Selecting Human Error Types for Cognitive Modelling and Simulation

Tina Mioch, Jan-Patrick Osterloh, Denis Javaux

Abstract This paper presents a method that has enabled us to make a selection of error types and error production mechanisms relevant to the HUMAN European project, and discusses the reasons underlying those choices. We claim that this method has the advantage that it is very exhaustive in determining the relevant error types and error production mechanisms, and that the final objects are selected according to explicit requirements, without missing relevant error types and error production mechanisms.

1 Introduction

It is a well known fact that human errors are the main contributing factor in aviation incidents and accidents (according to a Boeing study in 2004, human errors are involved in 62% of the accidents [1]). The main objective of the European project HUMAN is to build a methodology for human error prediction that is applicable in early phases of the design of a new system. The method foresees to simulate the interaction with the system by means of a dynamic cognitive model. By having a model of the crew, human behaviour is predicted, including specific types of errors considered as relevant by the manufacturers, cf. [5]. To achieve this objective, we rely on two functionally equivalent simulation platforms, a physical simulation plat-

Tina Mioch

TNO Human Factors, Kampweg 5, 3796 DE Soesterberg, The Netherlands, e-mail: `tina.mioch@tno.nl`

Jan-Patrick Osterloh

OFFIS Institute for Computer Science, Escherweg 2, 26127 Oldenburg, Germany, e-mail: `osterloh@offis.de`

Denis Javaux

NEXT STEP SOLUTIONS, Rue Daussoigne-Mehul 20, 4000 Liege, Belgium, e-mail: `denis.javaux@skynet.be`

form, comprising a full scale simulator usable for human-in-the-loop experimental simulations, and a virtual simulation platform. The virtual simulation platform will be used to produce predicted crew activities (with a special focus on human errors) on a series of dedicated experimental flight scenarios also used on the physical platform. The behaviour of the pilots in the experiments on the physical simulation platform is then compared with the behaviour of the cognitive model on the virtual simulation platform, to validate and improve the cognitive model.

A major input for the development of the cognitive model is the selection of the errors on which the HUMAN project will focus. The cognitive model should be able to predict errors that can occur during interaction with a target system, which in HUMAN is a cockpit system known as the Advanced Flight Management System (AFMS), which user interface is the Advanced Human Machine Interface (AHMI). We also need to determine the cognitive mechanisms that cause the selected error types. In this paper, we therefore differentiate between the concepts of error type (ET) and error production mechanism (EPM). ETs are the observable behaviour (phenotypes), while EPMs are the mechanisms by which error types occur (genotypes). This paper will describe the process by which the ETs and EPMs for HUMAN have been selected.

2 Method

To select ETs (Error Types) and EPMs (Error Production Mechanisms) appropriate for HUMAN we decided to apply the following strategy, based on 6 steps:

1. Definition of requirements for error taxonomies. As we use existing taxonomies as a source for the ETs and EPMs, and some in domains very remote from human factors (e.g., insurance companies), we need to make a selection, and therefore define requirements our source taxonomies have to comply with.
2. Literature review.
3. Selection of relevant taxonomies. A subset of the candidate taxonomies found in step 2 are selected, based on the requirements defined in step 1. These are our source taxonomies that are particularly relevant for our target domain and the framework of the HUMAN project.
4. Definition of requirements for ETs and EPMs. The source taxonomies contain plenty of ETs and EPMs, with only a subset of them useful for HUMAN. A selection therefore has to be made based on requirements specified for ETs and EPMs.
5. Documentation of ETs and EPMs in the selected taxonomies. To perform the selection of the final ETs and EPMs (step 6), each ET and EPM first has to be described along a series of dimensions, relevant to the requirements specified in step 4.
6. Selection of final ETs and EPMs. The ETs and EPMs found in the source taxonomies (step 4) and documented in step 5 are evaluated according to the requirements, ratings are provided, and the final selection is made.

Several general principles guided us during the performance of these six steps. First of all, the goal was to be as exhaustive as possible. For that reason, we first produce candidate taxonomies, and then candidate ETs and EPMs. Also, the selection processes should be as rational and explicit as possible. This is why we specify explicit requirements, for taxonomies as well as for ETs and EPMs. Last, the selection processes should be as 'democratic' and efficient as possible, making use of all resources available to us. Every potential contributor in the project is be required to provide inputs.

Step 1: Definition of Requirements for Error Taxonomies The selection of a scope in which to investigate human error taxonomies, error types and error production mechanisms is particularly important. The ETs and EPMs selected will become the central focus of HUMAN. The scope we selected for error taxonomies is 'The modern cockpit in a future ATM environment', since it relates to the overall scope of HUMAN itself. Aiming at something larger (e.g., other vehicle types than aircraft) would be too ambitious and yield selected ETs and EPMs which, despite being very interesting, are beyond the modelling and investigation capabilities of HUMAN.

Having determined the scope, we identified requirements for selecting the taxonomies actually relevant for HUMAN: Beside that the taxonomies have to be relevant within the defined scope, they should cover the most relevant ETs within the defined scope, and should include either observable error characteristics (phenotypes) for each ET, or the taxonomies should focus on understanding the cognitive process involved in the production of human error (genotypes) and the associated EPMs. In addition to this, the taxonomies should refer to ETs and EPMs that have either a significant frequency of occurrence, or whose occurrence is particularly safety threatening. In order to easier defer the EPMs, taxonomies with strong theoretical or methodological foundations should also be preferred, and they should be well-established and well-tested.

Step 2: Literature Review The first step was collecting papers on human error taxonomies and human errors in general, to get a better acquaintance on existing taxonomies. In this initial research, we did not limit our search to aviation, but also included other safety critical domains. In addition, we collected also literature about the relevancy of errors to the aviation. Listing this literature in this paper would exceed its scope. A very interesting and exhaustive error taxonomy literature survey can be found in [4]. The authors of this document categorized the taxonomies by their different foundations:

- *Task-based taxonomies* mostly describe lists of 'external error modes', which refer to the structure and elements of the external human task and classify the overt characteristics of the error. An examples for these taxonomies is [11].
- *Communication system models and taxonomies* mostly deal with mass communication, and are not primarily models of cognition. However, some of the models can be used to model communication within HUMAN, e.g. [10].

- *Information processing models and taxonomies* examine human performance by attempting to trace the information flow through several processing stages from information input to response output, e.g. [12].
- *Symbolic processing models and taxonomies* regard humans and computers as general purpose symbol manipulating systems, and are closely related to the information processing tradition, and related to artificial intelligence and cognitive science, e.g. [6, 9, 8].
- *Other models and taxonomies* bundle other taxonomies, not fitting in the above schema, e.g. the situation awareness error taxonomy of [2].

We used this categorization, but added also other interesting taxonomies that were not described in [4], e.g. [7, 3]. For each paper on our list of potential error taxonomies, we wrote small summaries, to make it accessible for all partners involved in the selection process. This list has then been used to select these taxonomies, fitting to our requirements and are of particular interest for HUMAN.

Step 3: Selection of Relevant Taxonomies Each partner involved in this task identified, independently of each other, the most relevant human error taxonomies, in accordance with the requirements. By doing this independently, we attempted to ensure that the selection process was exhaustive, as objective as possible, and that no interesting candidate was missed. Noticeably, most of the taxonomies finally selected were selected by all partners. We discussed the proposed taxonomies and their associated papers or reports extensively, taking the requirements explicitly into account. At the end, we decided to add to the commonly chosen taxonomies some of the taxonomies chosen by only a single partner, as this was in line with the requirement of being able to cover most of the relevant Error Types. After these discussions, we all agreed on the following 'source' error taxonomies:

- The *Phenotype* oriented taxonomies of [11]: very relevant within the defined scope, and including observable error characteristics.
- The *Genotype* oriented taxonomies of [8, 9, 6, 7, 2]: very relevant within the defined scope, and focusing on understanding the cognitive processes involved in human error production.

The lists above not only cover most of the categories of taxonomies in [4], but also additional ones, mentioned in separate literature. The only category we decided to leave out is the communication models and taxonomies, since we had no communication model for HUMAN. To reduce the risk of ignoring or even rejecting important ETs and EPMs, it was decided to ask the other HUMAN experts, namely human factor experts and pilots, to comment on the completeness of the selection and whether the error types in the source taxonomies occurred frequently in the cockpit. The experts approved the selection and did not propose any additional taxonomy or error types. In addition, we also performed cross checks with the FAA error list (FDAI Database).

Step 4: Definition of requirements for ETs and EPMs In order to select ETs and EPM for HUMAN in a very systematic way, with explicit procedures and decision

criteria, we derived the following requirements for ETs and EPMs: 1) The ET/EPM should be frequent in modern flight cockpits. 2) They must be relevant for the HUMAN target system (AFMS and AHMI). It must be possible to 3) detect the associated error types (ETs) in the data that we will gather in experiments with human pilots (PSP), and 4) the error production mechanisms (EPMs) behind the ET must be understandable (i.e., it must be possible to derive them from the observations on the PSP). Last but not least, 5) the effort for predicting the EPMs with the cognitive model must be compatible with the resources of HUMAN.

Step 5: Documentation of ETs and EPMs in the selected taxonomies With the list of selected taxonomies from step 3, we have a list of potential ETs and EPMs. We documented the potential error types on different dimensions, according to the requirements derived in step 4. Seven human factor experts and pilots evaluated the frequency and relevance of the ETs, and three cognitive modelling experts evaluated the possibility to detect the ETs and evaluated the EPMs on understandability and effort for predicting them.

Documentation of ET: *Entire task omitted*	
Description	
Occurs when the entire task to be achieved is omitted	
Example	
The crew do not execute the after takeoff checklist	
Theoretical affiliation(s), if any	
None	
References	
Swain (1982), Swain & Guttman (1983)	
Error production mechanism(s)	
- inadequate planning	- learned carelessness
- loss of information in working memory	- ineffective prospective memory
- routine capture	
Cognitive process(es) affected by the error type	
Plan execution, procedure execution	
Plausible frequency of occurrence with AHMI (in GECO)	**Relevance for cockpit crew in ATM environment**
Expert 1 — rare 1 **2** 3 4 5 frequent	Not relevant 1 2 3 **4** 5 relevant
Expert 2 — rare **1** 2 3 4 5 frequent	Not relevant 1 2 3 **4** 5 relevant
Expert 3 — rare **1** 2 3 4 5 frequent	Not relevant 1 2 3 **4** 5 relevant
Expert 4 — rare **1** 2 3 4 5 frequent	Not relevant 1 **2** 3 4 5 relevant
Expert 5 — rare **1** 2 3 4 5 frequent	Not relevant **1** 2 3 4 5 relevant
Expert 6 — rare 1 2 3 **4** 5 frequent	Not relevant 1 2 3 **4** 5 relevant
Expert 7 — rare 1 2 3 **4** 5 frequent	Not relevant 1 2 3 4 **5** relevant
Comments (optional)	
The relevance is derived from the fact that there are some tasks like transitional checklists that cover items that had been performed in the past. Such as the gear retraction during take-off. The associated after take-off checklist is read considerably later. As tasks like gear retraction are so obvious, any checklist containing mostly items of similar obviousness bears the risk of omission.	
Possibility to detect the error types	
Expert 8 — easy **1** 2 3 4 5 hard	
It obviously very easy: the whole task is missing (e.g. no take-off briefing)	
Expert 9 — easy 1 **2** 3 4 5 hard	
Based on behaviour data, Eye movements (as part of behaviour data), Comparison with normative behaviour	
Expert 10 — easy **1** 2 3 4 5 hard	
Seems to be easy to detect	

Fig. 1 Example of documentation of the different dimensions for the ET *Entire task omitted*

Figure 1 shows an example of documentation for one of the ETs, namely Entire task omitted. This error type belongs to the phenotypical, task-based taxonomy [11]. For each identified EPM, the two dimensions *predictability* and *understanding* are evaluated, for example for Learned Carelessness, see Figure 2. After the ratings have been collected, the mean values of all ratings constituted inputs for the selection process in step 6. Discussions were held when very conflicting ratings were provided by the experts.

Step 6: Selection of final ETs and EPMs The final selection is based on the documentation of the ETs and EPMs described above. The results of the documentation are aggregated into a single table, for each candidate ET and EPM, to help with the selection process. This allows us to derive associated cost (in terms of development effort), and the difficulty and risk (of failure) for the project for all ETs and EPMs. We used a series of formulas in the table to calculate the 'interestingness' of the ETs and EPMs in a formal, explicit, and as objective as possible way. An ET will be considered interesting if it is frequent, relevant for HUMAN and easy to detect on the physical simulation platform. A formula computes a value based on these three ratings and considers the ET as 'interesting' if the value obtained is above a threshold I_{ET}. An EPM will be considered interesting if it is easy to understand and to predict. Another formula computes a value based on the two respective ratings and considers the EPM as 'interesting' if the value obtained is above threshold I_{EPM}.

Another formula is used to calculate a recommendation for selecting ETs, i.e. an ET is recommended for selection, if it is above I_{ET}, and it has at least one contributing EPM with a calculated value above I_{EPM}. Even though these formulas provided a first set of candidate ETs, we developed two further formulas for the final selection by the experts: The first formula computes the 'Return on Investment' (ROI) for the HUMAN project associated with a given ET (and its underlying EPMs). ETs with a high level of interestingness and whose EPMs are interesting and easy to implement have a high ROI. Second, the ROI value obtained for each ET is then processed to determine if the ET should be considered of high priority (ROI above a threshold P_{ET}) and therefore addressed early in the project, or of a lesser priority, to be addressed at a later stage. The final selection has been performed by the whole HUMAN consortium. The implications of the selection of specific ETs and their EPMs, in terms of developments for the cognitive model and the experiments with human pilots have been discussed by the consortium. The final selection was interestingly close to the recommendations determined by the formulas.

Documentation of EPM: *Learned Carelessness*							
Predictability with the cognitive model:							
	SUM	4	MEAN	2		FINAL RATING	2
Expert 9	easy 1 **2** 3 4 5 hard						
Expert 10	easy 1 **2** 3 4 5 hard						
Observability understanding							
	SUM	13	MEAN	4,3		FINAL RATING	4,3
Expert 8	easy 1 2 3 **4** 5 hard						
	It's hard (4), if not very hard (5) to prove that a simplified action pattern observed in the simulator is the by production of learned carelessness. This can only be done with some confidence when the predictive capabilities we have regarding these implicit learning phenomena exactly predict the simplified pattern observed. In this case we may have good reasons to believe they are at play.						
Expert 9	easy 1 2 3 **4** 5 hard						
	- video based post-interviews - eye movements			- based on behaviour data (including historical data over several episodes) - rule identification based on behaviour data			
Expert 10	easy 1 2 3 **4** 5 hard						

Fig. 2 Documentation of the EPM *Learned Carelessness*

3 Results and Discussion

We claim that this method has the advantage that it is very exhaustive in determining the relevant error types and error production mechanisms, and that the final objects are selected according to explicit requirements.

The result of this method is the list of ETs and EPMs that are investigated in HUMAN, see Table 1. Please note that the list is of course specific to HUMAN, because they fit the specific requirements and objectives of the project.

	Learned Carelessness	Cognitive Lockup	Loss of Information in Working Memory	Action Priming	Routine Capture	Selective Attention	Saliency Bias	Recency Bias	Simple Omission	Distraction due to External Events
Priority	1	1	2	2	2	2	2	2	2	2
Entire task omitted	x		x		x				x	
Step(s) in task omitted	x		x		x				x	
Attentional Slips		x		x		x				
Application of bad rules	x	x			x					
Biases							x	x		
External activation Errors		x				x				x
Misordering of action sequences errors		x								
Failure to monitor or observe data		x				x			x	

Table 1 Overview of selected ETs and EPMs

As mentioned before, we have tried to be exhaustive, rational, explicit, democratic and efficient, at all steps of the selection process. We however encountered some difficulties. First of all, the definition of what constitutes a human error is controversial and the object of many discussions. For the HUMAN project, we have favoured a definition that relies on the notion of deviation from prescribed or acceptable activities. Error production mechanisms on the other hand are the mechanisms by which error types occur. The mechanisms are related to normal cognitive processes that do not perform optimally (variance of human performance), for a large variety of exogenous and endogenous causes, such as high workload, fatigue, inappropriate design of information display, and contribute to the occurrence of the error types.

In addition, the taxonomies obtained after surveying the literature were coming from very different sources, they were aimed at rather different objectives, and were relevant to multiple domains. The Error Types (ETs) and Error Production Mechanisms (EPMs) found in the source taxonomies were not always homogeneous or equivalent: identical or similar ETs or EPMs were sometimes described with different names, or identical names were used for different things. The levels of granularity within the taxonomies were also sometimes very different.

The distinction between ETs and EPMs is not a clear one. EPMs cause ETs, but sometimes EPMs cause other EPMs. The more we progressed in the project,

the more we understood that we were faced with causal chains or trees, not solely with a simplistic dissociation between two categories (ETs and EPMs). It was too late, given the constraints of European projects, to get back and redo our initial structuring of ETs and EPMs. We therefore decided to pay more importance to the EPMs than to the ETs for the remaining of the project. The EPMs are central to the project, since they are the error mechanisms we have to describe and implement in the cognitive model.

We have tried to respond in the most optimal way, given our local constraints, to each of these difficulties, always having in mind the principles mentioned above. We believe the resulting list of targeted ETs and EPMs to be far more appropriate to the peculiarities, limits and constraints of our project than it would have been if the selection process had been made in a more implicit and subjective way, by a local set of 2 or 3 participants in charge of this specific task.

References

1. Boeing Commercial Airplanes: Statistical Summary of Commercial Jet Aircraft Accidents, Worldwide Operations, 1959-2003. Seattle, WA: Boeing Commercial Airplane (2004)
2. Endsley, M. R.: Situation Awareness and Human Error: Designing to Support Human Performance. Proceedings of the High Consequence System Surety Conference, Albuquerque, NM. (1999)
3. Hollnagel, E. :Cognitive Reliability and Error Analysis Method (CREAM) Elsevier Ltd. (1998)
4. Isaac, A., Shorrock, S. T., Kennedy, R., Kirwan, B., Andersen, H., Bove, T.: Short Report on Human Performance Models and Taxonomies of Human Error in ATM (HERA) (No. 1.0). Brussels, Belgium: EUROCONTROL (DIS/HUM) (2002)
5. Lüdtke, A., Osterloh, J.-P., Mioch, T., Rister, F., Looije, R. Cognitive Modelling of Pilot Errors and Error Recovery in Flight Management Tasks. Proceedings of the 7th Int. Working Conference on Human Error, Safety, and System Development, Brussels, Belgium, Sept. 23-25, (2009)
6. Norman, D. A.: Design rules based on analyses of human error. Commun. ACM, 26. 254-258. (1983)
7. Orasanu, J., Martin, L.: Errors in Aviation Decision Making: A Factor in Accidents and Incidents. Proceedings of HESSD 98: 2nd Workshop on Human Error, Safety, and System Development. Seattle, WA, USA, pp. 100-106, (1998)
8. Rasmussen, J.: Information Processing and Human-Machine Interaction: An Approach to Cognitive Engineering New York, NY, USA, Elsevier Science Inc. (1986)
9. Reason, J.: Human error. Cambridge University Press. (1990)
10. Shannon, C. E., Weaver, W.: The Mathematical Theory of Communication. University of Illinois Press, Urbana, IL. (1949)
11. Swain, A.D. and Guttmann, H.E.: A handbook of human reliability analysis with emphasis on nuclear power plant applications. NUREG/CR-1278, USNRC, Washington, DC 20555. (1983)
12. Wickens, C.D., Hollands, J.G.: Engineering Psychology and Human Performance. Prentice Hall, Upper Saddle River, New Jersey, USA, (1999)