Ergonomics

ISSN: 0014-0139 (Print) 1366-5847 (Online) Journal homepage: https://www.tandfonline.com/loi/terg20

What is on your mind? Using the perceptual cycle model and critical decision method to understand the decision-making process in the cockpit

Katherine L. Plant & Neville A. Stanton

To cite this article: Katherine L. Plant & Neville A. Stanton (2013) What is on your mind? Using the perceptual cycle model and critical decision method to understand the decision-making process in the cockpit, Ergonomics, 56:8, 1232-1250, DOI: 10.1080/00140139.2013.809480

To link to this article: https://doi.org/10.1080/00140139.2013.809480

Published online: 26 Jun 2013.

Ergonomics, 2013
Vol. 56, No. 8, 1232–1250, http://dx.doi.org/10.1080/00140139.2013.809480

What is on your mind? Using the perceptual cycle model and critical decision method to understand the decision-making process in the cockpit

Katherine L. Plant* and Neville A. Stanton

Transportation Research Group, Faculty of Engineering and Environment, University of Southampton, Southampton SO17 1BJ, UK

(*Received 13 October 2011; final version received 17 May 2013*)

Aeronautical decision-making is complex as there is not always a clear coupling between the decision made and decision outcome. As such, there is a call for process-orientated decision research in order to understand *why* a decision made sense at the time it was made. Schema theory explains how we interact with the world using stored mental representations and forms an integral part of the perceptual cycle model (PCM); proposed here as a way to understand the decision-making process. This paper qualitatively analyses data from the critical decision method (CDM) based on the principles of the PCM. It is demonstrated that the approach can be used to understand a decision-making process and highlights how influential schemata can be at informing decision-making. The reliability of this approach is established, the general applicability is discussed and directions for future work are considered.

Practitioner Summary: This paper introduces the PCM, and the associated schema theory, as a framework to structure and explain data collected from the CDM. The reliability of both the method and coding scheme is addressed.

Keywords: critical decision method; perceptual cycle model; schema theory; aeronautical decision-making

1. Introduction

The purpose of this paper is to introduce the perceptual cycle model (PCM; Neisser 1976), and its incorporation of schema theory, as a way of examining the process of aeronautical decision-making. A qualitative method is presented in which the PCM is used as a coding scheme to analyse data elicited from the critical decision method (CDM; Klein, Calderwood, and Macgregor 1989). In previous work, Plant and Stanton (2012a) presented the PCM as a way to structure accident report data in order to infer causal explanations for erroneous decision-making. This study seeks to explore whether the PCM can be applied to interpret CDM data and how reliable this approach is. After introducing the relevant theoretical and methodological perspectives, the remainder of the paper is split into three parts: Section 3 analyses a critical incident in relation to the PCM, Section 4 addresses the reliability of this method and Section 5 provides a general discussion of the practical implications and applications of this approach.

1.1. Aeronautical decision-making

Aeronautical decision-making is a form of naturalistic decision-making (NDM; Klein, Calderwood, and Macgregor 1989) in which decision makers have domain expertise and make decisions in contexts which are usually characterised by limited time, goal conflicts and dynamic conditions. Aeronautical decision-making is an important area of research as McFadden and Towell (1999) have argued that pilot judgement in decision-making and the handling of emergency situations are usually the deciding factor as to whether an incident will become an accident. Decisional errors have consistently been found to account for a high proportion of pilot error (Diehl 1991; Orasanu and Martin 1998; Shappell and Wiegmann 2009). However, NDM is complex because there is often no clear standard of correctness and there is a loose coupling between event outcome and decision process, so that outcomes cannot be used as reliable indicators for the quality of the decision (Orasanu and Martin 1998). For example, good decisions may be overwhelmed by events outside of the decision maker's control, resulting in a bad outcome. With hindsight, it is easy to conclude that a poor decision was made. When examining NDM, it is, therefore, important to focus on the adequacy of the decision-making *process*, rather than using the outcome as a benchmark for 'decision correctness'. To do this, it is necessary to understand how and why the actions and assessments undertaken by an operator made sense to them at the time (Dekker 2006). The cognitive origins and processes that underlie decision-making have traditionally been overlooked (Maurino 2000; Hobbs and Williamson 2002). We propose that

*Corresponding author. Email: K.Plant@soton.ac.uk

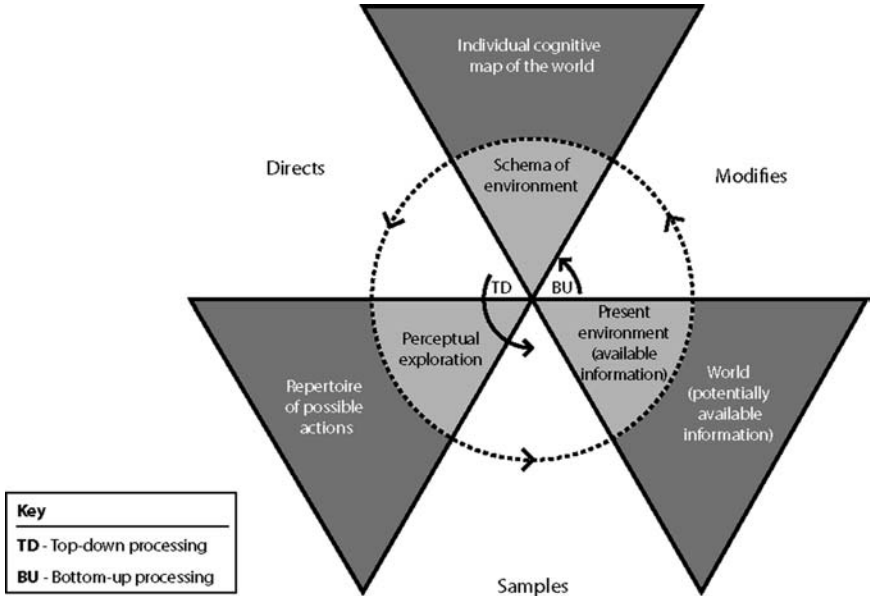

Figure 1. The PCM (*Source*: Adapted from Neisser (1976)).

Neisser's (1976) PCM is a suitable framework to model the cognitive basis of decision-making and with that comes the ability to identify areas vulnerable to potential mistakes and establish ways to mitigate these before errors occur.

1.2. The PCM

The PCM (Neisser 1976) is based upon the idea of a reciprocal, cyclical, relationship between operator and environment. The perceptual cycle models interaction between person and world, with heavy emphasis on the role of schemata (see Section 1.3). As illustrated in Figure 1, Neisser presented the view that human thought is closely coupled with a person's interaction in the world, both informing each other in a reciprocal, cyclical relationship. World knowledge (schemata) leads to the anticipation of certain types of information (top-down processing, TD in Figure 1); this then directs behaviour (action) to seek out certain types of information and provides a way of interpreting that information (bottom-up processing, BU in Figure 1). The environmental experience (world) can result in the modification and updating of cognitive schemata and this in turn influences further interaction with the environment.

Smith and Hancock (1995) have argued that the usefulness of the PCM explanation lies in the interaction between operator and environment, rather than considering the two separately. They used the PCM as the foundation for their perspective of situation awareness (SA), arguing that it is impossible to comprehend SA without an understanding of the interaction between operators and their task environment. The same view can be taken when comprehending decision-making; without an explanation of the interaction between the operator and their work context, a causal understanding of why decisions were made is not achievable (Dekker 2006). This paper describes a method that has been used to unearth the perceptual cycle process of a decision maker when dealing with a critical incident.

1.3. Schema theory

Schemata are a fundamental part of the PCM and, therefore, it is important to define what is meant by the term schema in relation to the umbrella term of mental representations. Schemata are internal knowledge structures that are based on similar experiences which capture the common features of this experience (Lieberman 2012). Bartlett (1932), who is generally considered the founder of schema theory, described schemata as mental templates which are neither entirely new behaviour nor an exact replication of old behaviour. Richardson and Ball (2009) stated that mental models are the most commonly cited type of mental representation. Johnson-Laird (1983) argued that mental models are constructed in working memory at the time of input which gives rise to visualised images. As such, mental models are more temporary than schemata; the latter are used to denote the representation of knowledge in long-term memory. The schemata of long-term memory are thought to inform the mental models held in working memory (Richardson and Ball 2009). Similarly, Wilson and Rutherford (1989) argued that activated schemata are utilised by mental models in a dynamic manner.

Schemata are conceptualised as having ‘slots’ which are used to structure the information linked to them. Although schemata represent abstract concepts, they are built from specific instances and allow abstract knowledge to be derived at the time of retrieval by sampling from domain-specific instances. As described in the PCM, when an environmental experience is encountered, relevant experiences (schemata) are retrieved to help develop an appropriate response. In terms of decision-making, there is general consensus that NDM is strongly schema driven (Klein 1993). For example, Elliott (2005) has argued that schemata form the lenses through which the decision maker views the problem. Regardless of the domain, decision makers will use the schemata they possess to make decisions. Often the use of schemata in decision-making is advantageous; they act as natural standard operating procedures (SOPs) to direct decision makers to make appropriate responses to environmental stimuli based on previously successful experiences. Problems arise, however, when the activated schema is inappropriate for the current situation. In this vein, Norman (1981) applied schema theory to the study of error by suggesting that schema-triggering situations may be wrongly interpreted, leading to inappropriate responses. The role of schemata in decision-making research has been widely demonstrated, in both ‘normal’ situations (e.g. Walker, Stanton, and Salmon 2011) and erroneous situations (e.g. Stanton and Walker 2011; Plant and Stanton 2012a).

Studying the role of schemata in decision-making comes from the perspective of the individual level of analysis, i.e. schemata are unique individual cognitive structures. In Ergonomics research, there has been notable criticism about the reductionist nature of the individual level of analysis, when compared with more systemic levels of analysis (Reason 2000). However, we argue here that as human operators operate at the ‘sharp end’ of a system, there is still a need to understand the cognitive processes that underpin decision-making. The PCM places the operator (with their individual schema) into the environment in which the decisions take place and in doing so recognises that cognition is distributed within a system (Rafferty, Stanton, and Walker 2010). As such, the PCM accounts for the operator–environment interactions that occur. Furthermore, it is widely acknowledged that people make decisions according to the principle of local rationality, i.e. behaviour (which in hindsight appears erroneous) is rational when viewed from the perspective of the current work context (Reason 1990; Dekker 2006).

Schema theory, however, is not without its criticisms (see Plant and Stanton 2013, for a review). One of the notable issues with schemata, as acknowledged by Walker, Stanton, and Salmon (2011), is that ‘*gaining insight into mental representations . . . is experimentally and conceptually challenging*’ (p. 879). Plant and Stanton (2013) discussed how the internal nature of schemata means they cannot be directly measured, but only inferred through the empirical consequences of them and qualitative methods are likely to be more suitable for this. The difficulties of eliciting and representing schemata have previously been a barrier to the application and utility of schema theory. From the argument presented, it would appear that there is value in exploring the principles of the PCM, in which schema theory is a central tenet, as a way to structure and understand the process of aeronautical decision-making. The CDM has been utilised here as the qualitative data collection method in order to explore the role of the PCM in aeronautical decision-making. An explanation of the CDM follows.

1.4. The CDM

1.4.1. Description

The CDM (Klein, Calderwood, and Macgregor 1989) is one of the most commonly used cognitive task analysis methods (Stanton et al. 2005). The CDM achieves knowledge elicitation through the use of cognitive probes as a tool for reflecting on strategies and reasons for decisions during non-routine situations. Since its development, the CDM has been extensively used in a variety of domains including emergency dispatch management (Wong, Sallis, and O’Hare 1997), critical care nursing (Crandall and Gretchell-Reiter 1993) and aviation (O’Hare et al. 1998). Specifically, this method focuses on eliciting knowledge for behaviours Klein, Calderwood, and Clinton-Cirocco (1986) classed as ‘recognition-primed decisions’, i.e. decisions for which alternative actions are derived from a recognition of critical information from the environment and prior knowledge. The emphasis on both prior knowledge and information from the environment makes the CDM a potentially suitable method to elicit data that can be analysed from the perspective of the PCM (which also places emphasis on prior knowledge, i.e. schemata, and the environment, i.e. world information). Furthermore, the CDM was selected as the data collection method because it is a theory-driven approach based on the assumption that expertise emerges most clearly during non-routine events (Klein, Calderwood, and Macgregor 1989). The schema theory literature suggests that the use of schemata is most evident in experts as opposed to novices, as experts have more experiences from which to assimilate schemata (Elio and Scharf 1990).

1.4.2. Procedure

The CDM achieves knowledge elicitation by asking people to discuss previous incidents they were involved with. The process of eliciting information is via cognitive probes in a retrospective semi-structured interview. Crandall, Klein, and Hoffman (2006) described the four phases for conducting a CDM interview: (1) incident identification, (2) timeline construction, (3) deepening probes and (4) 'what if' queries. Appendix provides an expanded explanation of each phase and a full list of the deepening probes. A shorter version of the method is permitted when time with experts is limited. For this, Crandall, Klein, and Hoffman (2006) have suggested that the probes are asked in relation to the whole incident as opposed to each phase of the incident. It is acknowledged that the probes have been modified over the years and researchers are encouraged to modify the list as necessary for their individual research projects (Klein and Armstrong 2005; Crandall, Klein, and Hoffman 2006). In the light of this research project, five additional probes were added that expanded certain areas of the original CDM probes to increase their relevance to the elicitation of schemata (i.e. drawing out the role of experience and expectations). These additional probes are shown in italics in Appendix. Interested readers are directed to the following texts for the full CDM procedure: Crandall, Klein, and Hoffman 2006; Klein and Armstrong 2005; Stanton et al. 2005.

1.4.3. Evaluation of the CDM

The main limitation with the CDM is its reliance on verbal reports (Klein, Calderwood, and Macgregor 1989). How far a verbal report accurately represents the cognitive processes of the decision maker is questionable (Stanton et al. 2005). People can misrepresent their own decision-making strategies and goals, especially as this is reported retrospectively and there are known issues associated with memory alteration and decay (Klein and Armstrong 2005). Klein, Calderwood, and Macgregor (1989) argued that it is essential to acknowledge the biases associated with a method of this nature and work to reduce them. As such, to increase confidence in the reliability of this method, a retest reliability study was undertaken. The result of this is reported in Section 4. This approach, however, is not always practical and, therefore, inter-rater reliability is usually calculated in relation to coding schemes used to analyse CDM data. For example, Klein, Calderwood, and Macgregor (1989) assessed the inter-rater reliability of the method in terms of how reliably a decision point could be identified from the unstructured portion of the interview. Percentage agreement between the two coders was over 80%, suggesting that decision points can be reliably identified with CDM data. Section 4 describes an inter-rater reliability study undertaken to assess the reliability of the PCM coding scheme used to analyse the CDM data in this study. Although there are issues associated with the CDM, the benefits are often seen to outweigh them (Klein, Calderwood, and Macgregor 1989). For example, the retrospective nature of the method means that the events of interest have occurred, so there is no need to create artificial simulations that are limited in contextual richness, nor is there a need to wait for non-routine events to occur (Klein, Calderwood, and Macgregor 1989). Klein, Calderwood, and Macgregor (1989) have argued that the semi-structured nature of the interview results in less time to gather relevant information, but retains the freedom to explore interesting data. The standardised probes also result in more reliable data.

2. Method

2.1. Methodological perspective

The main impetus for this paper was to explore a method that allows the role of the perceptual cycle in decision-making to be understood. As discussed in Section 1.1, it is essential to understand the decision-making *process* rather than just the decision outcome. From the description of the perceptual cycle and the role of schemata in this process (Sections 1.2 and 1.3, respectively), the PCM and associated schema theory appear to be useful theoretical perspectives from which to study the decision-making process. The issue of inferring schemata has been raised (Section 1.3) and the use of the CDM justified (Section 1.4).

The method undertaken was based upon the principles of thematic analysis by which text data are classified into meaningful themes. Here, deductive thematic analysis was conducted, in which themes (patterns in the data) are generated from existing theory (Boyatzis 1998). In accordance with the objectives of this paper, the coding scheme was based on the categories of the PCM and developed in line with Boyatzis' (1998) five criteria of how to structure a meaningful code. Table 1 provides the criteria along with the coding scheme used in this analysis.

This form of qualitative data analysis is open to criticism about the reliability of the results. As Klein, Calderwood, and Macgregor (1989) acknowledged, the CDM is a qualitative data gathering technique and such data cannot escape subjective interpretation. Assessing the reliability of the data interpretation, however, can go some way to making the claims more objectively verifiable. Both inter-rater reliability and intra-rater reliability are critical in thematic analysis, where computing it combines the richness of qualitative information with the precision of quantitative methods (Boyatzis 1998). Singletary (1993) argued that '*if the coding is not reliable, the analysis cannot be trusted*' (p. 294).

Table 1. The criteria for a meaningful code (Boyatzis 1998) and the PCM coding scheme (definitions adapted from Neisser's (1976) PCM explanation).

Criteria for a meaningful code (Boyatzis 1998)	Coding scheme used in this study		
1. A label/name	Schema	Action	World
2. Definition of what the theme concerns	Mental structures held by individuals that organise their representations of the world. Schemata are heavily influenced by experiences and expectations	The process or statement of doing something, or the intention to do something	Externally available information in the world (environment)
3. Description of how to know when the theme occurs	Statements relating to the use of prior knowledge and experience, i.e. things based on experience, expectation or 'knowing' things (this could be implied information through the discussion of training and/or SOPs)	Statements of doing an action or discussion about potential actions that could be taken	Statements relating to potential or actual information existing in the world (environment). Can be physical things, conditions or states of being
4. Description of any exclusions	References to mental information made in the context of an action should be coded as action. For example '*talking about known training procedures*' should be coded as action, even though the statement refers to prior knowledge of training procedures	Only code as action if the statement is referring to explicit actions made by the pilot/crew. For example, the statement '*light came on*' is an action but in this context this statement provides the pilot with information about the state of something in the world (i.e. light is on) and would, therefore, be coded as 'world'. The statement '*I turned the light on*' is coded as 'action'	n/a
5. Example	'*my expectation was that the engine would take a while to start in the rain*' '*. . . knowledge of training helped . . .*' '*. . . I had no other experience to base it on*'	'*I turned on the engine*' '*. . . we'll turn it off and on again . . .*' '*. . . Cancel the training and head back to base . . .*'	'*Caution light came on*' '*it was raining*' '*. . . the primary flight display screen was on . . .*'

2.2. Study participant and procedure

The CDM procedure outlined in Section 1.4.3 was used with a Search and Rescue helicopter pilot. The first interview (T1) occurred 9 months after the critical incident (T0, August 2009). The second CDM interview (T2, conducted to assess the test–retest reliability of the CDM, see Section 4.1) took place 25 months after T1. The participant was a male helicopter pilot (aged 34 years at T1). He was voluntarily recruited through an advert for participants placed on the British Helicopter Association website. At T1 and T2, a face-to-face interview was conducted which lasted for approximately 1 hour. At T1 the participant had approximately 2500 flying hours. The incident occurred when flying an AW139, at T0 the pilot had approximately 300 hours on this type of aircraft. At the time of the incident, the helicopter was being flown for a Search and Rescue training exercise. This is a domain that has recently been used by other researchers, such as Baber et al. (2013), as it eloquently exemplifies the NDM environment in a rotary wing context. Ethical permission for this and the subsequent studies reported in this paper was granted by the Research Ethics Committee at the University of Southampton.

The nature of the interview was explained to the participant, i.e. he would be required to recall an incident and the interview would last approximately 1 hour depending on length of answers provided. A critical incident was defined as being 'a non-routine or un-expected event that was highly challenging and involved a high workload'. Due to time constraints, a shorter version of the CDM was conducted. The CDM interviews were audio recorded and transcribed. In accordance with guidelines on qualitative data analysis, text was chunked into meaningful segments of approximately one sentence or less in length (Strauss and Corbin 1990).

2.3. Methodological questions to answer

The theoretical arguments that have been presented have raised several methodological questions that this paper will seek to answer.

1. Is the PCM able to account for the decision-making process of a pilot dealing with a critical incident, using data collected from the CDM?
2. How reliable is the CDM in terms of test–retest reliability?
3. How reliable is the method of qualitative data analysis in terms of inter- and intra-rater reliability?

3. Incident analysis

3.1. Incident synopsis

The pilot interviewed was the Pilot Not Flying (PNF) on the day of the incident and was, therefore, responsible for inputting data into the aircraft systems and fault diagnosis. Other crew on board included the Pilot Flying (PF), who had primary responsibility for the aircraft, and the rear winch crew. The PNF stated that the winch crew would have been aware of the situation via the intercom, but they were not involved in the diagnostic decision-making process. It should be acknowledged that the data presented in this paper are from the perspective of the PNF as the level of analysis was defined at the individual. Stanton et al. (2010b) have discussed the debate that exists over establishing the appropriate level of analysis in the context of SA research. The authors have argued that as long as the level of analysis is defined and declared, then research at the levels of the individual, the whole system and all the areas in-between have a valid place in Ergonomics research. Owing to this being an individual level of analysis, the roles of the additional crew members were not explicitly addressed. Issues associated with, what is likely to be, a composite account of the incident are addressed in the discussion (Section 5). Below is a synopsis of the incident amalgamated from the CDM interviews at T1 and T2:

> ... Finished winching in the English Channel and were then going to go to cliff winch ... put in the waypoint, or what I thought was the waypoint, into the navigation system ... I typed in the three digit code ... then we looked and all the screens, well the 4 primary screens, went blank. There was no navigation information and no primary flight information, except for the secondary back up system which gives an attitude, a speed and a height. It was a clear day with good conditions in uncontrolled airspace. Once we got over the initial shock ... the immediate thought was 'what's it done now?' blaming the systems rather than any action I made ... by this stage we were expecting electronic problems, can guarantee one on most flights. We initially started looking for circuit breakers, to look if any had popped ... lots of electronics on this aircraft, usually the case of finding the right circuit breaker. None had popped ... went to the flight cards ... there wasn't one for that, it is not the sort of thing that is expected ... we tried turning up the brightness of the screens to see if that made a difference. Nothing else we could do ... agreed we would head back to base to get the aircraft serviceable ... headed north, we were in the Channel so if you head north you will hit the mainland somewhere, we knew once we picked that up we were okay. After the event we realised I had reverted to the old waypoint for the previous aircraft we flew. Shut the engine down and started it back up and the waypoint had cleared and the screens were back.

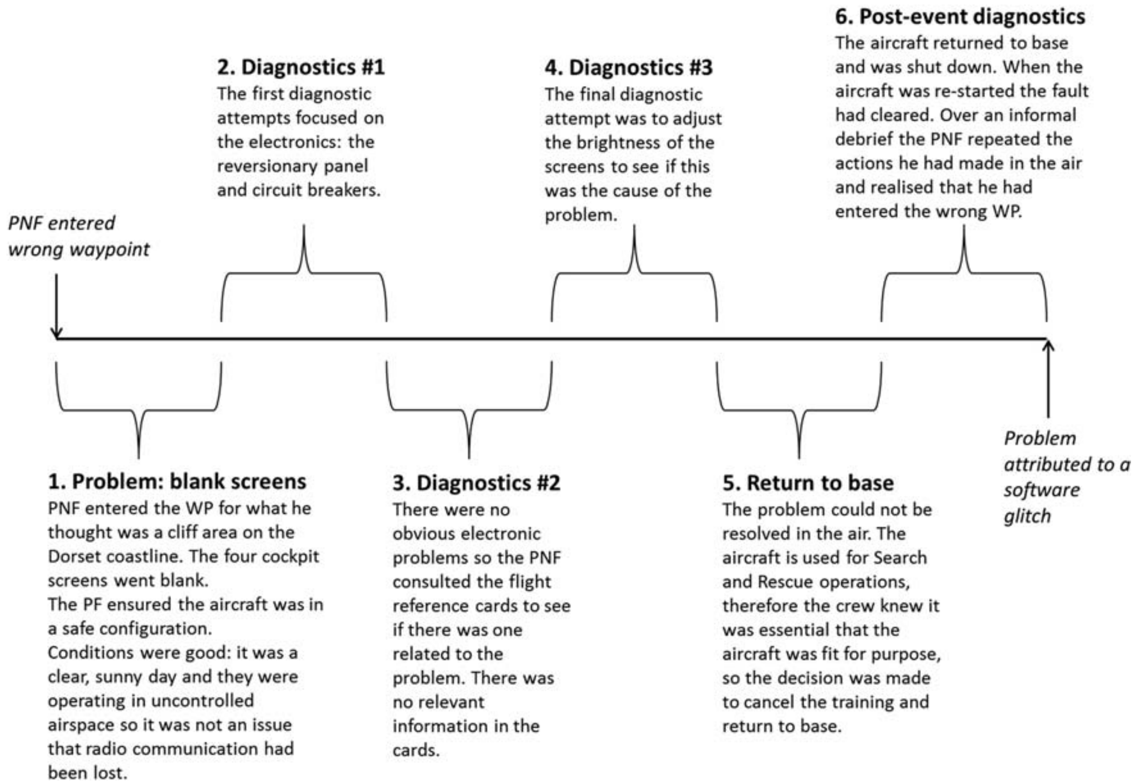

Figure 2. Timeline of the critical incident.

3.2. Thematic analysis of critical incident

The coding scheme presented in Table 1 was used to analyse the data obtained from the CDM interviews. In line with the CDM procedure described in Section 1.3, a timeline of events was created and is presented in Figure 2.

For each event in the timeline, the relevant data were coded with the PCM coding scheme to understand the perceptual cycle the pilot engaged in when dealing with the critical incident. Table 2 presents the data from each phase of the incident, organised by each element of the PCM. This is illustrated in Figure 3.

3.3. Incident summary

The fundamental principle behind Neisser's (1976) PCM is that although it is structured, information does not move in a simple linear flow but rather in a cyclical, interactive, process. This is illustrated by the helical representation in Figure 3. The Schemata held by a person direct their actions and exploration in the world, i.e. expectations lead to anticipating certain types of information in the environment which is actively sought out. Action requires information if it is to be carried out effectively and that produces more information for the perceiver. The state of the world is then encompassed back into, and can have a modifying effect on, the individual's schemata and thus future interactions in the world. It is clear that in this incident the pilot's perspective was compounded by a host of issues that are associated with the schemata he held about the aircraft he was co-piloting. The pilot's assumption about the problem was based on his experience and expectations of the aircraft's electrical system, which led the pilot to engage with various diagnostic attempts in order to solve the problem of the blank screens. It is during this diagnostic phase that the pilot's reliance of his schemata for the situation is evident. A schema is an organised mental pattern of thoughts or behaviours to help organise world knowledge (Neisser 1976). Decision-making literature suggests that decision makers make an initial assessment of a situation by looking for familiar patterns or prototypes (Klein, Calderwood, and Clinton-Cirocco 1986). Among other things, this strategy saves cognitive resources by generating appropriate options and responses. On the basis of the pilot's experience of the electrical glitches often encountered in the aircraft, the pilot was expecting this fault to be of similar nature, i.e. he was looking for familiar patterns.

Table 2. The perceptual cycle process for each phase of the critical incident.

Incident phase	Code	Description
Onset of problem (blank screens)	W_1	During a winch training exercise the PNF was required to enter a waypoint (WP) into the navigation system
	S_2	PNF held a schema for the WP names developed from aircraft he had previously flown: '*I reverted to the old WP for the S61 ... we have WP lists but was doing it from memory*'
	A_3	PNF keyed in the wrong WP (this resulted in the screens going blank)
	Misc.	The PF ensured the aircraft was in a safe configuration
Diagnostics (electronics)	W_4	Blank screens, aircraft in safe configuration, AW139 has more electronics than previous aircraft
	S_5	In his interview the PNF stated; '*we get a lot of electronic problems ... can guarantee one on most flights, through our experience we understand which ones are important ... aircraft often has little glitches*'. This demonstrates that the PNF held a schema for the situation that the problem was likely to be related to the electronics: '*... knew from ground school that there were two separate* [electronic] *systems, going to two separate screens*'
	A_6	'*I played around with the* [reversionary] *panel to see if that made a difference, manually applied the feeds from the other side*'
	W_7	Screens were still blank
	S_8	'*the training we now do in the simulator is to check the circuit breakers*' The PNF held a schema developed from training experiences
	A_9	Looked to see whether any circuit breakers had popped
Diagnostics (flight reference cards)	W_{10}	Screens still blank, not an immediately obvious electronic problem (e.g. no popped circuit breakers)
	S_{11}	Schemata acquired through training
	A_{12}	Checked the flight reference cards to see whether there was one relevant to this situation
Diagnostics (screen brightness)	W_{13}	Screens still blank, no relevant flight reference cards
	S_{14}	'*We know that the screens* [brightness] *are sometimes turned down for a night flight ... they would appear blank in daylight*'. The pilot stated that this was highly unlikely to be the cause of the problem as the screens had been on previously, but he was exhausting all other options
	A_{15}	'*I tried playing with the switch* [for screen brightness] *to see if that made a difference*'
Decision (return to base and shut down)	W_{16}	Screens blank, diagnostic attempts had not located the problem, aircraft was flying fine, it was a clear day and they were in uncontrolled airspace
	S_{17}	Both pilots were familiar with the location they were in, the PNF stated: '... in the English channel, south of the Isle of Wight, if head north I know we'll hit the mainland and be okay ... almost certainly knew it was a glitch rather than a major system fault'
	A_{18}	The PNF assisted the co-pilot when returning to base (e.g. eyes out navigation)
	Misc.	A collective decision was made to return to base, the aircraft was flown back by the Pilot Flying
Post-event diagnostics	W_{19}	Base at base with the engineers
	S_{20}	Talking about what had happened, '*realised I put in the WP for what I believed to be Anvil Point, was used to flying the S61 with a different set of WPs ... put in what I thought was the correct WP*' [the WP had changed on the new aircraft]
	A_{21}	Told the engineers what I had done
	Misc.	When the aircraft was shut down and powered back up the problem cleared and the screens returned

The error made by the pilot of entering the wrong waypoint (WP) can also be explained by schema theory. Schemata are knowledge structures based on a set of similar experiences and they capture the common features of this experience. Minsky (1975) discussed frames as a form of static schema and argued that they have slots to be filled with information. These slots have a default value assumed, i.e. the most common representation is used if alternative information is not provided. In the case of entering the wrong waypoint, the pilot reverted to the default value for his waypoint schema, i.e. the waypoint that was often used (correctly) in the previous aircraft but was incorrect for his current aircraft. Incidentally, the problem was caused by a software glitch. In the current aircraft, which was relatively new to the pilot, the entered waypoint was assigned to a location in New Zealand. The software glitch resulted in the screens being shut down if the waypoint location was over 10,000 miles away.[1]

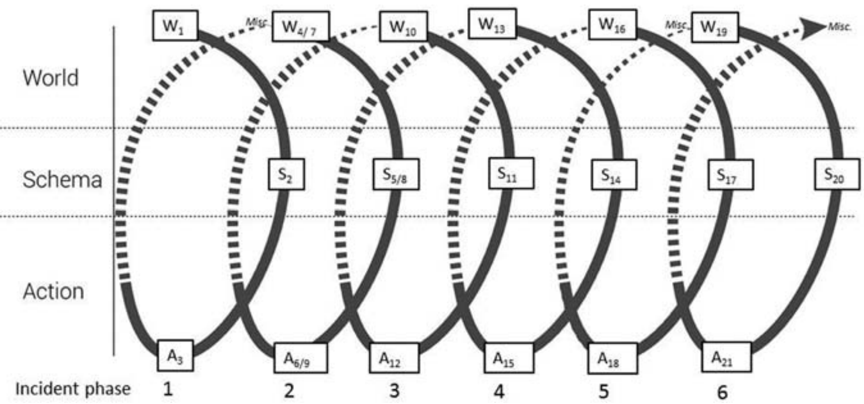

Figure 3. The progression of the incident through the perceptual cycle, descriptions for each annotation are listed in Table 2.

4. Tests of reliability

4.1. *Reliability of the CDM*

As detailed in Section 1.4.3, the reliability of the CDM has been questioned as the method relies on retrospective verbal reports; however, little appears in the literature assessing the retest reliability of the CDM. Taynor, Crandall, and Wiggins (1987) conducted one of the few studies available to determine the retest reliability of the CDM. Data collected from eight firefighter commanders at time intervals of 3 days, 3 months and 5 months from when the incident was first recalled were compared. Correspondence of information across the different reporting times averaged 86%. In this study, retest reliability was assessed with a time interval of over 2 years.

4.1.1. Procedure

To assess the retest reliability of the CDM, the procedure detailed in Sections 1.4.2 and 2.2 (also see Appendix) was used on two different occasions. At T2 the participant was asked to recall the incident as if the researcher had never heard about it. The CDM interview at T1 was broken down into 135 text segments and the CDM interview at T2 consisted of 174 segments. The tabulated text segments were subjected to open coding, in which they were analysed to identify themes into which they could be grouped. Theme identification was assisted by the names of the phases provided by the participant in the timeline construction stage of the CDM interview. Constant comparison technique was used, whereby each text segment was compared with previous items to see whether the same or a different phenomenon was described. A second rater re-coded 20% of the text segments using the theme descriptions provided in Table 3.

4.1.2. Results

Percentage agreement between the criterion coder and inter-rater coder was calculated at 92%, well above the suggested threshold of 80% agreement (Jentsch and Bowers 2005). T1 and T2 data were compared in terms of text segments in each theme. This is what Hoffman, Crandall, and Shadbolt (1998) described as a moderate check of reliability, i.e. that the general gist and details of an incident are the same, as opposed to an exact check of reliability whereby the same details are recalled in the same order at different time intervals. The point of this retest study was to assess the reliability of the CDM; a highly reliable method will produce similar data at T1 and T2. Therefore, the area of relevance and interest in this retest study is in understanding the differences between data at T1 and T2. If data at T1 and T2 widely vary, then the reliability of the CDM will be questionable. Table 3 presents the theme name, theme description and any key differences between the two data-sets.

4.1.3. Preliminary discussion

The high level of inter-rater agreement (92%) between the criterion coder and inter-rater coder suggests that the themes used to classify the data were accurately representing what the data were describing. However, it is the data within each theme that are of interest in this retest study. Data obtained from the CDM interview at T1 and T2 were compared in terms of content within each theme. Table 3 details the themes identified in the data and where differences occurred. There was no

Table 3. Theme name, description and comparison of T1 and T2 data to assess the retest reliability of the CDM.

Theme	Description of theme	Differences between data at T1 and T2
Conditions	Text relating to environmental conditions, such as weather and time of day	None
Aircraft status	Text relating to the current status of the aircraft in terms of performance and configuration	None
Waypoint location	Text relating to waypoints, e.g. inputting waypoint data and the associated navigation system, the role of waypoints and waypoint names	Differences (see Table 4)
Location information	Text relating to area information, e.g. location of aircraft, potential location of aircraft and location of other key landmarks	None
Communication information	Text discussing communications between aircraft crew and external operators such as the coastguard. Includes text related to the radio and other communication systems	Differences (see Table 4)
Problem	Text relating to the onset of the problem (blank screens)	None
Immediate actions	Text discussing immediate actions taken by crew at the onset of the problem	Differences (see Table 4)
Decision	Text relating to the decisions taken by the crew, e.g. return to base	None
Options	Text discussing alternative options available to the pilot, other than the primary decision that was made	None
Diagnostics (general)	Text relating to general diagnostic information, i.e. discussions about the need to understand what was wrong, as opposed to any specific diagnostic attempts	None
Diagnostics (electronics)	Text about specific diagnostic attempts regarding the electronic systems	None
Diagnostics (flight reference cards)	Text about specific diagnostic attempts regarding the flight reference cards	Differences (see Table 4)
Diagnostics (circuit breakers)	Text about specific diagnostic attempts regarding the circuit breakers	Differences (see Table 4)
Diagnostics (screen brightness)	Text about specific diagnostic attempts regarding screen brightness	Differences (see Table 4)
Post-event diagnostics	Text discussing what happened after the incident was contained such as establishing the cause of the problem	None
Similar incidents	Text referring to whether similar incidents have arisen since the incident and how these were dealt with	Differences (see Table 4)
SAR role	Text discussing the SAR role of the helicopter and how this influenced the decision-making process, e.g. choosing to terminate training and the importance of having a serviceable aircraft	None

difference between data at T1 and T2 in 10 of the 17 themes identified; some of the differences identified in the remaining eight themes were marginal. Table 4 provides a more detailed description of the differences between data at T1 and T2.

As expected, this retest study demonstrated that there are some differences in data elicited by the CDM at T1 and T2. Literature on human memory would argue that differences occur because memory alters and is distorted with the passage of time (Lieberman 2012), and this would explain some of the differences that were found between data at T1 and T2. This is one of the primary concerns with using the CDM for data collection. Furthermore, schema theory suggests that schemata change and assimilate overtime based on recent world experiences and as such can have a modifying effect on subsequent actions and decisions, thus what is reported in a retrospective interview. As discussed in Table 4, this is evident in some of the differences found between the two data-sets. However, considering the amount of time elapsed between the two CDM interviews (2 years), the data are generally very similar and the salient points of the incident were reflected in both interviews: the pilot discussed inputting the wrong waypoint, losing all four cockpit screens, various diagnostic attempts that were undertaken and the decision to abort the training and return to base. This retest study was conducted nearly 3 years after the incident, a reliability study with a time elapse of this length has not, to our knowledge, been conducted with CDM data previously. The overall conclusion that we have drawn from this is that given such a considerable amount of time elapsed and during that time the pilot would have dealt with a number of other incidents, incident recall was remarkably robust. This can be attributed to the use of the structured CDM probes, as they allow the same questions to be asked and, therefore, similar responses are elicited.

Table 4. Differences between T1 and T2 CDM data.

Theme	T1 data	T2 data	Discussion/implication
Waypoint location	St Albans Head named as the waypoint	Anvil Point named as the waypoint	Both locations are cliff points on the Dorset coastline approximately 5 miles apart. In a post-interview debrief, the pilot was asked about the discrepancy. He stated that St Albans Head was the correct location. The T1 interview took place 9 months after the incident, whereas the T2 interview occurred 34 months after the incident. It is unsurprising that nearly 3 years later the pilot confused cliff names, in the elapsed time both locations were visited numerous times. This exemplifies that the retest study is seeking to establish what Hoffman et al. (1998) have described as 'moderate reliability', i.e. the general gist is the same (the training exercise occurred at a location on the Dorset coastline), but not necessarily presented in the same order or using exactly the same language
Communication information	*Theme not present*	Pilot mentioned that due to the loss of screens they were not able to tune the radios, but the intercom still worked so they were able to talk to the rear-crew	* The ascription of '*theme not present*' only occurred in T1 data, i.e. everything mentioned at T1 was also mentioned at T2, but T2 also had additional information. In terms of text segments, there were more data in T2 (174 segments) than in T1 (135 segments), so it is unsurprising that more themes are present in T2 data. Reasons for this are open to much speculation; for example, unbeknown to the researcher the pilot may have had more time available at T2 and, therefore, provided a more detailed interview. The process of having recalled the incident once, even though it was 2 years previously, may have also triggered subsequent memories and at T2 the pilot was more familiar with both the interview format and the researcher
Diagnostics (screen brightness)	*Theme not present*	Adjusting the brightness to see whether the screen had been turned down was discussed as a diagnostic attempt within the context of a more detailed discussion about the electronics of the aircraft at T2. For example, there was more mention of the way the reversionary panel worked	See *. **Furthermore, this difference is potentially an example of the pilot being more familiar with the researcher. Owing to resource constraints, the researcher conducting the interview was the same at T1 and T2. Although at T2 the pilot was instructed to recall the incident as if the researcher had never heard it, there may have been some unconscious processes at play by which the pilot felt he could go into more detail because 'the researcher had heard it before'. Furthermore, the pilot may have held an unconscious bias that after working in the field of Human Factors aviation research for a further 2 years, the researcher was more familiar with aviation terminology; hence, he went into more detail about the electronics
Similar incidents	*Theme not present*	The pilot discussed how the incident had occurred since and how it is now dealt with	See *. Furthermore, at T1 the pilot had not experienced the loss of the four screens before, at T2 it had happened to him on subsequent occasions;hence, the 'similar incident' theme occurring in T2 data but not in T1 data

Immediate actions	The pilot said the aircraft was put into a safe configuration, i.e. the autopilot was put on	The pilot stated the autopilot was already on and aircraft was already in a safe configuration	In terms of a moderate test of reliability, the data relating to immediate actions in the two data-sets are relatively similar, the phrases 'safe configuration' and 'autopilot' were used in both interviews. The discrepancy lies in the order of events but this can be explained by the role of the pilot: at the time of the incident, the pilot was PNF, i.e. he was the co-pilot, responsible for tasks such as navigation and data input. The action of ensuring that the aircraft was in a safe configuration would have been carried out by the PF, so it is unsurprising that a slight discrepancy arose in relation to actions that were not carried out by the PNF
Diagnostics (flight reference cards)	The pilot stated the crew looked in the flight reference cards	Less explicitly referenced, alluded to in relation to general emergency checks	A similar explanation to the screen brightness theme (**) can be applied here. Checking the flight reference cards (discussed in more detail at T1 than at T2) is a fundamental part of dealing with a critical incident. The pilot may have held an assumption that it was not necessary to go into this in as much detail at T2
Diagnostics (circuit breakers)	Much discussion about checking the circuit breakers	Less explicitly referenced, they were discussed in relation to other electronic diagnostic attempts	This can be explained in terms of schema theory; schema theory proposes that schemata are not static entities but instead dynamic and active. This is represented in the PCM, whereby world or environmental information can have a modifying effect on schemata (see Figure 1). In the post-interview debrief, the pilot stated that training around the time of the incident and T1 was extremely focused on the importance of checking circuit breakers; hence, it was fresh in his mind. This demonstrates how information in the world (e.g. recent training about circuit breakers) can have a modifying effect on the schemata held for a particular situation (e.g. checking circuit breakers to deal with the incident) and result in certain actions (e.g. identifying the diagnostic attempt of checking circuit breakers at the T1 interview)

4.2. Reliability of the PCM coding scheme

Aside from establishing the reliability of the CDM, the reliability of the PCM coding scheme also needs to be considered. This is what Boyatzis (1998) termed 'rater-expert reliability', whereby a set of correct (or more correct) answers to the judgement situation have already been assigned (by an expert). Additional coders are judged against the standard set by the expert in a blind condition, i.e. raters are unaware of the experts' coding decisions (Walker 2005). Burla et al. (2008) argued that calculating inter-rater reliability is an established method to analyse the quality of a coding scheme. Here, inter-rater reliability of the PCM coding scheme was conducted on 65 text segments from the T1 CDM interview.

Reliability scores were calculated based on percentage agreement, i.e. number of agreements divided by the number of times the coding was possible, multiplied by 100; this is in accordance with the literature that has suggested that this is the most suitable way to calculate reliability scores with data of this nature. For example, Boyatzis (1998) has argued that percentage agreement should be used when themes to be coded (in this case three) and numbers of observed situations (segments, in this case 65) are few. Three themes and 65 segments are relatively few in relation to other coding situations in which infinite themes can be identified from many hours of interview data. Furthermore, Boyatzis (1998) argued that if data resulting from the coding are nominal, percentage agreement scores are appropriate. Within the literature, there are conflicting accounts as to the acceptable level of percentage agreement as there are no established standards (Lombard, Snyder-Duch, and Bracken 2002; Marques and McCall 2005). There is a general consensus that a level of 80% agreement and above indicates an acceptable level of reliability (Jentsch and Bowers 2005); therefore, this will be the benchmark for assessing reliability in this study. Percentage agreement will also be used to assess the intra-rater reliability of the coding scheme, i.e. the same codes are assigned by the same coder on two separate occasions.

4.2.1. Participants

A repeated-measures within-subjects design was used in the study to assess inter-rater reliability and intra-rater reliability of the coding scheme presented in Section 2 (see Table 1). The study required completion of a coding task conducted twice, exactly 4 weeks apart, with the same group of participants. Twenty post-graduate research students at the University of Southampton were voluntarily recruited from an e-mail advert. Seventy-five percent of the sample was male. The mean age of the participants was 27 years (range between 21 and 40 years, $SD = 4.5$). All participants rated English as either their first language (45% of sample) or as good as their first (55% of sample). The participants were paid £20 for taking part in the study, which was paid on completion of the second coding task.

4.2.2. Materials

A presentation was given that included a description of the PCM with details of the coding scheme (Table 1) and six coding examples. The participants were provided with a coding guide that included a description of the PCM and examples of the coding scheme, a practice question–answer sheet and an answer booklet which included a page of demographic questions and the coding task. The coding task consisted of extracts of the incident transcript in continuous prose to allow the interview to be read and understood in context, and then each extract was broken down into segments which were to be coded (65 text segments in total). The segments were provided in a table under each extract; the first column contained the segment, the second column was for the answer (schema, action or world) and the third column was for comments in case the participant wanted to explain the reason for the code they assigned to any of the segments.

4.2.3. Procedure

The experimenter gave a presentation to the participants, in which the nature of the coding task was explained. Six practice questions were included in which participants were instructed to complete the answers for each example and answers were given to the group. Throughout the presentation, participants had the opportunity to ask questions. At the end of the presentation, participants were invited to complete the consent forms, answer the demographic questions and complete the coding task which took approximately 30 min. Participants were told that there was no right or wrong answer, nor was there a certain number of each code or a pattern to the coding scheme. Participants were given the option to withdraw from the study at any time. The study was held at the same time over 2 days (Tuesday and Thursday of the same week) to increase opportunity for participation. Exactly the same procedure was conducted 4 weeks after the first study, either Tuesday or Thursday. To control for extraneous variables, participants were discouraged from switching group days. All participants returned to their appropriate second study time.

4.2.4. Results

In the first coding session, inter-rater reliability between the participants and the criterion coder averaged 88% (range between 72% and 94%, SD = 5.5). In the second coding session, inter-rater reliability was 84% (range between 65% and 95%, SD = 10.6). Intra-rater reliability was calculated as an average of 83% (range between 60% and 92%, SD = 10).

In a review of methods measuring inter-rater reliability, Lombard, Snyder-Duch, and Bracken (2002) argued that the biggest limitation of calculating percentage agreement is the failure for it to account for agreement that would occur by chance. To account for this criticism, a random allocation procedure was used. A random number generator was used to select the order of the 65 text segments. The three codes were assigned a number (schema = 1, action = 2, world = 3) and the number generator was instructed to generate 65 integers between 1 and 3. These numbers were assigned, in the order they appeared, to the 65 segments. This procedure, therefore, assigned the codes generated at random to the text segments ordered at random. This is akin to chance allocation and the procedure was repeated five times. To assess for chance, percentage agreement between the criterion coder and the random allocation was calculated. This averaged 31% (range between 22% and 39%, SD = 6.4).

4.2.5. Preliminary discussion

In the inter-rater results, percentage agreement was 88% at the first coding session and 84% in the second session. Both results are above the suggested 80% threshold for acceptable agreement. This demonstrates that different people using the PCM coding scheme on different occasions are coding CDM data in a consistent manner. In addition, the intra-rater percentage agreement of 83% demonstrated that the same people on different occasions are also using the coding scheme in a consistent manner. These results suggest that a high level of reliability is associated with the PCM coding scheme for coding CDM data. Furthermore, the use of the random number generator to calculate chance agreement (31%) established that the results obtained from the inter-rater reliability study were considerably higher than from chance alone, increasing confidence in the use of the coding scheme. The high-reliability results are not surprising when considered in the context of the theory-driven coding scheme that was implemented. Boyatzis (1998) has argued that a theory-driven code is likely to achieve consistency of judgement as interpretations are a direct commentary on the theory, rather than individual interpretation. In addition, Klein, Calderwood, and Macgregor (1989) suggested that CDM interviews were easier to code and led to reliable data because of their semi-structured nature and use of standardised probes.

5. General discussion

5.1. Summary

The aim of this study was to demonstrate how the PCM can be used as a means to explore the decision-making process by using it to structure and understand data elicited from the CDM. To achieve this, a critical incident interview was conducted with a helicopter pilot using the CDM procedure (Klein, Calderwood, and Macgregor 1989). As discussed in the introduction to this paper, aeronautical decision-making is a complex form of NDM because there is not always a clear coupling between the decision process and the decision outcome. Therefore, it is important to understand the adequacy of the decision *process*, rather than just the outcome. To do this, it is necessary to understand why actions and assessments made sense to an operator at the time the decisions were made (Dekker 2006). Neisser's (1976) PCM offers this process understanding by placing the operator in the environment in which they work. In this study, it was demonstrated that each phase of the incident could be understood in terms of the perceptual cycle the pilot was engaged in.

In order to address the concern levelled at the method about reliability, the retest reliability of the CDM was examined. This reliability study took place over a 34-month period. It was demonstrated that the CDM was able to produce consistent interview data at T1 (9 months after the incident) and T2 (34 months after the incident). To our knowledge, this is the only study that has examined the reliability of the CDM over such a long period of time. Previous studies, for example Taynor, Crandall, and Wiggins (1987), examined reliability only up to 5 months after the incident. The results presented in this paper inspire confidence in the use of the CDM to elicit consistent data and help to alleviate some of the concerns about the use of the method in relation to the fallibility of human memory.

It is acknowledged by Klein, Calderwood, and Macgregor (1989) that the qualitative data generated by the CDM require some form of subjective interpretation usually through thematic analysis of the data. Deductive thematic analysis was conducted here as themes in the data were generated from existing theory (the PCM). There is consensus in the literature that assessing inter-rater reliability and intra-rater reliability is an established method to analyse the quality of a coding scheme. In this study, both inter-rater reliability and intra-rater reliability were demonstrated to be above the threshold level for acceptable agreement. This suggests that the coding scheme based on the principles of the PCM is a reliable way to analyse CDM data from the perspective of understanding data in the context of schema theory and the perceptual-cycle process of decision-making.

5.2. General applicability

A central tenet of the PCM is the use of schemata in information processing and decision-making. As introduced at the start of this paper, schemata are generally advantageous, as they help organise the mass of world information available to decision makers and can reduce cognitive expenditure by directing attention and influencing action. However, they can also leave decision makers vulnerable to making schema-driven errors. Therefore, it is important to understand their role in decision-making. This is especially pertinent in the context of aviation where rigorous training procedures and SOPs are designed to structure the decision-making process by offering a limited set of choices (Simpson 2001). It could be hypothesised that pilots will not display a high use of schemata because of the proceduralised nature of aviation. However, in this study, the pilot was found to utilise a number of schemata in order to assist his decision-making process. It is possible to infer the consequences of decision-making that is largely schema driven. In general, the use of schemata aids perception and decision-making. For example, Morris and Leung (2006) found that mental workload was not significantly increased, when task demand increased, if pilots could revert to pre-existing schemata. However, the reverse of this is that schemata can influence the production of inappropriate actions and decisions; for example, over reliance on pre-existing but inappropriate schemata has been shown to lead to fixation on certain cues at the expense of others (Stanton et al. 2010a; Plant and Stanton 2012a). In the incident presented here, the pilot entered the wrong waypoint as a result of schema-driven decision-making. Furthermore, the majority of the diagnostic attempts were influenced by the schemata held for the aircraft's electronic system. Fortunately in this instance, there were no adverse effects and the incident was relatively minor. However, a situation can be envisaged whereby schemata will influence decision-making at the expense of making an appropriate decision.

The main impetus for this paper was to explore a method that could assist with understanding the role of the PCM in decision-making. The CDM was utilised as the method of qualitative data collection. The combination of the CDM for data collection and the PCM coding scheme for data analysis is, as far as the authors are aware, a novel one. Schemata are notoriously difficult to elicit (Smith and Hancock 1995; Walker, Stanton, and Salmon 2011). Schema theory is intuitively appealing but has not always received unanimous acclaim, which is in part due to the difficulties associated with eliciting data that can be interpreted in terms of schema theory (Plant and Stanton 2013). This study has demonstrated that the probes in the CDM produce data that can be understood in terms of schema theory. The PCM accounts for the processes involved (i.e. why decisions made sense to the operator at the time) and it encompasses the whole system but remains human-centred by placing the operator into the environment in which they work. Literature suggests that events preceding accidents are usually non-routine or novel situations (Reason 1990); therefore, understanding decision-making when dealing with critical incidents has the potential to shed light into the cognitive processes involved when accidents happen.

With any form of methodological exploration come lessons learnt and recommendations. Here, this primarily concerns the reliability of both the CDM and the PCM coding scheme. The test–retest reliability of the CDM was conducted to address the criticism about the reliability of the method and the fact that, to our knowledge, a retest reliability study of the CDM has not been conducted for at least 20 years and not with a time elapse of longer than 5 months (Taynor, Crandall, and Wiggins 1987). Conducting a CDM retest is time consuming and may not always be practically possible if time with the interviewee is limited to one session. The study by Taynor, Crandall, and Wiggins (1987) and the retest study conducted here both found the CDM to be reliable for eliciting similar data on different occasions. Furthermore, Klein, Calderwood, and Macgregor (1989) have argued that the use of standardised probes is likely to result in more reliable data. As such, we do not suggest that every CDM interview requires a retest; however, based on the data differences found from the retest, we recommend that time is allowed at the start of the CDM interview for practice and familiarisation with the cognitive probes. Other qualitative data collection techniques, such as verbal protocols, advocate practice as part of the procedure (Walker 2005). Some of the differences found between T1 and T2 have been attributed to the fact that the interviewee was more familiar with the format and procedure at T2 (see Table 4). We encourage a 15-min practice session to familiarise with the probes in the context of an everyday activity before the main incident interview. Furthermore, if resources are available to conduct a retest study, we advocate the use of a different researcher to do this as some of the differences found between the T1 and T2 data were attributed to familiarity with the researcher (see Table 4). The coding scheme used in this study was shown to uphold both inter-rater reliability and intra-rater reliability and allowed CDM data to be analysed in relation to the perceptual cycle process. A similar coding process was used by Plant and Stanton (2012a) in relation to accident report data; therefore, we argue that the coding scheme is mature enough to be utilised by other interested researchers with any form of qualitative data.

5.3. Avenues of future work

This study has presented many avenues for future work, some of which have already been alluded to throughout the paper. One of the primary areas to focus future research concerns the level of analysis. As previously discussed, the CDM interview came from the perspective of the PNF and as such the roles of the additional crew members were not explicitly addressed. The pilot was instructed to recall the incident from the perspective of his personal decision-making process: detailed information about the decision-making of other crew members would have been highly speculative, particularly in relation to the use of schemata. Speculating on the role of other people's prior knowledge and previously held experiences in decision-making is not appropriate in the context of schema theory: schemata are entirely unique to the individual who holds them. Neisser (1976) explained that '*schemata are developed by experience*; *everyone*'*s experiences are different* . . . *since every person*'*s perceptual history is unique we should all have unique cognitive structures*' (p. 187). We acknowledge that exploring decision-making solely from an individual perspective, particularly in a complex system such as aviation, is too simplistic (Plant, Stanton, and Harvey 2013). However, as the motivation behind this paper was methodological exploration, the use of one critical incident, from the perspective of one crew member, was deemed appropriate. We are now in position to examine the perceptual cycle process from a larger level of analysis and work in this vein is already underway (see Plant, Stanton, and Harvey 2013).

In addition, future endeavours will seek to establish the predictive value of this approach as thus far only retrospective analyses have been considered (see Plant and Stanton 2012a, 2012b). There has been a call in the aviation literature for proaction rather than reaction with regard to safety (Maurino 2000; McFadden and Towell 1999). Similarly, Hancock, Hancock, and Warm (2009) argued that predictive knowledge is the key to making the next step in scientific progress, without being able to predict behaviour, behavioural modification will not be possible. Schema theory allows predictions to be made about memory (Plant and Stanton 2013). For example, based on the principles of schema theory, it can be predicted that when new situations are encountered, existing schemata will be utilised to help interpret the new situation. If an appropriate schema cannot be found, then it will be harder for a new situation to be interpreted and a 'best-fit' schema may be utilised instead. Exploring why decisions are made from the perspective of schema theory and the predictions this affords has the potential to provide an understanding of schema triggers and reasons for inappropriate activation of schemata or the failure for specific schemata to activate (Norman 1981). Using the PCM and schema theory can help predict what *can* go wrong based on a description of what *should* happen, thus advancing a proactive approach to the study of aeronautical decision-making.

6. Concluding remarks

The value of this research lies in the application of an analysis method based on the PCM (coupled with the CDM as the method of data collection) to understand the processes involved in aeronautical decision-making. The PCM analysis structures the analysis of qualitative data in such a way that the integrating elements of the PCM: schemata, actions and world information, are accounted for and attributed to decisions. It is widely acknowledged that schemata influence decision-making (Simpson 2001), although it is less widely understood how this process happens and the potential implications of this. The use of the PCM to understand decision-making data allows for the adequacy of the decision, and decision process, to be assessed, as opposed to just the decision outcome. This aligns with the aims of NDM research (Orasanu and Martin 1998). In the incident presented here, potential inadequacies in the decision-making process were apparent in the pilot's reliance on his schema for an electrical fault. However, the over-all decision that was made (albeit that this decision was made collectively amongst the crew), to abort the training and return to base, was entirely adequate for the situation as a successful outcome was achieved. The decision-making process, structured via the PCM (Table 2 and Figure 3), demonstrated what information (both external and internal) the pilot used to deal with the critical incident. From this study, it would appear that CDM data can be classified within the context of the PCM and that this classification can be used to help understand and explain aeronautical decision-making. The application of the PCM coding scheme to analyse CDM data is, to our knowledge, a novel application. The motivation for such research is to develop more process-driven decision-making research in order to establish causal accounts of why pilots make decisions which will increase understanding about the potential consequences of those decisions.

Acknowledgements

The authors wish to thank the participants in the study, particularly the pilot who provided the incident description. We also extend our thanks to the reviewers whose comments have greatly improved the quality of the paper.

Note

1. A company report was logged on the safety quality integrated database system and the issue fed back to the relevant technical authorities. The cause of the problem reported in this paper (i.e. WP more than 10,000 miles away) is the cause *attributed* by the pilot only.

References

Baber, C., N. A. Stanton, J. Atkinson, R. McMaster, and R. J. Houghton. 2013. "Using Social Network Analysis and Agent-Based Modelling to Explore Information Flow Using Common Operational Pictures for Maritime Search and Rescue Operations." *Ergonomics* 56: 889–905.

Bartlett, F. C. 1932. *Remembering: A Study of Experimental and Social Psychology*. Cambridge, MA: Cambridge University Press.

Boyatzis, R. E. 1998. *Transforming Qualitative Information: Thematic Analysis and Code Development*. Thousand Oaks, CA: Sage.

Burla, L., B. Knierim, J. Barth, K. Liewald, M. Duetz, and T. Abel. 2008. "From Text to Codings. Intercoder Reliability Assessment in Qualitative Content Analysis." *Nursing Research* 57 (2): 113–117.

Crandall, B., and K. Gretchell-Reiter. 1993. "Critical Decision Method: A Technique for Eliciting Concrete Assessment Indicators From the Intuition of NICU Nurses." *Advances in Nursing Science* 16: 42–51.

Crandall, B., G. A. Klein, and R. R. Hoffman. 2006. *Working Minds: A Practitioner's Guide to Cognitive Task Analysis*. Cambridge, MA: MIT Press.

Dekker, S. W. A. 2006. *The Field Guide to Understanding Human Error*. Aldershot: Ashgate.

Diehl, A. 1991. "The Effectiveness of Training Programs for Preventing Aircrew "Error"." In *Proceedings of the Sixth International Symposium on Aviation Psychology*, 640–655. Columbus, OH: The Ohio State University.

Elio, R., and P. B. Scharf. 1990. "Modelling Novice-To-Expert Shifts in Problem Solving Strategy and Knowledge Organisation." *Cognitive Science* 14: 579–639.

Elliott, T. 2005. "Expert Decision-Making in Naturalistic Environments: A Summary of Research." In *Technical Report DSTO-GD-0429*, 1–61. Defence Science and Technology Organisation.

Hancock, P. A., G. M. Hancock, and J. S. Warm. 2009. "Individuation: The *N*=1 revolution." *Theoretical Issues in Ergonomics Science* 10 (5): 481–488.

Hobbs, A., and A. Williamson. 2002. "Skills, Rules and Knowledge in Aircraft Maintenance: Errors in Context." *Ergonomics* 45 (4): 209–308.

Hoffman, R. R., B. Crandall, and N. Shadbolt. 1998. "Use of the Critical Decision Method to Elicit Expert Knowledge: A Case Study in the Methodology of Cognitive Task Analysis." *Human Factors* 40 (2): 254–276.

Jentsch, F., and C. Bowers. 2005. "Team Communication Analysis." In *Handbook of Human Factors and Ergonomics Methods*, edited by N. A. Stanton, A. Hedge, K. Brookhuis, E. Salas, and H. Hendrick, 50.1–50.5. Boca Raton, FL: CRC Press.

Johnson-Laird, P. 1983. *Mental Models*. Cambridge, MA: Harvard University Press.

Klein, G. A. 1993. "Twenty Questions, Suggestions for Research in Naturalistic Decision Making." In *Decision Making in Action: Models and Methods*, edited by G. A. Klein, J. Orasanu, R. Calder-wood, and C. E. Zsambok, 389–403. Norwood, NJ: Ablex.

Klein, G. A., and A. Armstrong. 2005. "Critical Decision Method." In *Handbook of Human Factors and Ergonomics Methods*, edited by N. A. Stanton, A. Hedge, K. Brookhuis, E. Salas, and H. Hendrick, 35.1–35.8. Boca Raton, FL: CRC Press.

Klein, G. A., R. Calderwood, and A. Clinton-Cirocco. 1986. "Rapid Decision Making on the Fire Ground." In *Proceedings of the Human Factors and Ergonomics Society 30th Annual Meeting*, Vol. 1, 576–580. Norwood, NJ: Ablex.

Klien, G. A., R. Calderwood, and D. Macgregor. 1989. "Critical Decision Method for Eliciting Knowledge." *IEEE Transactions on Systems, Man and Cybernetics* 19 (3): 462–472.

Lieberman, D. A. 2012. *Human Learning and Memory*. Cambridge, MA: Cambridge University Press.

Lombard, M., J. Snyder-Duch, and C. C. Bracken. 2002. "Content Analysis in Mass Communication. Assessment and Reporting of Intercoder Reliability." *Human Communication Research* 28 (4): 587–604.

Marques, J. F., and C. McCall. 2005. "The Application of Inter-Rater Reliability as a Solidification Instrument in a Phenomenological Study." *The Qualitative Report* 10 (3): 439–462.

Maurino, D. E. M. 2000. "Human Factors and Aviation Safety: What the Industry Has, What the Industry Needs." *Ergonomics* 43 (7): 952–959.

McFadden, K. L., and E. R. Towell. 1999. "Aviation Human Factors: A Framework For The New Millennium." *Journal of Air Transport Management* 5: 177–184.

Minsky, M. 1975. "A Framework for Representing Knowledge." In *The Psychology of Computer Vision*, edited by P. Winston, 211–277. New York: McGraw-Hill.

Morris, C. H., and Y. K. Leung. 2006. "Pilot Mental Workload: How Well Do Pilots Really Perform?" *Ergonomics* 49 (15): 1581–1596.

Neisser, U. 1976. *Cognition and Reality*. San Francisco: W.H. Freeman and Co.

Norman, D. A. 1981. "Categorization of Action Slips." *Psychological Review* 88 (1): 1–15.

O'Hare, D., M. Wiggins, A. Williams, and W. Wong. 1998. "Cognitive Task Analyses for Decision Centred Design and Training." *Ergonomics* 41 (11): 1698–1718.

Orasanu, J., and L. Martin. 1998. "Errors in Aviation Decision Making: A Factor in Accidents and Incidents." In *Proceedings of the Second Workshop on Human Error, Safety and System Development*, April 1–2 1998, Seattle, WA.

Plant, K. L., and N. A. Stanton. 2012a. "Why Did the Pilots Shut Down the Wrong Engine? Explaining Errors in Context Using Schema Theory and the Perceptual Cycle Model." *Safety Science* 50: 300–315.

Plant, K. L., and N. A. Stanton. 2012b. "'I Did Something Against All Regulations': Decision Making in Critical Incidents." In *Proceedings of the 4th International Conference on Applied Human Factors and Ergonomics*, 21–25th July 2012, San Francisco, CA.

Plant, K. L., and N. A. Stanton. 2013. "The Explanatory Power of Schema Theory: Theoretical Foundations and Future Applications in Ergonomics." *Ergonomics* 51 (6): 1–15.
Plant, K. L., N. A. Stanton, and C. Harvey. 2013. "The Role of the Perceptual Cycle in Teams." In *Proceedings of the Eleventh International Conference on Naturalistic Decision Making*, Marseille, France, 21–24th May 2013. Arpege Science Publishing.
Rafferty, L. A., N. A. Stanton, and G. H. Walker. 2010. "The Famous Five Factors in Teamwork: A Case Study of Fratricide." *Ergonomics* 53 (10): 1187–1204.
Reason, J. 1990. *Human Error*. Cambridge, MA: Cambridge University Press.
Reason, J. 2000. "Human Errors: Models and Management." *British Medical Journal* 320: 769–770.
Richardson, M., and L. J. Ball. 2009. "Internal Representations, External Representations and Ergonomics: Towards a Theoretical Integration." *Theoretical Issues in Ergonomics Science* 10 (4): 335–376.
Shappell, S. A., and D. Wiegmann. 2009. "A Methodology for Assessing Safety Programs Targeting Human Error in Aviation." *International Journal of Aviation Psychology* 19 (3): 252–269.
Simpson, P. A. 2001. "Naturalistic Decision Making in Aviation Environments." Report No. DSTO-GD-0279. DSTO Aeronautical and Maritime Research Laboratory, Australia.
Singletary, M. W. 1993. *Mass Communication Research: Contemporary Methods and Applications*. Boston, MA: Addison-Wesley.
Smith, K., and P. A. Hancock. 1995. "Situation Awareness is Adaptive, Externally Directed Consciousness." *Human Factors* 37 (1): 137–148.
Stanton, N. A., L. A. Rafferty, P. M. Salmon, K. M. Revell, R. McMaster, A. Caird-Daley, and C. Cooper-Chapman. 2010a. "Distributed Decision Making in Multihelicopter Teams: Case Study of Mission Planning and Execution From a Non-Combatant Evacuation Operation Training Scenario." *Journal of Cognitive Engineering and Decision Making* 4 (4): 328–353.
Stanton, N. A., P. M. Salmon, G. H. Walker, and D. P. Jenkins. 2010b. "Is Situation Awareness All in the Mind?" *Theoretical Issues in Ergonomics Science* 11 (1–2): 29–40.
Stanton, N. A., P. M. Salmon, G. H. Walker, C. Baber, and D. Jenkins. 2005. *Human Factor Methods. A Practical Guide for Engineering and Design*. Aldershot: Ashgate.
Stanton, N. A., and G. H. Walker. 2011. "Exploring the Psychological Factors Involved in the Ladbroke Grove Rail Accident." *Accident Analysis and Prevention* 43: 1117–1127.
Strauss, A., and J. Corbin. 1990. *Basics of Qualitative Research: Grounded Theory Proecedures and Techniques*. London: Sage.
Taynor, J., B. Crandall, and S. Wiggins. 1987. "The Reliability of the Critical Decision Method," KATR-863(B)-87-07F. Prepared Under Contract MDA903-86-C-0170 For the US Army Research Institute Field Unit, Leavenworth, KS, by Klein Associates, Inc., OH.
Walker, G. H. 2005. "Verbal Protocol Analysis." In *Handbook of Human Factors and Ergonomics Methods*, edited by N. A. Stanton, A. Hedge, K. Brookhuis, E. Salas, and H. Hendrick, 30.1–30.8. Boca Raton, FL: CRC Press.
Walker, G. H., N. A. Stanton, and P. M. Salmon. 2011. "Cognitive Compatibility of Motorcyclists and Car Drivers." *Accident Analysis and Prevention* 43: 878–888.
Wilson, J. R., and A. Rutherford. 1989. "Mental Models: Theory and Application in Human Factors." *Human Factors* 31: 617–634.
Wong, W., P. Sallis, and D. O'Hare. 1997. "Eliciting Information Portrayal Requirements: Experiences with the Critical Decision Method." *The Information Science Discussion Paper Series* 97.

Appendix

The four phases for conducting a CDM interview and the table of CDM deepening probes (from Crandall et al. 2006).

1. *Incident identification*: The interviewee identifies several candidate incidents and with the help of the interviewer an appropriate incident is selected for deepening. The selected incident will depend on the goals for data collection. For example, if the research question is interested in understanding the role of schemata in decision-making, then incidents in which knowledge and experience affected the outcome of an incident are favoured. The interviewee is required to provide a brief account of the story from beginning to end.
2. *Timeline construction and verification*: This phase is aimed at getting a clear and refined overview of the incident structure by identifying key events that occurred in order to expand the initial account. This structure provides a framework for the remainder of the interview. Diagramming the timeline can be a useful exercise as often mistakes or gaps are identified and additional detail can be added. This exercise can also identify crucial decision points in which the interviewee experienced a major shift in their understanding of the situation or take action that affected the events.
3. *Deepening*: This phase is the most challenging as it allows an opportunity to get inside the head of the expert. This phase goes beyond the time elements and basic facts and attempts to understand the perceptions, expectations, goals and judgements of the expert. This is where the CDM probes are used to deepen the understanding of the event. Following the advice in Stanton et al. (2005), if the participant gave a yes or no answer, prompts such as 'why' were used in order for answers to be expanded. A full list of the CDM probes is provided below. Probes should be defined prior to the analysis to ensure the output is compliant with the aims of the study (i.e. eliciting data that could help determine the role of schemata during critical decision-making). It is acknowledged that the probes have been modified over the years and researchers are encouraged to modify or add to the list as necessary for their individual research projects (Klein and Armstrong 2005; Crandall et al. 2006). In the light of the research goals here, five additional probes were added that expanded certain areas of the original CDM probes to increase their relevance to the elicitation of schemata (i.e. drawing out role of experience and expectations). These additional probes are italicised in the table of probes. In the full version of the CDM, the probes are asked for each individual event identified in the timeline. A shorter version of the CDM can also be conducted in which the probes are asked only once, in relation to the whole incident (as opposed to each individual segment).
4. '*What if*' *queries*: The final stage of the CDM interview provides an opportunity to consolidate the interviewer's insight into the interviewee's experience, skill and knowledge. The incident is used as a starting point for the interviewer to pose various

hypothetical scenarios (about the overall incident or individual segments) in order to establish how the outcome of the incident may have been altered. On the basis of time constraints, Crandall et al. (2006) have acknowledged that this stage is not always essential to achieve project goals and as such can be omitted in a shorter CDM interview.

It is advised that, owing to the intimate and sometimes emotional nature of the interview, the interview is not ended too abruptly (Crandall et al. 2006). Interviewee's need to be debriefed at the end and time should be allocated to answering any questions they may have.

CDM 'deepening' probe questions (items in italics are additional probes added for this research project)

CDM deepening probes	
Cues	What were you seeing, hearing, smelling and noticing?
	For each phase detail, any physical events (e.g. alarm sounding) that defined the phase
	For each phase detail, the mental events (thoughts, perceptions) that defined the phase
Information	What information did you use when making the decision or judgement?
	How and where did you get this information, and from whom?
	What did you do with the information?
	At any stage were you uncertain about the reliability or relevance of the information that you had available to you?
	Was there any additional information you might have liked to assist with formulating the decision?
Analogy	Where you reminded of any previous experience?
	What was it about the previous experience that seemed relevant for this case?
SOPs	Does this case fit a standard or typical scenario
	Is it a type of event that you are trained to deal with?
Goals and priorities	What were your specific goals and objectives during the incident?
	What was most important to accomplish at this point in the incident?
Options	What other courses of action were considered or were available to you?
	How was this option chosen and the others rejected?
	Was there a rule that you were following in choosing this option?
Experience	What specific training or experience was necessary or helpful in making this decision?
	Was the decision you made comfortably within your experience?
	Did you experience influence the decision that you made?
Expectations	Were you expecting this sort of incident to arise during the course of the flight?
	Were you expecting to make this sort of decision during the course of the event?
	Did your expectation influence your decision-making process, if so, how?
Mental models	Did you imagine the possible consequences of this action?
	Did you create some sort of picture in your head?
	Did you imagine the events and how they would unfold?
	At any stage did you find it difficult to process and integrate the information available to you? (if yes, how did you overcome this?)
Decision-making	*What was the primary decision that you made?*
	What features were you looking for when formulating your decision?
	How much time pressure was involved in making the decision?
	How long did it take you to actually make this decision?
	At any stage were you uncertain about the appropriateness of your decision?
Guidance	Did you seek any guidance at this point in the incident?
	How did you know to trust the guidance you got?