Modelling Naturalistic Decision Making Using an Artificial Neural Network: Pilots' Responses to a Disruptive Passenger Incident

Duggan, S & Harris, D

Original citation:
Duggan, S & Harris, D 2001, 'Modelling Naturalistic Decision Making Using an Artificial Neural Network: Pilots' Responses to a Disruptive Passenger Incident' *Human Factors and Aerospace Safety*, vol 1, no. 2, pp. 145-165.

ISSN 1468-9456

Publisher: Ashgate

Details of the definitive version are available at [URL to be supplied by Publisher]

Human Factors and Aerospace Safety 1(2), 145-166

Modelling Naturalistic Decision Making using an artificial neural network: Pilots' responses to a disruptive passenger incident

Sarah J. Duggan and Don Harris
Cranfield University, UK

Abstract

This paper describes a study conducted within a naturalistic decision making paradigm in which a disruptive passenger threatens the safety of a hypothetical flight. Sixty-five professional members of flight crew participated in a series of semi-structured interviews during which they described their decision-making process for dealing with this situation. An artificial neural network was used to model the decisions made on the basis of the situation assessment activities undertaken to produce an empirically verifiable model of the participants' decision-making process. Cross-validation of the results showed that decision outcomes could be very accurately predicted on the basis of this model. It is suggested that neural networks may be a viable way of modelling naturalistic decisions.

Introduction

Flying a modern, highly automated civil aircraft is no longer simply a problem of skilled psychomotor performance but is one of flight management and real-time decision-making. Fortunately, many of the decisions that are required of flight crew are routine, well structured and quite familiar. However, the crew will also

Correspondence: Don Harris, Human Factors Group, College of Aeronautics, Cranfield University, Cranfield, Bedford, MK43 0AL, United Kingdom or d.harris@cranfield.ac.uk

be required to solve less-routine problems. Decision-making is not a simple task: it involves situation assessment, choice amongst alternatives and assessment of risk (Orasanu, 1993).

Many non-routine problems are quite simple or can be solved by reference to checklists carried on the flight deck. These serve as essential aids for the management of the majority of in-flight technical failures encountered. However, the vast majority problems do not relate to equipment malfunctions. The sources of these problems are often external to the flight deck. The information about the cause of these problems is usually incomplete and/or probabilistic and the problems themselves are ill structured. Often there is no readily identifiable 'correct' solution, only satisfactory and less-satisfactory outcomes. Furthermore, the quality of many of these outcomes can only be described in retrospect.

When subject to even the most cursory analysis, it can be seen that the types of decision often required of flight crew are quite different to the problem solving scenarios encountered by participants in laboratory-based research. Much laboratory-based research has been couched within the classical decision-making (CDM) experimental paradigm, which itself has its roots in economic theory (Lehto, 1997). From a historical perspective, two major approaches within CDM research emerged: the 'preference and choice' approach and the 'statistical inference' approach, however both were based, either directly or indirectly, on formal logic and probability theory. In these types of study participants were often placed in a highly artificial situation attempting to solve a relatively simple, well structured problem with the benefit of being in possession of complete information. Often the decision required of them was simply the choice between several pre-determined alternatives. Nevertheless, the results of such experiments could be quantified and subject to statistical analysis. However, as a result of the methods by which these studies were undertaken to generate such data, the CDM paradigm was criticised as being a poor description of everyday decision-making and especially of decision-making on the flight deck. Later researchers have also criticised these approaches as focussing simply on the decision event and not the decisional process (Orasanu and Connolly, 1993).

As a result of these shortcomings of CDM theory, naturalistic decision-making (NDM) has become the dominant paradigm to describe pilot problem-solving behaviour. NDM recognises that the manner in which humans make decisions in the laboratory and in real life is quite different. Orasanu (1993) observed that decisions in the aviation domain are particularly well-suited to study using an NDM approach as they tend to be ill-structured and set in a dynamic, time-pressured environment where the consequences of a poor decision can be dire. In many cases there can be conflicting or competing goals (for example the trade-off between safety and efficiency) which may be the product of personal biases or organisational values. Furthermore, in aviation, a decision is not an end in itself, it is merely often just the pre-cursor to another decision. Finally, there is more than a single individual contributing to or implicated in most decisions taken (Orasanu and Connolly, 1993).

The context and quality of information have major influences on the manner in which a problem is approached in the 'real world'. The cognitive processes employed by the decision-maker in these situations are vastly different to those by which it is inferred people make a decision within the CDM paradigm. Exactly how these two paradigms differ depends upon the author and CDM/NDM model proposed. CDM models tend to favour an approach where several alternatives are evaluated in parallel by the decision maker (encompassing their own biases). The one with the greatest *perceived* chance of success or highest *perceived* utility is finally selected (e.g. minimax decision making models, subjective expected utility models or multi-attribute utility models – see Lehto, 1997 for an overview). NDM approaches suggest that after a brief period of situation assessment, a course of action is chosen that will potentially produce a *satisfactory* outcome. For example, Klein (1989) suggests that if a situation is not immediately recognised, the decision-maker mentally rehearses what the likely effect of certain actions will be. These options are then evaluated serially. Once a satisfactory outcome is thought likely on the basis of this cognitive activity, that course of action is pursued. This general approach to decision-making is also evident in other NDM models (e.g. Connolly, 1988; Montgomery, 1989).

While NDM theory has contributed greatly to understanding the decision-making processes in real-world situations it has had critics who have pointed out some of its limitations (e.g. Howell, 1997). Perhaps its greatest shortcoming lies in the analytical methods employed. The majority of NDM models have been developed qualitatively from observation and analysis of experts' decision-making processes. However, in common with most qualitative research, NDM's biggest strength is also its greatest weakness. The rich description of the decision-making process almost precludes any element of quantification or prediction. However, for NDM models to be accepted, they must be empirically testable, hence they need to be embedded within a theory. They need to offer predictions about outcomes (decisions) on the basis of context and information (see Hammond, 1993). A testable model can also be replicated and the bounds of its generalisability established. The method for the quantification of parameters and their analysis in CDM allows for such things, however, NDM models do not. At the moment they are descriptive rather than predictive.

Conventional statistical analytical techniques, either univariate or multivariate, cannot deal with the complexities and richness of the NDM process, where many inputs may simultaneously result in many outputs. 'Conventional' statistics cannot cope with complex branching of logical conditions and cannot provide a tool to model the complex relationships between inputs and outputs, which *could* be taken as reflecting the decision-making process. 'Conventional' statistics can only cope with additive and/or multiplicative relationships between predictor and criterion variables (or in the terms of decision making, inputs and outputs). It is also difficult to predict more than a single criterion variable and under no circumstances is logical branching allowed within an analysis. One technique for data analysis can cope with these complexities, though: neural networks.

Neural networks (NNs) are emerging as a technique within the social sciences to describe and model complex problems (Hair, Anderson, Tatham and Black, 1998; Garson, 1998). NNs allow the simultaneous prediction of many outcomes from many inputs and allow models to be developed which implicitly include complex 'if...then' expressions. They are particularly well suited to applications with noisy, missing, overlapping, non-linear and non-continuous data (Moore, 1998) and can also handle highly unstructured data. NNs do not, however, provide for probabilistic 'goodness-of-fit' tests of the models developed (*cf.* structural equation modelling) nor do they allow for statistical tests of difference. Nevertheless, they do potentially provide a way of building an empirically describable and verifiable model of the relationship between inputs and outputs. Perhaps as important for the purposes of NDM is the observation made by Haykin (1994) that NNs also allow for contextual information to be incorporated into a model. This model-building approach implicit in NN analysis also reflects an emerging trend in psychology and human factors which is away from statistical tests of difference and toward developing process-engineering types of models of human behaviour (Moray, 1999). Dowell, Smith and Pidgeon, (1997) have also suggested that NDM should be viewed as an engineering process.

Leven and Levine (1996) argue that NNs are more than just computational devices and that they can provide a basis for explaining multi-attribute decision-making. In the past there has been conflict within decision science, between quantitative (CDM) and NDM approaches. NDM studies have lacked mathematical and theoretical foundations, whereas quantitative approaches have neglected less easily quantifiable factors that are important in the decision making process. Leven and Levine (1996) showed that the NN approach can encapsulate both CDM and NDM approaches.

NNs are based upon a simplified model of the hypothesised manner in which the brain operates. The most common form of artificial NN is the multi-layer perceptron (simplified elements of which are illustrated in figure 1). The basic processing unit in a NN is the node. A NN contains many nodes, some of which represent input values and some of which represent output values, plus often a middle ('hidden') layer. It is the nodes in the hidden layer that allows relationships to be more complex than simple one-to-one relationships between input and output. Each node is a self-contained unit that acts in parallel with other nodes. Every connection between nodes has a weight (ω) that is applied to the incoming data. Each node then creates a summated value of these products, to which is applied an activation function which creates the output value from the node. The activation function usually takes the form of a simple function which dictates that the output of a node is of the binary form 'fire' (produce an output) or 'inhibit' (no output).

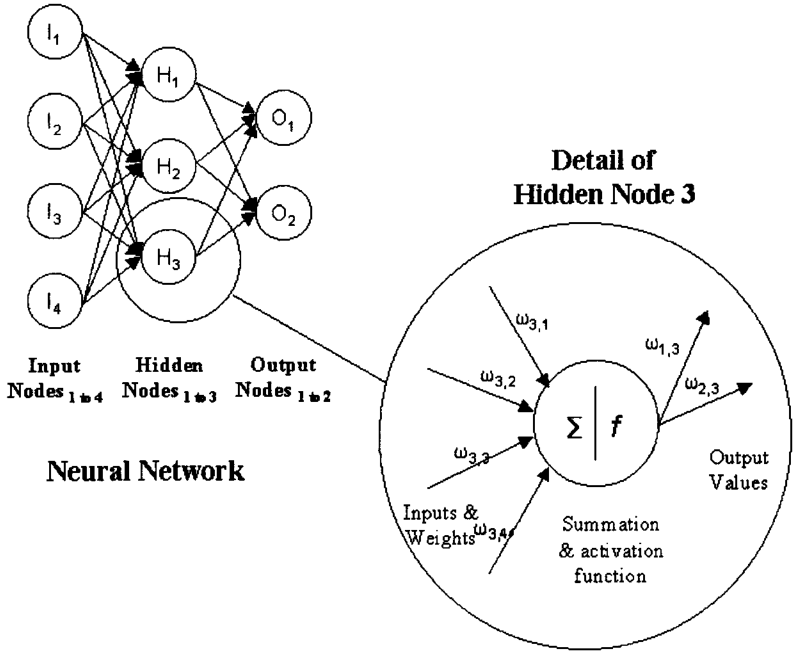

Figure 1 Elements of the multi-layer perceptron type of NN (adapted from Hair, Anderson, Tatham and Black, 1998; and Garson, 1998)

The essential feature of NNs is that they learn the relationship(s) between inputs and outputs (or in the current application 'information' and 'decisions') and self-correct. All NN applications require two sets of data, a set for supervised learning, where the outcomes are known and a cross validation data set to test the model derived. In the supervised learning set, the model commences with a 'best guess' and applies a set of weights. Predicted outputs are compared with actual, known outputs and the weights in the model are corrected to reflect the sign and magnitude of the error observed. This continues iteratively until the overall error rate falls below some pre-defined criterion. The model is then applied to the input data from the cross-validation data set and the output predictions from the NN are compared to the actual outputs. In this way the NN model is validated.

Garson (1998) describes many studies that demonstrate the successful application of NNs, many of which are models for making economic decisions and predictions (cf. early CDM models). In particular, he suggests NNs are well suited to pattern-matching applications. As alluded to earlier, the first aspect of any decision-making task is that of situation assessment. Situation assessment is itself a precursor for situation awareness which in turn is the precursor for all decision-making (Lipshitz, 1993; Nobel, 1993; Prince and Salas, 1997). The NDM models proposed by Hammond (1988), Klein (1989) and Nobel (1993) are all essentially pattern-matching models of decision-making, (although it is arguable if Hammond's work can actually be considered part of the NDM paradigm; it is, however, strongly related to these other NDM models in its approach). Nobel (1993) particularly emphasises role of situation assessment. In these NDM models, information is gathered about the situation and compared with mental models (reference problems) of similar situations held in long-term memory. These may be formulated from actual experience or vicariously through such methods as formal training or informal discussions with colleagues. These situation assessment and subsequent, cognitive pattern-matching activities form the basis for the subsequent decisions and actions taken. It would thus seem that NNs might provide a method by which an empirically verifiable model of situation assessment and NDM may be constructed. This paper tests this assertion by modelling the decision-making processes of pilots responding to a (hypothetical) 'air rage' (disruptive passenger) incident.

Method

Scenario requirements and development

An unstructured problem scenario with indirect, incomplete and inferential information was developed. The scenario developed had to be suitable for administration to flight crew using a semi-structured telephone interview in order to gain a large sample with diverse experience from a geographically widely-dispersed population and within a reasonable time frame. It also had to have no prescribed 'correct' solution. The scenario must not involve a well-defined course of action prescribed by such flight deck items as the checklists in the quick reference handbook. As a result of these criteria, interviewees would be required to make a decision based upon only their assessment of the situation.

The scenario chosen was an 'air rage' scenario. This was selected for several reasons. Firstly, while many airlines promulgate guidelines and advice to their pilots about prospective courses of action to take when there is an unruly passenger in the aircabin, there are no set checklists to follow in such circumstances. Secondly, the interviewee would require no instrument readings or air traffic control instructions. The information the interviewee would receive over the telephone from the interviewer would be similar in content and mode of communication as that they would receive from the cabin crew in an actual incident.

'Air rage', or disruptive passenger behaviour has become an issue of increasing concern for aviation safety in the late 1990s. The UK Flight Safety committee (1998) suggested that up to 6,000 disruptive passenger incidents might occur every year on UK-registered aircraft. Not all lead to a prosecution, however in the 12-month period ending March 1998, British Airways reported 266 incidents (Jack, 1998). In a similar period United Airlines reported 450 incidents (Longmuir, 1998). American Airlines reported a 200% increase in disruptive passenger behaviour between 1994 and 1995 (Hicks and Morrison, 1997). These incidents can compromise the safety of the aircraft and the crew themselves. Cabin crew have been attacked by passengers with broken bottles and in July 1999, a pilot on a Japanese Airlines aircraft was stabbed to death when a passenger broke into the flight deck. On other occasions disruptive passengers attempting to open doors or overwing emergency exits during flight have compromised the safety of everyone on the aircraft. In response, in 1995 the UK Air Navigation Order was revised (and further revised in 1999) to make it an offence to intentionally interfere with the duties of cabin crew. This is now punishable by up to two years imprisonment; endangering the safety of an aircraft carries a maximum sentence of five years.

When faced with a disruptive passenger incident the Commander of the aircraft has various choices. S/he can elect simply to attempt to diffuse the situation; they can have the passenger physically restrained and/or they can make an immediate diversion to the nearest suitable airfield if they believe that there is any danger to passengers, crew or the aircraft itself. In the latter case, the costs incurred may be recovered from the passengers responsible.

The interview scenario was developed with the aid of two senior training captains from a major international airline both of whom had experienced in-flight incidents involving a disruptive passenger and also a senior member of cabin crew. The scenario and interview protocol was subject to several iterations using a further flight crew before commencing the study proper.

Participants

All participants were recruited from the Cranfield University volunteer pilot panel. All interviewees were in possession of a full UK Airline Transport Pilot's Licence and were flying heavy passenger-carrying, commercial transport aircraft at the time of the study. In total, 65 participants were interviewed.

Scenario

All interviewees were presented with the following information. They were told that they were flight crew on a four-hour flight from airport A to airport B. The aircraft was flying over a country in mainland, Western Europe. Prior to take-off there had been a minor technical problem with the aircraft that had delayed departure. Many passengers were unhappy about this delay and several had been drinking. On this particular flight there was an all-female cabin crew compliment. The 'air rage' incident commenced about halfway through the flight. At the time of the incident there was a diversionary airport available (airport C), approximately 30 minutes flying time away. Facilities at airport C for passenger handling were rudimentary, however, due to poor weather in the vicinity, it was the only airport with a suitably equipped runway of adequate length available.

The remainder of the scenario was as follows, however, this information was not volunteered to the interviewee unless it was specifically asked for. The disruptive passenger was a moderately large man, travelling as part of a larger group of passengers. He was deemed to have become abusive toward the cabin crew after having (apparently) consumed a large amount of alcohol. He was not on medication.

Procedure

The interviewee was contacted prior to the telephone interview to establish if they were willing to participate in the study. If they indicated that they were willing to do so, they were telephoned on a later occasion at a mutually agreed time.

The interviewer gave a brief overview of the purposes of the study before commencing the interview. Demographic details for each interviewee (age, sex, rank, flying experience, type-ratings etc.) were available from the Cranfield University volunteer pilot panel database. The agreement of the interviewee to record the interview was also elicited at this time.

The interviewee was given the initial scenario information described in the previous sub-section. They were asked what actions they would take to assess the situation (i.e. to gather information) and what decisions they would make about actions to control the situation. To obtain the information described in the latter part of the scenario participants had to specifically ask for it. If they asked a question that required information not contained within the scenario, the interviewer, who was essentially playing the part of the Purser reporting the incident to the flight deck, replied that she did not know. To help the interviewer maintain consistency in her responses, a scenario 'flow diagram' was available to her at all the times during the interview with all the explicitly 'known' information about the passenger and the incident on it. The interviewer only prompted for clarification of certain points or actions when required.

Interviews typically lasted between five and 15 minutes. At the end the interviewee was asked if they had any questions and thanked for their time.

Results

Sample characteristics

In the final sample, 61 interviewees were male and four were female. Twenty were First Officers and 45 were Captains. All interviewees were in possession of a full UK ATPL (Airline Transport Pilots Licence) at the time of the study. Mean flying experience (in years) was 20.3 (with a standard deviation of 11.3 years): mean flight experience (in hours) was 9,027.7 hours (standard deviation of 6,069.1).

Treatment of data

Subsequent to the interviews the content of the transcripts was categorised into either 'situation assessment' variables (where the interviewee had actively alluded to either searching for information or procuring information about the situation by other means) or 'subsequent actions/decisions' aimed at dealing with the situation. Various sub-categories within each of these broad categories were further identified from content analysis of the interview transcripts. This coding framework encompassed all the 'situation assessment' activities and 'action/decisions' made by the interviewees. The researchers independently verified the existence of these categories within the data set.

The contents of the interview transcripts from each participant were then re-categorised using the coding framework derived. The pilot's reports were coded for containing evidence of having acquired (or not acquired) a certain element of information from their information seeking activities or having made (or not made) a certain decision subsequent to these actions. The coding of the interview transcripts into this framework was subject to cross checking by the researchers. Table 1 contains a description of the categories elicited from the interviews and their frequencies of occurrence within the data set.

Initial qualitative analysis and discussion of the data

It can be seen from figure 2 that the vast majority of pilots interviewed (72.3%) suggested they would acquire situation assessment information from multiple sources before making their decision and implementing their actions. However, relatively few pilots would elect to obtain situation assessment information from more than three sources. The results in table 1 suggest that the most frequent situation assessment activity undertaken was to get a full description of the situation from the cabin crew. Just over one-quarter of all in the interviewees said that they would send the other pilot back to the aircabin to appraise the situation even though many company's operating procedures strongly suggest that it is undesirable to leave a solitary member of crew on the flight deck in such circumstances. The two approaches tended to be mutually exclusive, which suggests that there may be either a whole-crew-based approach to dealing with the problem or a flight deck-based approach. Although it is not directly evident from the results presented (as this would require presenting an *extremely* large multi-way contingency table) only in eight cases did the decision-maker gather information both from the cabin crew and also send a pilot back to investigate. Interestingly, in five of these eight cases, it was established that the passenger was a large man. Given that there was an all-female cabin crew compliment on board, this action is perhaps not too surprising. If it was established that there were off-duty crew on board (a piece of information solicited in approximately one-quarter of all instances) it was unlikely that a member of flight deck crew would be sent to investigate (this only happened on five out of a possible 17 occasions). In four out of the six occasions that it was established that the passenger was a large man, the interviewee also appraised the facilities at the diversionary airport! This option to divert was also frequently associated (11 out of 15 occasions) when the pilot *did not* investigate if there were other off duty crew travelling.

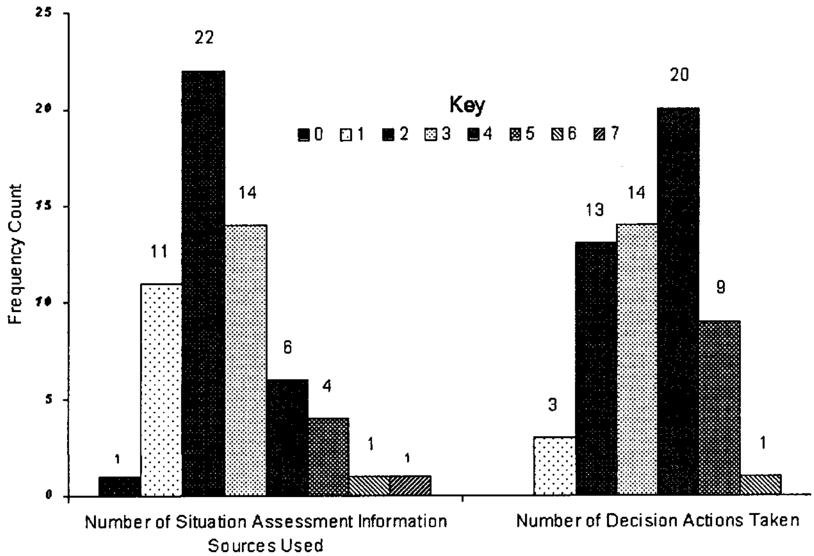

Figure 2 Frequency distributions for the number of situation assessment activities undertaken by interviewees and the number of subsequent actions implemented on the basis of the decisions made

Simple apparent associations between input variables are reasonably easy to identify using a multi-way contingency table as long as the data set is not too large or is it desired to look at the associations only between two or three variables. However, it becomes more difficult to identify relationships such as the last one in the previous paragraph, where the presence of one situation assessment variable is apparently associated with the *absence* of another. The problem becomes even greater where it is desired to establish associations between many situation assessment variables (input nodes) and subsequent decisions. In these cases *not* gathering situation assessment information may be equally as important in determining a decision as actually gathering data. It can also be seen from figure 1 that only in very few cases would a pilot follow a single course of action on the basis of their assessment of the situation.

Table 1 Frequency of situation assessment actions (input nodes) and subsequent decisions and actions (output nodes) derived from the interview transcripts

Situation assessment actions (input nodes)	*n* (%)	Subsequent decisions/ actions (output nodes)	*n* (%)
Establish if the passenger is travelling alone	47 (72.3)	Get friends or other able-bodied passenger to help	47 (72.3)
Get a full description of the situation from the cabin crew	25 (38.5)	Issue a specific warning to the disruptive passenger	44 (67.7)
Send a pilot back to the aircabin to appraise the situation	17 (26.2)	Contact the company	27 (41.6)
Appraise facilities at nearer, alternate airport (airport C)	15 (23.1)	Organise police at destination airport	26 (40.0)
Establish if there are any off-duty crew travelling that may help	15 (23.1)	Make a general warning on the public address system	19 (29.2)
Collect personal details about the passenger	7 (10.8)	Divert to nearer alternate airport (airport C)	17 (26.2)
Establish if he is a large man	6 (9.2)	Refuse to serve any more alcohol	9 (13.8)
Establish if he is on medication	6 (9.2)	Physically restrain the passenger	9 (13.8)
Appraise the facilities in the country being overflown for dealing with the problem	6 (9.2)	Move him away from other passengers	4 (6.2)
Ascertain if there were any problems prior to boarding	5 (7.7)		
Get a summary of the problem from the disruptive passenger	4 (6.2)		

Not too surprisingly the most common courses of action were enlisting the help of the passenger's friends or other able bodied people on the aircraft and issuing a formal warning (see table 1). On 30 occasions, the former of these decisions was made when the interviewee had previously sought to establish if the passenger was travelling alone and/or if they had any off-duty crew in the aircabin, which is logical. Issuing a warning was not strongly related to any one situation assessment variable.

Interestingly, the decision to divert to airport C was not highly related to appraising the facilities there beforehand. However, it was noticeable from the data that pilots who elected to divert collected far less situation assessment data than those who continued on to their final destination (on average they obtained information from 1.9 sources versus the pilots who elected to carry on, who collected information from 2.7 sources). A similar but less pronounced pattern was observed with pilots who would elect to restrain the passenger. They also gathered less situation assessment in formation (2.2 versus 2.5 sources investigated). Only in one case did an interviewee elect to restrain the disruptive passenger *and* divert. It would seem that these options are almost mutually exclusive. In all cases where the disruptive passenger was either restrained or the pilot elected to divert, the police at the arrival airport were notified. It would seem that the more extreme methods for dealing with the 'air rage' incident are taken on the basis of the least situation assessment information. These data suggest that in these cases the pilot seem to decide that 'they have seen enough' and rapidly make a 'conservative' (or 'safe') decision on how to deal with the situation.

NN analysis and interpretation of the data set

Having coded all 65 interview transcripts in the manner described in the earlier section describing the treatment of the raw data, the data set was randomly split into two portions of approximately two-third and one-thirds. The larger set of 42 cases was used as the training set for the neural network; the smaller set of 23 cases was used as the data set to cross validate the solution derived (see Garson, 1998).

The data set was then entered into the binary version of the NN analysis programme Neuroshell™ (1990). This version is specifically designed for the analysis of binary, categorical data. Neuroshell is based on the multilayer perceptron and utilises a back-propagation method for controlling the learning rate and assessing the convergence of the NN model.

The 'situation assessment' variables were used as the input variables (nodes) to the model and 'subsequent actions/decisions' were used as the output nodes. Five cases were deleted from the analysis as their pattern of input variables exactly duplicated that of another case in the training set. The software user manual (Neuroshell, 1990) suggests that the deletion of duplicate cases is likely to produce a better final solution.

Nine nodes were used in the 'hidden layer' based on the formula for the estimation of the optimum number of hidden nodes suggested in the user manual

(Neuroshell, 1990). This is double the square root of the number of input nodes plus the number of output nodes. This number is also well within the maximum number of hidden nodes suggested by Garson (1998), which is half the number of cases in the training data set. Too few hidden nodes and the solution will fail to converge; too many and the solution produced tends to overfit the training data set (*i.e.* the answer produced will be specific to the training data set and will not generalise to the cross validation data set: in this latter case, the correct classification rate for these cases will be poor). The NN model converged using just these nine hidden nodes with all cases having an error of <0.05. The total error rate (error sum of squares) was 0.59.

Table 2 Weights from the situation assessment variables (input nodes) to the hidden nodes and their associated activation functions (biases), for the NN model derived from the derivation set of data

Situation assessment actions (input nodes)	Weight to hidden node number:-								
	1	2	3	4	5	6	7	8	9
Bias	7.4	6.5	-4.3	-3.5	-23.4	4.1	1.3	0.7	-0.7
Establish if the passenger is travelling alone	-7.7	19.1	12.6	-7.9	9.4	11.2	2.3	-3.2	2.3
Send a pilot back to the aircabin to appraise the situation	-3.7	-9.0	-6.5	6.0	13.2	11.7	6.9	2.6	0.2
Establish if there are any off-duty crew travelling that may help	8.1	5.1	-21.0	-4.0	2.4	-7.8	-6.2	2.8	-1.8
Get a full description of the situation from the cabin crew	-2.9	-3.0	6.4	7.8	8.1	-18.2	-8.8	-5.1	0.3
Get a summary of the problem from the disruptive passenger	-23.0	-9.9	13.9	2.2	15.6	-4.5	-5.0	-5.4	2.4
Appraise facilities at nearer, alternate airport (airport C)	16.2	-1.5	-17.2	12.3	15.6	0.8	-1.9	-3.2	-2.7
Ascertain if there were any problems prior to boarding	-4.0	10.3	-13.1	21.8	4.7	8.4	7.1	-4.0	-2.3
Establish if he is a large man	-27.3	-11.1	19.8	5.4	-16.7	6.4	2.4	-1.5	4.2
Establish if he is on medication	13.4	-7.6	-3.4	-12.6	-1.1	-6.9	-0.3	2.3	-0.2
Collect personal details about the passenger	-2.2	0.1	24.7	-11.0	-8.7	-1.0	5.3	-7.6	3.7
Appraise the facilities in the country being overflown	-5.8	0.0	1.0	-9.6	-12.9	-1.9	-2.3	2.9	0.5

The figures in table 2 correspond to the weights on the paths from the input nodes to the hidden nodes (as described in figure 1). As the input nodes were of a binary format, coded '1' for gathering situation assessment information and '0' for not doing so, the weight also represents that path's contribution to making a hidden node fire (or not). A large positive weight will encourage a hidden node to fire: a large negative weight will help inhibit the node. The biases (calculated by the NN program) represents the threshold above which hidden node will fire.

Table 3 Weights from the hidden nodes to the and 'subsequent actions/decisions' variables (output nodes) and their associated activation functions (biases), for the NN model derived from the derivation set of data

	Weight to output node:-								
Hidden node number	Get friends or other able-bodied passenger to help	Move him away from other passengers	Refuse to serve any more alcohol	Contact the company	Organise police at destination airport	Make a general warning on the public address system	Issue a specific warning to the disruptive passenger	Physically restrain the passenger	Divert to nearer alternate airport (airport C)
Bias	0.8	1.6	-8.9	-0.7	-14.1	-7.7	2.8	-2.4	-1.3
1	-0.9	-3.8	-1.7	-0.4	2.8	4.1	-6.0	3.8	4.1
2	2.5	0.4	4.5	1.5	3.2	1.5	3.2	-0.5	-5.7
3	1.3	-1.6	-3.1	-6.9	-0.5	4.7	-7.6	-0.8	5.2
4	-1.7	1.0	1.5	2.7	6.6	0.8	2.6	-3.2	0.8
5	2.8	-1.5	5.5	-0.2	-0.6	5.5	-1.2	-0.8	-1.0
6	-1.7	0.7	-8.5	-12.3	2.5	-8.9	2.9	5.3	11.4
7	0.3	1.1	6.7	13.3	-4.0	9.6	-3.7	-2.1	-11.6
8	-3.6	-1.8	6.3	-1.8	5.4	-0.3	1.4	-5.4	-0.6
9	0.6	-3.3	6.9	2.9	10.3	-2.6	4.0	1.6	-2.2

The hidden nodes also produce a binary output. Table 3 shows the connection weights between the hidden nodes and the output nodes (decisions and actions taken). As the output from the hidden nodes is also binary, the weights are again essentially the contribution to making an output node 'fire'.

Table 4 Frequency and percentage of 'subsequent actions/decisions' variables (output nodes) categorised as being either correct or incorrect in the cross validation data set

Subsequent actions/decisions (Output nodes)	Correctly Classified		Incorrectly Classified	
	Hits	*Correct rejections*	*Misses*	*False alarms*
Get friends or other able-bodied passenger to help	20	3	0	0
Move him away from other passengers	1	22	0	0
Refuse to serve any more alcohol	1	22	0	0
Contact the company	8	15	0	0
Organise police at destination airport	11	12	0	0
Make a general warning on the public address system	9	14	0	0
Issue a specific warning to the disruptive passenger	15	8	0	0
Physically restrain the passenger	2	21	0	0
Divert to nearer alternate airport (airport C)	8	15	0	0

When analysing the veracity of the predictions made by the NN model a signal-detection theory-based approach was used. The predictions made for each output node (decision) in the cross-validation data set were categorised as 'hits' (correct predictions of actions taken); 'misses' (action taken but NN predicted that this category of action would not be taken); 'false alarms' (where it was predicted that an action would be taken but actually it was not taken); and correct rejections (predictions that a category of action would not be taken which was proven to be correct).

As can be seen in table 4, the NN described in tables 2 and 3 produced a 100% correct classification rate for all the output variables (decisions) based upon the input variables (situation assessment information). There were no 'misses' or 'false alarms' for any output variable.

Discussion

No previous study has examined the decision-making of commercial pilots in a disruptive passenger incident. The discussion of the interviewees' situation assessment activities and subsequent actions in the hypothetical disruptive passenger incident has already been presented the previous section describing the qualitative analysis of the data gathered. The emphasis in this section will be on assessing the utility of artificial NNs for the modelling of NDM. To re-iterate, as the study was conducted within an NDM paradigm the interviewees were restricted in neither the sources of information that they could interrogate to assess the situation nor in the number or type of actions that they could pursue as a result of the information gathered. A certain degree of artificiality has been introduced into the study through the subsequent content analysis and coding of the interview transcript data to prepare it for analysis using the NN shell, however, as far as possible, it has been attempted to maintain this analysis within the NDM tradition.

Before proceeding any further it is worth issuing a word of warning about the evaluation of the results from a NN. While it may be tempting to infer cause and effect relationships between situation assessment variables and decisions (as in the description of selected results in the previous section) the interpretation of individual variables in a NN should be approached with caution. The neural model *as a whole* should be interpreted for its efficacy in predicting outcomes rather than interpreting the contribution of individual variables. In this instance the NN correctly predicted 100% of all the decisions for each of the nine output variables that each pilot in the cross-validation data set actually made. This is extraordinarily high and is uncommon even for a neural network.

Some insight may be gained about the manner in which a given situation assessment NN input variable affects a subsequent decision by studying the weights to and from the relevant hidden nodes. However, it again needs to be emphasised that variables can only be evaluated in the context of the other variables and their stated relationships. The value in an NN lies in the model as a whole, *not* its individual components or specific relationships within it. In this case it would be wrong to suggest that one input (situation assessment) variable is more important than another in determining the final course(s) of action chosen (cf. the weights in multiple regression). The following description merely describes the manner in which a variable operates with a NN.

Consider the weights to the hidden nodes from the situation assessment variable appraising facilities available at airport C. If this situation assessment activity was made, the input node took a value of unity. This value is then multiplied by the various weights to each of the hidden nodes. Its effect begins to spread throughout the NN. The sums of all the weights from each input variable (situation assessment) variable are computed at each hidden node. If the sum of the input functions to the hidden node exceeds a critical value (essentially the bias in this case as thresholds were set to 0.01) the node 'fires' and an output from the hidden node is made. For simplicity, only the larger weights in this illustration will be considered. From the node associated with the pilot appraising the facilities available at the diversionary airport, it can be seen that this input variable will help to activate hidden nodes 1, 4 and 5, and suppress (inhibit from firing) hidden node 3 (see table 2). From table 3 the influence of hidden node 4 (for example) can be established on the outcome (decision) variables. Taking just the larger weights associated with this node to illustrate further this principle, this element will help 'fire' the output variable that suggests the pilot will organise the police at airport C pilot but will simultaneously help suppress the firing of the output node suggesting that the pilot will call for the passenger to be physically restrained.

While the above explanation gives some insight into how the inputs, hidden nodes and outputs in the NN operate it is essential that only the overall results should be interpreted, not individual paths between nodes. The NN is as good as the combined effect of its components. The greatest problem faced by the researcher lies in the interpretation and evaluation of NN models. The efficacy of a model can only be evaluated in its ability to predict accurately outcomes. In this case the evaluation of the accuracy with which each output (decision) variable was predicted was evaluated using a signal detection theory based approach. This allowed the nature of any misclassification of a variable (as a 'miss' or a 'false alarm') to be identified.

The adequacy of the input variables to the NN is a little harder to assess, however, a poor input variable (one that makes little or no contribution toward predicting the NN output values) will adversely affect the whole of the network. In the instance of including a poor input node into a NN model it is likely that either the solution will not converge or that the predictions made by the model will be inaccurate upon cross validation. Unfortunately, there is no other way to identify poor input nodes other than by trial and error. The extremely high correct classification rate of the decisional outcomes in the cross-validation data set (see table 4) would suggest that the situation assessment variables elicited in the interviews, taken as a set, are good predictors of the decisions a pilot will make when dealing with a disruptive passenger, once the NN model described in tables 2 and 3 is applied.

The main issues in a NN such as the one produced in this study are basically concerned with the content and the criterion validity of the model produced. The criterion (predictive) validity is relatively easy to establish through the use of the signal detection theory based approach described previously. The greatest challenge lies in assessing the content validity of the NN model. In this case the question of content validity applies to both situation assessment inputs and decision/action outputs of the NN model. Kerlinger (1973) suggests the researcher should ask themselves the question 'is the substance or content of this measure representative of the universe of content of the property being measured?' (p. 458). In the case of the NN modeller it is essential to establish if the input and output nodes (situation assessment and decisions/actions) are an exhaustive set of all the pieces of information that the decision maker would interrogate and a reasonably representative set of the universe of possible (likely) subsequent actions. It should be noted that at any one point in time a human being actually has an almost infinite number of possible behaviours open to them (some are more likely thank others, though)! As a result it is not possible to define the entire universe of behaviours on the output side of the model. Content validity can never be totally established. It can only be reasonably assured by employing appropriate data gathering and analytical methodologies. In this case it has been attempted to ensure content validity by employing a reasonably large sample of suitably qualified pilots with a range of experience and by employing a data gathering technique commensurate with the NDM paradigm.

The present study does, however, have some shortcomings in that it was not conducted within the context of either a flight deck or a real 'air rage' incident. Context, time pressures and multiple actors are important aspects of decision making, as recognised by many NDM theorists (e.g. Klein, 1989; Orasanu 1993; Orasanu and Connolly, 1993). While the study presented the information to the interviewee in a manner similar to the way it would be presented in a real incident, some other aspects of the situation were different. For example, the interviewee did not have the opportunity to discuss options with other colleagues on the flight deck and the time-scale for the interview did not reflect the time-scale of an actual incident. All the sources of situation assessment information and the decisions subsequently made were proffered in a short period of time. The opportunities for collecting information, implementing a decision and monitoring the situation, as would be done in a real-life situation, were limited. However, the methodology employed was not all that different to that employed in many other NDM studies which have utilised retrospective introspection about the manner in which a decision was reached and were necessary given the methodology employed.

This study does strongly suggest, though, that the use of an artificial NN may be a way of empirically modelling NDM. NNs have been used with success in similar applications in the past, although not with such a high correct classification rate on the cross-validation sample. This may be an artefact of only a relatively small number of input and/or output variables being used in the final model. Models that have greater numbers of situation assessment variables and decisional outcomes may show a lower correct classification rate on cross validation. This requires further investigation. The 'richness' and 'complexity' of the NNs produced by pilots of different levels of experience may provide some insight into the key factors in the decision making process. Cellier, Eyrolle and Mariné (1997), in an overview of expertise in decision making in dynamic environments observed that novice and expert representations of systems differed considerably. In some instances experts had a far more complex internal representation of the system than the novices did and in other instances quite the opposite was true. Experts had a far simpler, functional representation. It is possible that these representations of system complexity could be reflected in the NNs produced by experts and novices. This approach may be well worth considering to help validate the use of NNs as an approach for the modelling of naturalistic decisions.

Despite the need for considerably more work to evaluate the utility of using NNs, to model naturalistic decisions the present work has demonstrated that multiple decisional outcomes can be accurately predicted from a model based upon multiple sources of situation assessment information. The question remains, however, if NNs can be used for other categories of NDM tasks other than what is essentially a pattern-matching decision making task, for example 'consequential choice' and 'reassessment' type NDM tasks (Lipshitz, 1993). However, in the present safety-related scenario, the model has performed commendably in predicting the decisions made in an unstructured situation.

References

Cannon-Bowers, J.A., Salas, E. and Pruitt, J.S. (1996). Establishing the boundaries of a paradigm for decision-making research. *Human Factors, 38*, 193-205.

Cellier, J.-M., Eyrolle, H. and Mariné, C (1997). Expertise in dynamic environments. *Ergonomics, 40*, 28-50.

Connolly, T. (1988). Hedge-clipping, tree felling and the management of ambiguity. In, R. Boland, L.R. Pondy and H. Thomas (Eds) *Managing the challenge of ambiguity and change*. Chichester, UK: Wiley.

Department of Transport (1995). The Air Navigation (No. 2) Order. London: Her Majesties Stationary Office.

Dowell, J., Smith, W. and Pidgeon, N. (1997). Design of the natural: an engineering process for naturalistic decision making. In R. Flin, E. Salas, M. Strub and L. Martin (Eds.) *Decision making under stress*. Aldershot: Ashgate.

Garson, D. G. (1998). *Neural Networks: An Introductory Guide for Social Scientists*. London: Sage.

Hair, J.F., Anderson, R.E., Tatham, R.L. and Black, W.C. (1998). *Multivariate Data Analysis (5^{th} Edition).* Upper Saddle River, NJ: Prentice-Hall.

Hammond, K.R. (1993). Naturalistic decision making from a Brunswikian viewpoint: Its past, present and future. In, G.A. Klein, J. Orasanu, R. Calderwood and C. E. Zsambok (Eds.) *Decision making in action: Models and methods* (pp. 205-227). Norwood, NJ: Ablex.

Hammond, K.R. (1988). Judgement and decision making in dynamic tasks. *Information and Decision Technologies, 14,* 3-14.

Haykin, S. (1994). *Neural Networks: a Comprehensive Foundation.* New York, NY: Macmillan.

Hicks, B. and Morrison, R. (1997). Passenger Related Safety Hazards. *ASRS Directline 9.* Washington D.C. NASA.

Howell, W.C. (1997). Progress, prospects and problems in NDM: A global view. In, C.E. Zsambok and G. Klein (Eds.) *Naturalistic decision making* (pp. 37-46). Mahwah, NJ: Lawrence Erlbaum Associates.

Jack, M. (1998). Disruptive Passengers: an Airline Strikes Back. *Presentation at the United Kingdom Flight Safety Committee Seminar on Passenger Behaviour,* (3 November) London: Heathrow.

Kerlinger, F.N. (1973). *Foundations of Behavioral Research (2nd Edition).* New York: Holt Rinehart and Winston.

Klein, G. (1989). Recognition Primed Decisions. In W. Rouse (Ed) *Advances inman-machine system research, volume 5* (pp. 47-92). Greenwich, CT; JAI.

Lehto, M.R. (1997). Decision making. In, G. Salvendy (Ed) *Handbook of Human Factors and Ergonomics (2nd Edition)* pp. 1201-1248. New York: John Wiley.

Leven, D. and Levine, D.S. (1996). Multiattribute decision making in context: A dynamic neural netweork methodology. *Cognitive Science, 20,* 271-299.

Lipshitz, R. (1993). Converging Themes in the Study of Decision Making in Realistic Settings. In, G. Klein, J. Orasanu, R. Calderwood and C. Zsambok (Eds) *Decision Making in Action: Models and Methods* (pp. 103-137). Norwood, NJ: Ablex Publishing Corporation.

Longmuir, S. (1998). Testimony given at US Committee on Transportation and Infrastructure Subcommittee on Aviation Hearing, 11 June. Cited in, R.Bor (1997). Unruly passenger behaviour and in-flight violence: a psychological perspective. *Travel Medicine International, 17,* 5-9.

Montgomery, H. (1989). From cognition to action: the search for dominance in decision making. In, H. Montgomery and O. Svenson (Eds.) *Process and Structure in Human Decision Making* (pp. 182-187). Chichester, UK: Wiley.

Moore, B. (1988). ART1 and Pattern Clustering. In, D. Touretzky, G. Hinton and T. Sejnowski (Eds.) *Proceedings of the 1988 Connectionist Model Summer School* (pp. 174-185). San Mateo, CA: Morgan Kaufmann.

Moray, N. (1998). The psychodynamics of human-machine interaction. In, D. Harris (Ed) *Engineering Psychology and Cognitive Ergonomics - Volume 4* (pp. 225-235). Aldershot: Ashgate.

Neuroshell User Manual - 3rd Edition. (1990). Frederick, MD: Ward Systems Group.

Nobel, D. (1993). A Model to Support the Development of Situation Assessment Aids. In, G. Klein, J. Orasanu, R. Calderwood and C. Zsambok (Eds.) *Decision Making in Action: Models and Methods* (pp. 287-305). Norwood, NJ: Ablex Publishing Corporation.

Orasanu, J. (1993). Decision making in the cockpit. In, E.L. Wiener, B.G. Kanki and R.L . Helmreich (Eds.), *Cockpit Resource Management* (pp. 137-172). San Diego: Academic Press.

Orasanu, J. and Connolly, T. (1993). The reinvention of decision making. In, G. Klein, J. Orasanu, R. Calderwood and C. Zsambok (Eds.) *Decision Making in Action: Models and Methods* (pp. 3-20). Norwood, NJ: Ablex Publishing Corporation.

Prince, C. and Salas, E. (1997). Situation assessment for routine flight and decision making. *International Journal of Cognitive Ergonomics, 1*, 315-324.