and vigilance cognitive domain.

when both radar systems and aircraft avionic systems were significantly less advanced than today. And whilst the issue of traffic growth in the UK generally refers to regulated airspace, the problem also extends to unregulated airspace.

A report by the Civil Aviation Authority's Directorate of Airspace Policy recommended that National Air Traffic Services (NATS) review existing radar separation standards in unregulated airspace. This led to a study to evaluate the proposed reduction of vertical separation minimum from 5000 to 3000 ft Mode-C and the lateral separation minimum from 5 to 3 NM.

The following studies were conducted:

- Review of aircraft proximity (Airprox) reports occurring over a 4-year period using TRACEr to determine the current frequency of incidents and their causes.
- Literature review of pertinent issues.
- Predictive human error analysis (HEA) of the provision of a RAS, using TRACEr HEI, fault tree analysis (FTA) and event tree analysis (ETA).
- Real-time simulations to evaluate all options, including analysis of errors observed and recorded.

A variety of methods were used during the studies. The following sections outline the methods used to study human error aspects.

8.2. Retrospective application—incident analysis

One of the primary measures of safety performance within NATS is the Airprox reporting system, part of the Mandatory Occurrence Reporting system. An Airprox is defined as 'A situation in which, in the opinion of a pilot or a controller, the distance between aircraft as well as their relative positions and speed have been such that the safety of the aircraft involved was or may have been compromised' (CAA, 1996).

Thirty-one Airprox reports pertaining to incidents occurring between 1991 and 1995 were analysed using TRACEr. The aim of this analysis was to determine the possible effects of separation standards on incidents involving aircraft receiving a RAS, as well as the types of errors that controllers currently make. Of these reports, 25 involved aircraft in receipt of a RAS from civil controllers and six involved aircraft in receipt of a military service.

Each report reviewed entailed a loss of separation within unregulated airspace between 2000 ft and FL240. Twenty-four of the 31 original reports were classified by the Joint Airprox Working Group or Joint Airprox Assessment Panel as 'No risk of collision', whilst seven of the reports were deemed to compromise safety (three military and four civil). None of the incidents were classed as having an 'actual risk of collision'. TRACEr was used to analyse the reports, classifying task error, information, PSF, IEM and PEM. Possible effects of the current separation standards and reduced separation were judged.

The analysis revealed some types of error that could have a greater effect with reduced separation. The main errors of significance to the study were as follows:

- *Misjudgements of heading* (four errors). This observation is significant to reduced lateral separation, particularly where the controller is aiming for little more than separation minima.
- *Late detection of conflict* (four errors), where the controller concerned was late to notice a conflicting aircraft. This error could affect reduced separation more where the controller has first seen an unknown aircraft and then deals with other traffic before returning to check the unknown aircraft.
- *Delayed avoiding action instructions or traffic information* (four errors). With a potentially reduced period available to resolve conflicts, it is essential that the controller provides avoiding action once aware of the conflicting aircraft.

PSFs noted in the reports included Mode C/secondary surveillance radar (SSR) (flight level display) problems due to closer aircraft proximity (e.g. no Mode C, SSR label and aircraft symbol reflections, label overlap) and high workload (due to complexity, traffic load or staff shortages). These factors are indirect contributors to risk, and could influence performance in reduced separation conditions. Workload could reduce generally with reduced separation due to reduced communication, but increase significantly when the reduced separation minima are eroded. Problems of SSR garbling and label overlap could increase due to reduced aircraft proximity.

The separation standard was not thought to be a significant factor in 24 of the 31 reports (e.g. the controller failed to notice the presence of conflicting aircraft). In seven reports, it was unclear how reduced separation would have affected the incident, particularly with respect to the timings involved in controllers calling traffic.

In conclusion, the AIRPROX review suggested that the current separation criteria generally did not have a large effect on the incidents. However, in a small number of cases, separation standards did have a bearing on the incident. TRACEr helped to identify some key errors to consider in future analyses.

8.3. Real-time application—simulations

Two real-time simulations (one area control, one terminal control) were conducted in the high-fidelity area radar training simulator at the Central ATC School in RAF Shawbury to test the application of reduced separation minima. Three controllers participated in the terminal simulation and four participated in the area simulation. Each simulation lasted four days, with one of the following separation criteria assessed per day: 5 NM/5000 ft Mode-C; 5 NM/3000 ft Mode-C; 3 NM/5000 ft Mode-C; 3 NM/3000 ft Mode-C.

Among the methods used to analyse the simulation data, three methods provided information pertaining to human error potential: questionnaires; controller debriefs and open discussions; and observation and video analysis.

Seven erosions of separation minima occurred in each simulation. In the first (terminal) simulation, the erosions were most often associated primarily with lateral separation (e.g. misjudging a turn, boxed in by conflicting traffic). Other erosions were associated with visual monitoring and distraction. In the second (area) simulation, five of the erosions involved the late detection of a conflictor (e.g. the controller was aware of an aircraft, but the aircraft turned into confliction). The remaining erosions were due to either misjudgement of a turn or misidentification of an aircraft.

In the open discussion, the controllers noted that, whilst 3 NM made controlling easier, and reduced the frequency of 'avoiding action' turns, it was more difficult to gauge visually on the radar display. This is because the airways, being 10 NM wide, provide a gauge against which to estimate 5 NM.

TRACEr was used during and after simulations to help organise the questionnaire and observation/video data, and as an aid to probe discussions and debriefs.

8.4. Predictive application—human error analysis

Human error analysis can be directed backwards to consider an undesirable event and then determine what errors could lead to such an event, or forwards to predict what errors would be likely as a situation unfolds. Both approaches, known as fault tree analysis (FTA) and event tree analysis (ETA), were used in this study. TRACEr was applied to a hierarchical tasks analysis (HTA) of the process of providing a RAS, and employed to predict more independent errors that relate to the sub-tasks involved in separating aircraft to feed into the FTA and ETA. Fig. 4 shows a small extract of the fault tree to illustrate how outputs from the TRACEr HEI were used in the context.

Most of the errors identified in the analyses could occur with the current separation standards, but their effects or their frequency could be different with reduced separation minima. In many cases, time pressure is a major factor that hinders error detection and recovery. However, some errors might be prevented with reduced separation. The main types of errors were as follows:

Judgement: Misjudgements are a particular area of concern for reduced lateral separation, because traffic tends to manoeuvre more laterally than vertically. However, the controller could suggest climbs or descents that the aircraft cannot make, due to weight, weather, etc. Importantly, though, if the controller were to misjudge a required climb or descent, he or she could still opt to stop the climb or descent and turn the aircraft to achieve lateral separation.

Memory: Controllers could forget to issue a planned instruction (e.g. FL or heading) after a distraction, or may forget received information. Reduced separation generally allows less time to address resulting situations, and places more demand on the pilot to sight traffic or request and implement avoiding action. Other errors of memory could include forgetting to check the position of traffic previously observed at long range. With reduced separation minima, it is possible that controllers could delay such checks, knowing that traffic will take more time to cover the distance to separation minima.

Visual perception: Reduced lateral separation would mean that the controller would be required to judge 3NM on the radar screen. This would be more difficult than judging 5 NM. Reduced vertical separation would increase the visual demand in detecting gradual changes in digital Mode C FLs, both for unknown aircraft and in ensuring that RAS-supplied aircraft do not bust their cleared FLs. The visual demand on controllers is likely to increase if aircraft are separated at the proposed separation minima, because the controller would need to be more vigilant to notice any aircraft deviations, since erosions are likely to be more serious. This could result in the controller becoming fixated on these potential conflicts, particularly with 3NM lateral separation. Reduced separation also offers less time to see whether avoiding action has been successful.

Communication: Controllers could make a number of communication errors, for instance where the controller makes a slip of the tongue or omits information from an instruction. The pilot might 'step on' a message by trying to use the radio–telephone at the same time as the controller, or could confuse headings and FLs. Again, reduced separation, particularly 3 NM laterally, allows less time for detection and correction. However, if the controller is engaged in fewer communications, there should be fewer opportunities for error and possibly more time to detect and correct communication errors.

On the positive side, reduced lateral and vertical separation should allow more time for *planning*, as there would be fewer routine heading and level changes, fewer conflicts to resolve and lower RT load. This should help to prevent further problems in tactical control. Also, reduced separation could reduce needless turns that can create secondary conflicts.

8.5. Study conclusions

On analysing the findings of the methods above, and the other methods used in the study, the weight of evidence in this study suggested that reduced lateral

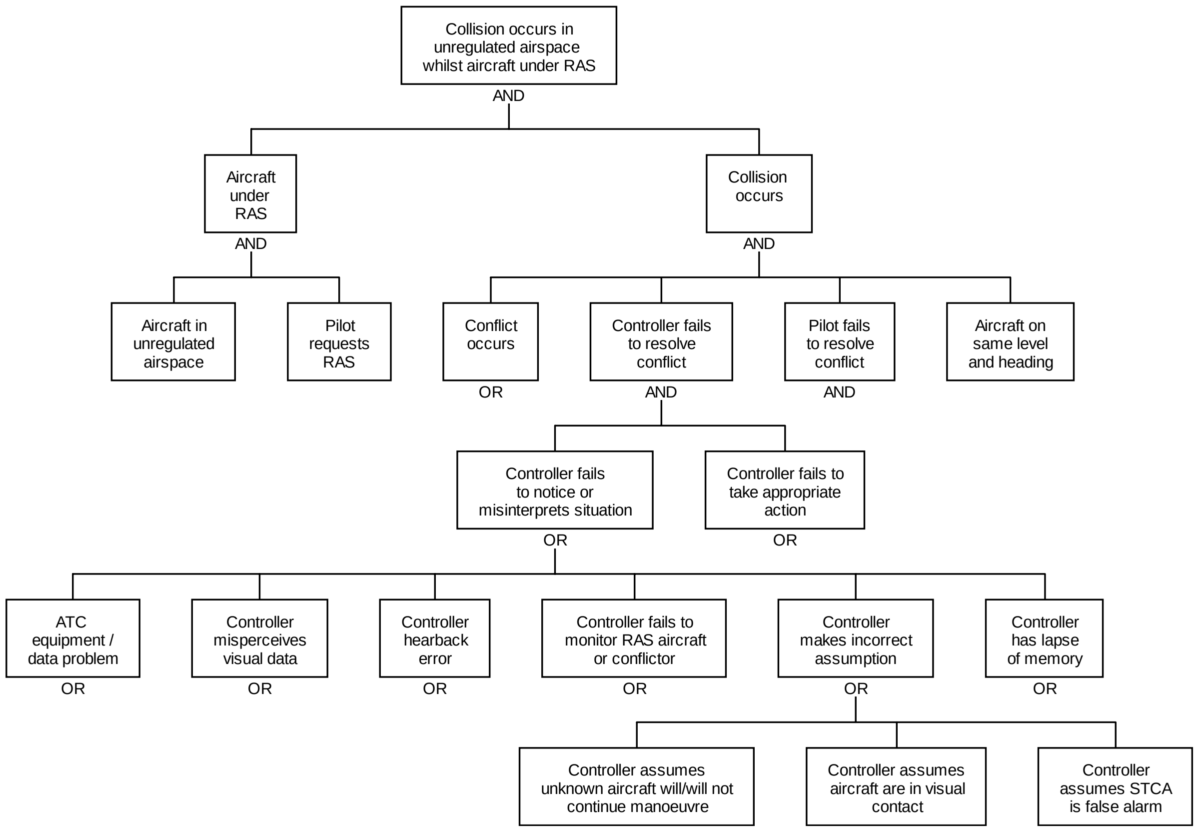

Fig. 4. Extract of the fault tree for 'Collision occurs in unregulated airspace whilst under RAS.' (Branches that end in 'AND' or 'OR' have been collapsed.)

separation to 3 NM was of greater safety concern, and offered fewer benefits, than reduced vertical separation to 3000 ft Mode-C. This was supported by independent quantitative collision risk modelling. Of the three reduced separation options evaluated, the 5 NM/3000 ft Mode-C option was proposed for further evaluation. Whilst the controllers were keen (subjectively) to see separation reduced to 3 NM, the TRACEr analysis helped to clarify more objectively the potential problems that could impact on safety. This shows the value of having an HEI technique, and a concomitant attention to consequences—without this, potentially significant errors may be ignored or 'down-played'.

TRACEr was instrumental in analysing data from multiple sources including incident reports and various real-time simulation measures, and in predicting further potential errors to populate an extensive fault tree and event trees. Overall therefore, TRACEr, a technique grounded in both ergonomics theory and the ATC context, was critical in informing a key operational decision—determining the safety of separation options. Following a final procedural hazard and operability (HAZOP) study, and extended (12-month) live trials at operational centres, the proposed reduced vertical separation option went into operation nationally in December 2000.

9. A note on validation issues

Whilst this paper does not constitute a validation of the technique, an early development study, focusing on inter-analyst reliability, was carried out on a prototype version of TRACEr (Shorrock, 1997). This version contained just one set of 116 'error types' (a mixture of PEMs and IEMs) within 10 cognitive domains. Nine human factors specialists individually classified 23 different events highlighted in four controller-reported Airprox reports. The number of analysts selecting the most frequently chosen categories for each event was calculated (i.e. the *mode* category), and the number of *different categories* selected per event was also calculated. For each event, on average, five out of nine analysts agreed on the same error type category, and responses tended to fall into four error categories on average. This level of agreement was considered reasonable because analysts could potentially chose *any* of 10 cognitive domains and any of 116 error types, even though only a subset would normally seem sensible for each event.

A total of 41 error types were used to classify the 23 events by the group as a whole. Over 98% of the error classifications used error types within TRACEr. A questionnaire was designed to evaluate TRACEr subjectively on some of Kirwan's (1992a) criteria (similar to those presented earlier). The responses suggested that the strongest areas of the technique were comprehensiveness, structure, acceptability of results, and usability. The main area of concern was resource usage (time and expertise).

The findings highlighted a number of areas of confusion in the way that participants interpreted and used the categories in this early version. Several key changes were therefore made at this stage, including a reduction in the number of cognitive domains, separation of 'error types' into IEMs and PEMs, and a reduction in the overall number of categories.

Whilst this study concerned a prototype of TRACEr, it showed some encouraging results and indicated some key areas of variance in the use of TRACEr, which were subsequently addressed. A further study is now being planned to test the reliability of TRACEr.

10. Conclusion

The TRACEr technique has been implemented as a retrospective incident analysis technique, and as a predictive HEI technique for use with new controller tools, procedures, etc. TRACEr uses the same concepts and method for incident analysis and error prediction, thus maximising the efficiency of the method, maximising learning from past events, and providing a better basis for comparison of findings. In the former usage, TRACEr has helped to unpack the nature of errors contributing to incidents in the UK more precisely. This led to a number of studies to determine how to prevent such errors from occurring, or to mitigate their consequences, at this stage principally via specialised training and procedures (such as a position hand-over checklist). TRACEr's philosophy of ultimately being oriented towards error reduction has therefore been put into practice.

In its prospective usage, TRACEr has, in several studies, predicted errors associated with new systems and procedures, which, in some studies, subsequently occurred during simulations and prototyping trials. The studies cited earlier also led to design changes, and the study described in this paper contributed significantly to decision-making on separation minima, providing capacity benefits. In several cases, TRACEr has helped to identify strategies to prevent or reduce error. This suggests that the technique is relatively well-focused on the domain of ATC, and has to an extent captured the context of ATC operational practices.

TRACEr is now being applied to further incidents in UK airspace, and it is hoped that it will be also applied to other interface design projects in the future. TRACEr has also been used as the basis for a EUROCONTROL project called HERA, the human error in ATC project (Isaac et al., 1999, 2000, 2002). This project aims to develop an incident error analysis technique that is commonly applied across European States, and ultimately possibly in the USA, following collaboration between EUROCONTROL and the Federal Aviation Administration and their human factors analysis and classification technique (HFACS) (Wiegmann and Shappell, in press). HERA has already been implemented in Swedish operational ATC (Josefsson, 2001). Finally, a version of TRACEr has also been developed more recently for use by incident investigators and air traffic controllers. This tool—TRACEr *lite*—contains five core TRACEr taxonomies (task error, IEM, PEM, information and PSF), which have been reduced in detail by use of knowledge elicitation methods. TRACEr *lite* is currently on trial at Manchester Area Control Centre. It is hoped that the TRACEr tool-set will provide an extra layer in NATS' safety nets, so that air traffic control retains its excellent record of safety.

Acknowledgements

The majority of work described in this paper was conducted whilst the authors were at National Air Traffic Services Ltd. The views expressed in this paper are those of the authors and do not necessarily reflect the opinion or policy of any related organisations. The authors would like to thank the NATS Human Factors Unit, and the controllers and system designers, as well as Professor John Wilson (University of Nottingham), for their enthusiastic support.

References

Air Accidents Investigations Branch, 2001. Incident involving Boeing 747-436, G-BNLY and Airbus A321, G-MIDF at London Heathrow Airport on 28 April 2000. Aircraft Incident Report No.: 1/2001 (EW/C2000/4/6).

Andersch, E.G., Staats, L.C., Bostom, R.N. 1969. Communication in Everyday Use. Holt, Rinehart & Winston, New York.

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

Baber, C., Stanton, N.A., 1996. Human error identification techniques applied to public technology: predictions compared with observed use. Appl. Ergon. 27, 119–131.

Bagnara, S., Di Martino, C., Lisanti, B., Mancini, G., Rizzo, A., 1989. A Human Taxonomy Based on Cognitive Engineering and on Social and Occupational Psychology. European Commission JRC, Ispra, Italy.

Berliner, D.C., Angelo, D., Shearer, J., 1964. Behaviours, measures and instruments for performance and evaluation in simulated environments. Presented at the Symposium on the Quantification of Human Performance. Albuquerque, New Mexico, August 17–19.

Civil Aviation Authority (CAA), 1996. Aircraft Proximity Reports Airprox (C)—(Controller Reported) Review Period—October 1995 to May 1996, Vol. 10. Civil Aviation Authority, UK.

Dougherty, E.M., 1993. Context and human reliability analysis. Reliability Eng. System Saf. 41, 25–47.

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, Tennessee.

Evans, A.E., Basra, G., Megaw, E., 1998. Generation of Human Error Probability Data for Air Traffic Control: a Feasibility Study. NATS R&D Report 9846. National Air Traffic Services, London.

Evans, A., Slamen, A., Shorrock, S.T., 1999. Use of human factors guidelines and human error identification in the design lifecycle of NATS future systems. Proceedings of the FAA/EUROCONTROL Technical Interchange Meeting. Toulouse, France, April 27–29 1999.

Flach, J.M., 1995. Situation awareness: proceed with caution. Hum. Factors 37 (1), 149–157.

Fleishman, E.A., Quaintance, M.K., 1984. Taxonomies of Human Performance: the Description of Human Tasks. Academic Press, London.

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

Hollnagel, E., 1998. CREAM: Cognitive Reliability Error Analysis Method. Elsevier, London.

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

Hollnagel, E., Cacciabue, P.C., 1991. Cognitive modelling in system simulation. Proceedings of the Third European Conference on Cognitive Science Approaches to Process Control. Cardiff, September 2–6 1991.

Hollnagel, E., Marsden, P., 1996. Further Development of the Phenotype-Genotype Classification Scheme for the Analysis of Human Erroneous Actions. European Commission JRC-ISIS, Varase, Italy.

Isaac, A., Kirwan, B., Shorrock, S.T., 1999. Human error in European Air Traffic Management: the HERA Project. Paper presented at the Third International Workshop on Human Error, Safety, and System Development—HESSD'99. Liege, Belgium, June 7–8 1999.

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

Isaac, A., Shorrock, S.T., Kirwan, B., 2002. Human error in European air traffic management: the HERA project. Reliability Eng. System Saf. 75 (2), 257–272.

Johnson, C., 1999. Why human error modeling has failed to help systems development. Interacting Comput. 11, 517–524.

Jones, D.G., Endsley, M.R., 1996. Sources of situation awareness errors in aviation. Aviat. Space Environ. Med. 67 (6), 507–512.

Josefsson, B., 2001. Personal communication.

Kirwan, B., 1992a. Human error identification in human reliability assessment. Part 1: overview of approaches. Appl. Ergon. 23 (5), 299–318.

Kirwan, B., 1992b. Human error identification in human reliability assessment. Part 2: detailed comparison of techniques. Appl. Ergon. 23 (6), 371–381.

Kirwan, B., 1994. A Guide to Practical Human Reliability Assessment. Taylor & Francis, London.

Kirwan, B., 1998a. Human error identification techniques for risk assessment of high risk systems—Part 1: review and evaluation of techniques. Appl. Ergon. 29 (3), 157–177.

Kirwan, B., 1998b. Human error identification techniques for risk assessment of high risk systems—Part 2: towards a framework approach. Appl. Ergon. 29 (5), 299–318.

Kontogiannis, T., 1997. A framework for the analysis of cognitive reliability in complex systems: a recovery centred approach. Reliab. Eng. System Saf. 58, 233–248.

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

Norman, D.A., 1981. Categorisation of action slips. Psychol. Rev. 88, 1–15.

Norman, D.A., 1986. Cognitive engineering. In: Norman, D.A., Draper, S.W. (Eds.), User Centered System Design: New Perspectives on Human-Computer Interaction. Lawrence Erlbaum, Hillsdale, NJ, pp. 31–62.

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.

Payne, D., Altman, J., 1962. An Index of electronic equipment operability: report of the development of the American Institute of Research. Report No. AIR-C-43-1/162, Pittsburgh, PA.

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

Pew, R.W., Miller, D.C., Feehrer, C.S., 1981. Evaluation of Proposed Control Room Improvements Through Analysis of Critical Operator Decisions. EPRI-1982. Electric Power Research Institute, Palo Alto, CA.

Profit, R., 2001. Keynote address. Proceedings of IBC's Fourth Annual Conference on Aviation Safety Management. London, May 31 to June 1 2001.

Rasmussen, J., 1982. Human errors: a taxonomy for describing human malfunction in industrial installations. J. Occup. Accidents 4, 311–335.

Rasmussen, J., Pedersen, O.M., Carnino, A., Griffon, M., Mancini, C., Gagnolet, P., 1981. Classification Scheme for Reporting Events Involving Human Malfunctions. Risø-M-2240, SINDOC(81)14. Riso National Laboratories, Roskilde, Denmark.

Rasmussen, J., 1986. Information Processing and Human–Machine Interaction: An approach to Cognitive Engineering. North-Holland, New York.

Reason, J., 1979. Actions not as planned. The price of automatization. In: Underwood, G., Stevens, R. (Eds.), Aspects of Consciousness, Vol. 1. Psychological Issues, Wiley, London.

Reason, J.T., 1987a. A framework for classifying errors. In: Rasmussen, J., Duncan, K.D., Leplat, J. (Eds.), New Technology and Human Error. Wiley, Chichester, UK.

Reason, J.T., 1987b. Generic error modelling system: a cognitive framework for locating common human error forms. In: Rasmussen, J., Duncan, K.D., Leplat, J. (Eds.), New Technology and Human Error. Wiley, Chichester.

Reason, J., 1990. Human Error. Cambridge University Press, Cambridge.

Rizzo, A., Ferrante, D., Bagnara, S., 1995. Handling human error. In: Hoc, J.M., Cacciabue, P.C., Hollnagel, E. (Eds.), Expertise and Technology: Cognition and Human Computer Interaction. Lawrence Erlbaum Associates, NJ.

Roske-Hofstrand, R., Murphy, E.D., 1998. Human information processing in air traffic control. In: Smolensky, M.W., Stein, E.S. (Eds.), Human Factors in Air Traffic Control. Academic Press, San Diego, CA.

Rouse, W.B., Rouse, S.H., 1983. Analysis and classification of human error. IEEE Trans. Systems Man Cybern. SMC-14, 539–549.

Sellen, A.J., 1994. Detection of everyday errors. Appl. Psychol.: Int. Rev. 43 (4), 475–498.

Shorrock, S.T., 1997. The development and evaluation of TRACEr: a technique for the retrospective analysis of cognitive errors in air traffic control. M.Sc. (Eng) Thesis, The University of Birmingham, September 1997.

Shorrock, S.T., Kirwan, B., 1999. The development of TRACEr: a technique for the retrospective analysis of cognitive errors in ATC. In: Harris, D. (Ed.), Engineering Psychology and Cognitive Ergonomics, Vol. 3. Aldershot, UK, Ashgate.

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

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

Swain, A.D., Guttmann, H.E., 1983. 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.

UK Airprox Board, 2000. UKAB Report No. 3—July 1999 to December 1999.

Vanderhaegen, F., 2001. A non-probabilistic prospective and retrospective human reliability analysis method - application to railway system. Reliab. Eng. System Saf. 71 (1), 1–13.

Wickens, C., 1992. Engineering Psychology and Human Performance, 2nd Edition. Harper-Collins, New York.

Wiegmann, D.A., Shappell, S.A., in press. Human error analysis of commercial aviation accidents: application of the human factors analysis and classification system (HFACS). Aviat. Space Environ. Med.

Wiener, E.L., Kanki, B.G., Helmreich, R.L. (Eds.), 1996. Cockpit Resource Management. Academic Press, New York.

Wioland, L., Amalberti, R., 1998. Human error management: towards an ecological safety model. A case study in an Air Traffic Control microworld. Proceedings of the Ninth European Conference on Cognitive Ergonomics. Limerick, Ireland, August 24–26 1998.

Zapf, D., Maier, G.W., Rappensperger, G., Irmer, C., 1994. Error detection, task characteristics, and some consequences for software design. Appl. Cognitive Psychol.: Int. Rev. 43 (4), 499–520.