Analysis of the user behaviour when interacting with systems during critical situations

Yuska P. C. Aguiar[1,2], Maria de Fátima Q. Vieira[1], Edith Galy[3], Charles Santoni[4]

[1]LIHM, Electrical Engineering Department, UFCG Campina Grande, Brazil
[2]Department of Exact Sciences, UFPB Rio Tinto, Brazil
[3]Université de Provence (Aix-Marseille I), Département de Psychologie Cognitive et Expérimentale Centre de Recherche PsyCLE Aix-en-Provence, France
[4] Université Paul Cézanne (Aix-Marseille III), Département Génie Industriel et Informatique, LSIS Marseille, France
{yuska@dce.ufpb.br; fatima@dee.ufcg.edu.br; edith.marie@univ-provence.fr; charles.santoni@polytech.univ-mrs.fr}

Abstract. Human error studies tend to focus on identifying the relationship between the human activity, its errors and consequences. Accidents and incidents report analysis has been the path followed by several authors in the human error studies field, as it will be discussed in this chapter. However, reports tend to detail technical aspect of the error occurrence but fail to explore the human-behaviour component that might have influenced it. In order to investigate the human behaviour and its relation with accidents and human errors the authors propose to observe individuals working during critical situations. This observation must adopt a methodological approach, and the authors advise to support it by an experimental protocol, to ensure a rigorous systematization of the data gathering and analysis of the human behaviour. This chapter presents a cognitive model conceived to support this approach, by investigating an individualǿs: characteristics; functional state; situation perception, decision-making approach and performance during task completion which accounts for the knowledge of the work situation. It also briefly presents its supporting experimental protocol, and discusses its application in the context of a decision making aid system, employed during maritime pollution crisis management.

Keywords: human behaviour, critical situations, cognitive models, operating risks (risk and cognition)

1 Introduction

It is widely accepted in the literature that the human error results from failures in the cognitive system. Those failures frequently happen during knowledge acquisition, which comprises a set of cognitive processes and activities that act on sensory information in order to interpret, classify and organize it. To prevent the error occurrence it is important to investigate the risk factors involved and understand the full context in which the error occurs, in order to identify which combined factors can influence the human behaviour and performance, and consequently the task outcome. Once being aware of the risk factors which can trigger the human error, and having identified the factors which may influence task performance, preventive measures can be put into place to: improve the process adopted when performing the task; the tools employed in doing it; and the human operator skills.

This research investigates the adoption of cognitive models to help understanding the cognitive process and anticipate risk-related behaviour in the work context. It focuses in analyzing the error context (when an operator of an industrial automated system is faced with unexpected work situations), combined with the analysis of typical time pressures, anxiety and altered emotional behaviour, in order to anticipate error occurrences. It follows a brief review of the literature which presents the human error as seen from different perspectives by different authors. These do not treat directly the proposed research problem, but give a foundation for the study.

Hollnagel (1993) defines the error as the consequence of a faulty action which leads into unexpected results. According to Rasmussen (1983), when a system presents an unsatisfactory response to a human action, which is different from the expected one, a human error has occurred. Yet, for Reason (1990; 1997), the human error occurs when the consequences of a fault cannot be assigned to external agents, exposing to risks: people, equipment and the surrounding environment. The authors adopt a combination of those views and assume that a cognitive model is at the basis of understanding the human interaction and the task outcome. The following cognitive models adopt complimentary approaches and share the notion that managing cognitive resources to perform a task and deciding which action to perform are influenced by multiple variables.

According to Norman and Draper (1986), the human cognitive process happens in seven stages: formulating objective; formulating intension; specifying action; performing action; perceiving the status of the environment and evaluating results. Therefore, an individual can assess whether an achieved result corresponds to the intended one realizing when there was an error.

The Interaction model proposed by Norman and Baucom (2002) defines an interaction as consisting of two phases: execution and evaluation. The execution phase translates objectives into intentions to perform a task of interest. This intention results into a sequence of actions (mental specifications) which are converted into actions in the physical world. The evaluation phase starts with the perception of the environment in response to the action performed. This perception is then interpreted and compared to the initial objectives.

The SRK (Skill-Rule-Knowledge) model, proposed by Rasmussen (1983), describes qualitatively different modes of information processing during task performance, accounting for a behaviour based on skills; rules and knowledge.

The cognitive model proposed by Endsley (1995) considers that task performance is determined by a decision made after the adequate understanding of the current situation. This decision is processed in three levels; element perception, element understanding and anticipation mechanisms, which allow to foresee the following state of the situation, given that the intended action is executed. Endsley model of mental activities follows a sequence that begins with detecting a sensory signal, followed by accessing the memory, performing logic and intuitive thinking, arriving at the decision making phase and completed by performing the action. System and task characteristics, such as complexity levels, as well as the risk involved in doing the task determine the work load and the levels of vigilance and cognitive control, thus determining the available cognitive resources or mental abilities (Amalberti 1996). Limiting the availability of such cognitive resources, through high spending elsewhere in the task, results in operator behaviour automation (Rasmussen et al 1994); which in turn can lower performance during task execution.

Although the influence of these factors on the cognitive processes, and consequently on the operator performance is widely accepted, there is no explicit representation of those variables in the cognitive models found in the literature, which would help anticipate the human behaviour which leads into error. Given the need for a deeper understanding of the human behaviour, especially when working with risk situations, this chapter proposes a model which accounts for the above variables in an attempt to help understand the human behaviour in such work environments, and thus anticipate the error prone human behaviour. This model´s main objective is to identify internal and external elements which affect operator behaviour and performance, allowing to measure the level of influence of each variable on the observed behaviour; thus enabling to propose changes in the working environment, particularly in the ergonomic aspect of the human interface component of the employed tools. The numbers of variables to account for in the study, as well as the analysis of the cross influence between those, are the main challenges of this research. Nonetheless an initial simplified model has been proposed and an experiment has been performed to test its impact on the human error study. Both the model and experiment are presented in this chapter.

1.2 Critical Systems Usability and Human Error Prevention

In the domain of critical systems, accident and incident report analysis is at the basis of the human error study. Although this is cited in the literature as the main approach taken in the field, as mentioned in the works of: (Rasmussen et al. 1981); (van Eekhout & Rouse 1981) and (Johnson & Rouse 1982), according to Rasmussen (et al. 1981), analyzing the human behaviour when faced with adversities during task performance is also an important source of information on the cognitive mechanisms and strategies employed during activity.

The authors of this chapter adopt the following definition for a critical system. This is any system whose failure could threaten human lives, the systemős environment or the existence of the organisation which operates the system. Failure in this context means any potentially threatening system behaviour other than failure to conform to a specification. Examples of typical critical systems are command and control systems such as air-traffic control and electricity supply systems, disaster management systems among others.

On the other hand, critical situations are not exclusive to critical systems. In this work context, critical situations arise when decisions must be made and actions must be taken under time pressures and cognitive resourcesô constrains; and the failure to do so might also threaten human lives, the systemős environment, or the existence of the organisation which operates the system.

Aiming to support the human error study by widening the range of information available beyond that obtained through report analysis, this chapter proposes the adoption of a method and protocol to support the observation and analysis of the human behaviour during the interaction in critical situations. It is proposed that an individual taking part in such experiment must be immersed into a work context that reproduces the conditions described in accident and incident reports. That is, an individual with a similar profile to the original human operator must be lead into performing the same task under similar conditions, whilst being observed according to the proposed protocol. This approach in observing the human behaviour when performing tasks in a critical situation is synthesized in the experimental protocol introduced later in this chapter. This protocol systematizes the observation planning, execution and documentation of the observation experiment.

From the usability point of view the investigation inherits the techniques employed in observing the interaction between systems and users when performing a predefined task under controlled conditions, i.e., during a usability test. During these tests, data is gathered to produce objective and subjective metrics, as part of a diagnostic about the system in use. In this research, cognitive psychology, supports the investigation of the relationship between the functional and emotional state of an operator in charge of a system, his (her) workload, and how the combination of these two influences: the error occurrence; the task outcome and the operatorǿs performance. This investigation relies on the adequate choice of tools, available to work-psychology studies, in order to support data gathering and analysis, relevant to the understanding of the human behaviour.

2. Understanding the Human Behaviour

To understand the human behaviour, that is, the reactions of an individual during a critical situation, it is necessary to analyze this situationǿs characteristics and the individualǿs functional state in order to identify the relevant human behaviour and its components. These reactions happen in several observable levels such as: performance, emotional, and physiological. The tools employed in detecting and measuring these reactions can be organized similarly into one of three categories, according to the targeted information: subjective, performance and psycho-physiological measuring tools. This section presents the tools employed during the research.

2.1 Subjective Measures

Subjective measures enable estimating parameters which are related to an individualǿs feelings and reactions induced by a situation (or context), such as: emotions, functional state, and workload perception.

Subjective Workload Measures There are two most commonly used techniques for measuring the subjective mental workload. The first one, NASA-Task Load Index (NASA-TLX; Hart & Staveland 1988), classes the workload into six subscales: Mental Demand, Physical Demand, Temporal Demand, Own Performance, Effort, and Frustration Levels. The other is the subjective workload assessment technique (SWAT; Reid & Nygren 1988) which describes the operator workload in three dimensions: Time

Load, Mental Effort Load and Psychological Stress Load. Some of the dimensions have been considered in both techniques (Miyake 2001): the Time Load and Temporal Demand dimensions; the Mental Effort Load and the Mental Demand and Effort dimensions; and the Psychological Stress Load and Frustration dimensions. Both techniques are largely used in the field of aeronautics, where Collet, Averty, and Dittmar (2009) found a positive correlation between the number of aircrafts to control and the NASA-TLX score, amidst traffic controllers, indicating a high sensitivity of NASA-TLX to small workload changes.

Subjective Functional State Measures. According to Thayer (1987), task performance is sensitive to the individual functional state, highlighting that arousal state underlies the behaviour. Therefore, the model of multidimensional activation (Thayer 1985; 1986) is composed of two dimensions, the energetic arousal and the tense arousal. The tense arousal is considered to be determined by danger, and to be largely cognitively mediated. Whereas the energetic arousal varies naturally according to circadian rhythm, and as a function of factors such as: time of the day, exercise, nutrition, and mental workload (Thayer 1987; Cariou, Galy & Mélan 2008). Each dimension level is relative to the others, because these are assumed to form a curvilinear relationship. That is, these are positively correlated at low levels, and negatively correlated at high levels (Mélan, Galy & Cariou 2007). Therefore a phenomenon considered dangerous has a greater psychological impact (leading to a greater tense arousal) when the energetic level is low; and a lesser impact when the energetic arousal is high. Consequently, considering these dimensions (tense arousal and energetic arousal) during a critical situation seems essential. The Short Form Activation-Deactivation Adjective Check List (AD-ACL) can be used to assess energetic and tense arousal (Thayer 1986). This is composed of four sub-scales: energy, tiredness, tension, and calmness. The ratio between the energy and tiredness scores is employed to assess energetic arousal and the ratio between the tension and calmness scores, to assess tense arousal.

Subjective Emotion Measures Scherer (2001) proposed a Component Model of Emotion (CME), by considering emotion like an episode of interrelated and synchronized state changes, resulting from the situation evaluation. The model components are: cognitive appraisal, physiological reactions, behaviour tendencies, motor expression, and subjective feeling (emotional experience). Two of this modelôs components: cognitive appraisal and emotional experience can be estimated by subjective measures.

Cognitive Appraisal consists of a situation evaluation under direct emotional responses (positive or negative). Demir, Desmet and Hekkert (2009) consider the following appraisal components: consistency of motives, intrinsic pleasure, expectation confirmation, standard conformance, agency, coping potential, and certainty. Scherer (2001) proposed the Geneva Appraisal Questionnaire (GAQ) to assess the result of an individual's appraisal process during a specific emotional episode. This tool allows estimating five dimensions: intrinsic pleasantness, novelty, goal/need conduciveness, coping potential and norm/self-compatibility. The emotions taken into account by GAQ are: anxiety, irritation, contentment, joy, sadness, disgust, fear, anger, and surprise.

Emotional Experience is characterised by emotion type and intensity. Two tools can be used to measure it: the Geneva Emotion Wheel (GEW; Scherer 2005) and EMOTAIX (Piolat & Bannour 2009). GEW is a verbal self-reporting instrument on which an individual is asked to indicate the emotional intensity felt in a particular situation, amid 20 emotion categories, such as: interest, irritation, contentment, joy, sadness, disgust, fear, anger, and surprise. Five degrees of intensity are available for each emotion category, and these are represented by circles of increasing sizes, as a function of the emotional intensity. EMOTAIX, in turn, allows analysing the emotional lexicon, used by an individual, when reporting the feelings experienced in a situation. This is a computer based application that works with corpus analysis **-** Tropes software (version 7). This tool enables to account for the emotional lexicon (2.014 references) according to hedonic dimensions (positive or negative valence) and according to 28 basic thematic categories, grouped in super and supra categories.

2.2 Performance Measures

Performance is the result of an established behaviour, more or less adapted to a situation, and corresponds to response accuracy and response latency. Chi and Lin (1997) demonstrated a trade-off between these performance criteria. The time needed to complete a task increases when accuracy requirements increase, whereas a decrease in accuracy occurs when task speed requirements increase. In the same line, Fournier, Wilson, and Swain (1999) proposed a method to arrive at a performance index, taking into account both response accuracy and response latency. They evaluated subjects' behavioural responses in a multi-task context by calculating a composite standardized Z-score for each subject. For each task, the ratio between the reaction time and the proportion of correct responses was weighted by one-quarter and, the correct ratios were summed up. Results revealed that the global performance decreased as the task demand increased and, that the performance improved with training, especially in high task demand conditions.

2.3 Psycho-physiological Measures

The managing of a critical situation causes changes to the autonomic system which consists of two systems: one controlled directly by the nervous system and another controlled by the hormonal system (adreno-medullary). These two systems have different functional roles (Folkow 2000). Whereas the former executes precise, rapid, and often highly differentiated adjustments, the latter independently modifies important metabolic functions. The two systems may mutually support each other, when massive and generalized system activation occurs, as it is the case during a critical situation. Measures commonly employed to evaluate autonomic system demand are respiratory, cardiovascular and electro-dermal measures. One major advantage of using physiological measures is the continuous availability of bodily data, allowing the reactions to be measured at a high rate and with a high degree of sensitivity, even in situations when open behaviour is relatively rare (Paas 1992). Physiological measures are also very sensitive to physical effort and, will reflect specific mental load or emotional variations for activities involving little or no physical effort (Brünken, Plass & Leutner 2003). The measurement of cardiac activity is a physiological technique employed in the assessment of mental workload and emotional aspects (i.e. anxiety). It has been demonstrated that the heart rate variability (HRV) (Backs 1995) shows a systematic and reliable relationship with task demand (Mulder & Mulder 1981; Tattersall & Hockey 1995). Similarly, electro-dermal measures are being used to estimate emotional feeling such as anxiety, contentment, joy and fear (Kreibig 2010), and it has been found that these are very sensitive even to low emotional changes.

The tools presented in this section were conceived to estimate an individualôs reactions in several levels. However, these levels are not parallel, but nested within each other. A study by (Galy, Cariou & Mélan 2012) highlighted the relationship between the indicators to be adopted in this study: 1) subjective measures (self-rated effort; energetic arousal), 2) performance measures (correct responses), and 3) physiological measures (heart rate variability). The model presented on this chapter was based on this study, which suggests that the measures displayed differential sensitivity to the three contextual factors investigated, and that some measures can be determinants of others. Specifically, the heart rate variability increased with high energetic arousal, whereas the participantsôself-rated the mental effort as sensitive to both: task difficulty and time pressure. Performance was determined both by energetic arousal and by the interaction between task difficulty and time pressure. Thus, the only satisfactory solution to apprehend the human behaviour in the critical situations is to diversify metrics hoping to achieve a more comprehensive view of the potential repercussions of the contextual factors on an individual, at all levels.

3 A Simplified Abstract Model of the Human Behaviour

During this research an abstract model was conceived to represent a set of variables and their relationship, aiming to support the understanding of the human behaviour during critical situations. The model includes six major components, each composed of elements of interest to investigate. These are represented in Figure 1: (i) an individual functional state, based on his or her state of alert state and anxiety level; (ii) an individual characteristics (profile, personality and attention); (iii) situation characteristics (stressing elements, tool's usability and task to be performed - test scenario); (iv) the situation awareness through workload and emotional state, both as perceived and expressed by the individual; (iv) decision making (situation cognitive evaluation) and (v) the achieved human performance is expressed in terms of: task duration; number of errors and final task state.

The relationship between components is defined by the influence of one component over another. For instance, the situation awareness is influenced by: the individual and situation characteristics and the individualôs functional state. Within situation awareness, the workload and the emotional state influence each other. The performance is influenced by the decision making, and both depend on the situation awareness.

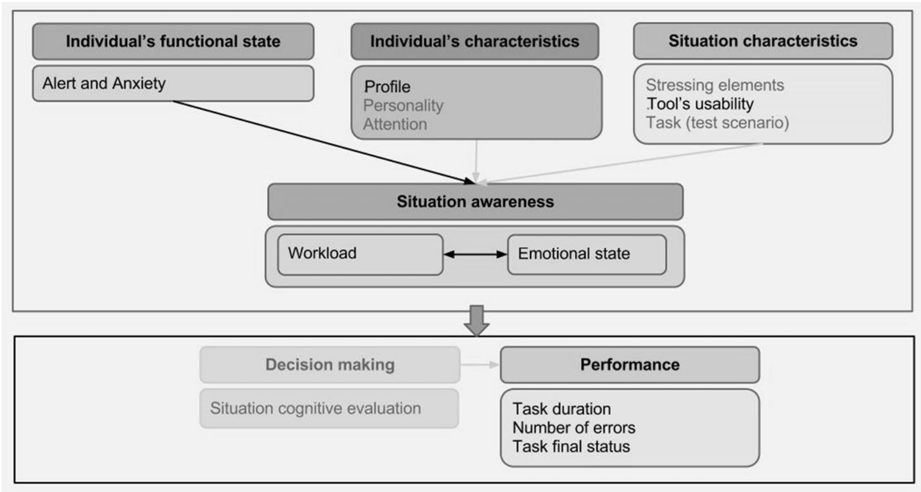


Fig 1: The abstract simplified model of the human behaviour

It must be pointed out that other variables, also accounted for in the simplified model are equally important for the understanding of the human behaviour, in spite of not being discussed on this chapter. These are: the individualǿs characteristics such as profile (age, gender, and schooling, knowledge on the system, the task and the work context); personality trait, attention and concentration levels and, situation characteristics such as usability of tools and systems used to perform the task, stressing events and decision making style, based on the cognitive evaluation of the situation. It must be highlighted that the scenario devised for the experiment concerns a specific task and a set of stressing factors to be presented to the individual taking part in the experiment along its course. Therefore, in spite of not being discussed in this chapter, the case study chosen scenario accounts for these variables and their relation with the user behaviour.

Cognitive resources are essential to an individual decision making process, and in this context these determine task performance. Their availability is determined by the individual functional state, represented in the model as alertness and anxiety levels. Cognitive resources consumed before task execution are spent with the individual perception of the situation in terms of the workload; and with the individualǿs emotional state. Situation awareness is influenced by the individual's and situation characteristics. The remaining cognitive resources can be obtained by subtracting the consumed resources from the available resource. The result is left for decision making which in turn influences an individualǿs performance.

In the model, human performance is characterized by the variables: duration of the task, task completion state and error rate, all measured during task performance. These variables are influenced by the user perception about the experiment itself, characterized by the inter-relationship between workload and emotional state. Whereas situational awareness is the result of the following individual characteristics: profile, functional state of alertness and anxiety, as well as situation characteristics coupled with the usability of the tools employed to perform the task.

The following section introduces the experimental protocol conceived to support the understanding of the human behaviour based on the above model. It guides the process of experiment planning and data gathering during the observation.

3.1 The Experimental Protocol - PEOI

In the human interface domain, the usability evaluation practice is based on the observation of individuals during task performance. This practice allows identifying the elements of the interface which prevent goals achievement as well as the potential solutions to the perceived problems. Since it also consists in observing human behaviour, the data gathered might be biased by the awareness of the observed individual. Therefore it is paramount to follow a protocol in all of the experimentôs phases: planning, conducting, analyzing and reporting the results, in order to ensure that consistent data will be available. Such detailed description of procedures and activities supported by specific documents compose an experimental protocol (Mayhew 1999).

This section briefly describes the experimental protocol adapted to support the observation and gathering of human-behaviour data from the interaction, during critical situations. The Experimental Protocol to Observe the Interaction (PEOI) reflects the model conceived to support the understanding of the human behaviour, presented above. Its application is supported by tools employed in the observation of the human behaviour in a controlled environment during the simulation of critical situations.

PEOIôs structure is based on the principles adopted during the observation of the human behaviour in the areas of product usability testing and work ergonomics in cognitive psychology. It originates from the practices adopted in the Human Interface Laboratory - LIHM, at the Federal University of Campina Grande (UFCG), in Brazil, and resulted from the compilation of twenty seven years of product usability testing experiences at LIHM, refined with contributions from the literature review on usability evaluation practices, such as those described in the work of Nielsen 1994; Mayhew 1999; Redish 2007. It is structured in phases, goals and artefacts. Its original version describes how to plan, execute and report on the observation of the interaction between users and products (Aguiar et al. 2011).

In order to adapt it for the observation of the user behaviour, the original version was reviewed to: (1) establish the relationship between phases, processes and activities; (2) define the roles of the actors involved in the experiment; (3) propose tools for data gathering and analysis, which can be adequate to handle the variables of interest for the human behaviour model.

The protocolôs process flow is organized in six phases, each one with a clear objective, and consisting of processes detailed into activities:

Phase 1: Experiment Planning, which is crucial for the correct observation.

Phase 2: Participant's training. This phase is optional, depending on the observation goals.

Phase 3: Experiment Elaboration and Validation. During this phase, the team must organize all the materials to be used during the experiment.

Phase 4: Driving the test and Data gathering.

Phase 5: Data preparation and Analysis. This phase consists of organizing the data gathered in the previous phase: questionnaires, interviews, audio and video recordings, as well as physiological data recordings. This is followed by the analysis of the situation characteristics and participantsô functional state, decision making and performance.

Phase 6: Presenting the results as a diagnostic containing the findings on participantsô behaviour.

An ensemble of roles and responsibilities are defined for all experiment participants. A role represents a set of responsibilities assigned to a participant. The roles were grouped into classes and represented in the protocol as a workflow.

To support the protocol application during all phases of an experiment, a set of artefacts and respective templates is available to the experimentôs team. The artefacts are grouped according to the phases: planning, application, data analysis and reporting on results. Some artefacts were conceived to ensure the ethics of the experimentôs procedures, including terms of agreement with the experiment conditions whereas others are terms of confidentiality over the tools and task procedures employed during the experiment and which belong to the stakeholders.

3.2 PEOI: Data gathering methods

Data gathering in the protocol consists of a combination of methods: interaction observation; interviews; questionnaires; physiological measurements; document analysis and video and discourse analysis. The choice of methods to be combined is a function of the observation goals and the phase of the protocol. The tools proposed for data gathering are illustrated in (Figure 2). These are grouped into four categories: individual's functional state; individual's characteristics; situation characteristics; situation awareness; decision making and performance.

The toolsô categories reflect the components in the Human Behaviour Model. Given the focus of this chapter, it follows a brief description of only the subset of tools applied in the case study described in section 4.

Individual's characteristics: individualôs profile data is obtained through the application of the questionnaire: User Objective and Cognitive Profile (POCUS) (Scherer & Vieira 2008). It gathers information on: personal, physical, professional, contextual, psychological and clinical aspects of an individual.

Situation characteristics: product usability level is measured by applying the questionnaire Webquest (Vieira et al. 2005). The usability level is based on an individualôs subjective satisfaction with a product (tool) employed to do a task during the observation experiment. This questionnaire investigates the following aspects: navigation and product use, documentation (and help mechanism), product feedback in response to its user actions and product-user interaction. The result is a satisfaction index level in the scale: extremely satisfied (0,67 to 1), very satisfied (0,33 to 0,66), marginally satisfied (0,01 to 0,32), neutral (0), marginally dissatisfied (-0,01 to -0,32), very dissatisfied (-0,33 to -0,66) and extremely dissatisfied (-067 to -1).

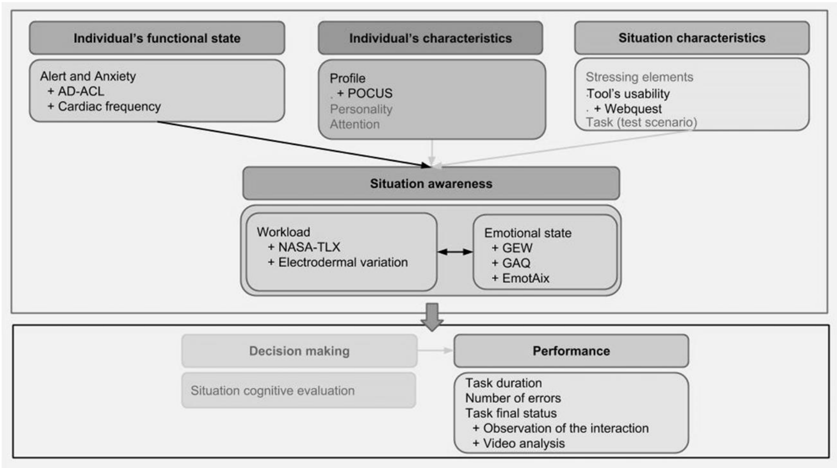


Fig 2: Data gathering tools

Functional state: to investigate the anxiety and the alert levels in an individual, two tools are employed: the questionnaire Activation-Deactivation Adjective Check List (AD-ACL), which allows evaluating the transitional state of alertness. To this transitional state it is assigned one of the levels: general activation, sleepiness, high activation and general deactivation. With this knowledge it is possible to determine: the actual state of alertness (arousal energetic); and the state of tension (tense arousal) which combined determine the levels of anxiety and alertness. The second tool, measures the heart rate variance, to evaluate the anxiety level. In this case, the anxiety level is extracted from the interpretation of the spectrum, in the fourth band of frequency: very low (0.01 to 0.04 Hz); low (0.04 to 0.15 Hz), high (0.15 to 0.40 Hz) and Ultra High (10-5 a10-2 Hz). The relation: low frequency/high frequency, allows evaluating the sympathovagal balance (Berntson, Quigley & Lozano 2007). When the balance is found, the anxiety is discarded; this happen when the results present a variation of +/- 2.33. According to Sztajzel (2004) this variation must be approximately 4.61, for healthy individuals.

Situation awareness: Two dimensions are assigned to an individual perception of a situation: the workload perception and the emotional state. To measure the workload, PEOI proposes to apply two tools: the questionnaire NASA-TLX, and the electrodermal variance (physiological measurement) during task performance. The NASA-TLX questionnaire identifies the individual perception on his/her workload. It covers different aspects of the workload, such as: behaviour (effort and performance); task demand (mental, physical and temporal); and subjective (frustration). If the sum of all aspects is higher or equal to 60 points, the individual is characterized as being within the acceptable workload level. In order to evaluate the emotional state, PEOI proposes the tools: GAQ and GEW questionnaires and a discourse analysis tool - EmotAix. GAQ investigates the relevance, implications, potential for coping and the compatibility of a situation with norms and standards; considering it all essential to know the user cognitive evaluation ability. This questionnaire presents the respondent with a list of emotions to select the one that best represents his/her emotional state during the situation. The valance, intensity and duration of the chosen emotion must also be informed. GEW presents the respondent with 20 emotions organized in pairs, disposed at a quadrant of a Cartesian plan. The dimensions are: High control / power appraisal versus low control/power appraisal and, Unpleasantness / Obstructiveness Appraisal versus Pleasantness / Conduciveness Appraisal. Each dimension is composed of five emotions, and every emotion is associated to five circles of different sizes associated to its intensity. The respondent must choose one (or maybe two) predominant emotion and the corresponding intensity of his/her emotional state. The tool EmotAix is employed for analyzing the individualôs discourse during an interview. This tool allows the identification of positive, negative and neutral emotions, as well as surprise; based on a vocabulary (dictionary) of emotions, feelings, mood, personality and temperament.

Performance: individual performance is measured using objective indicators during task performance observation. These indicators are: time spent on doing the task; obtained by comparing the estimated time versus the time actually taken; the number of human errors incurred when performing the task; and the task state at completion, that is task completed with or without fault or task unfinished, abandoned or interrupted. These data are gathered through direct observation of the variables of interest. The observation can be enhanced by video and audio recordings.

Besides the described tools for data gathering, the protocol also provides guidance on how to analyze the collected data, as will be shown next.

3.3 PEOI: Data Analysis

In the protocol, data analysis is performed in two steps. In the first one must select how to analyse each set of data, according to the tool used for gathering. In the second step a data analysis correlation is performed. The objective of the latter is to correlate data across the categories of the proposed Human Behaviour Model. During this second phase the variables are separated into two groups, internal and external variables, according to their pattern of cognitive resources consumption (Galy et al. 2012).

4 Case Study

This section presents a case study to illustrate the Human Behaviour Model application, and the supporting protocol (PEOI). Its main objective was to allow investigating the relevance of the collected data and analysis. The situation consisted in observing and analyzing the user behaviour when interacting with a product during a critical situation ï a crisis management, in a simulated environment. During the experiment the participants were asked to generate an intervention plan after a maritime accident, within a strict time constraint. The plan generation was supported by the decision making aid tool - *Generateur de Plans d'Intervention* (GENEPI) (Mercantini et al. 2010). The experiment consisted in simulating the situation, inspired in a report of a real crisis. The protocol PEOI was employed to observe the user interaction with the system GENEPI. The aim was to understand the relationship between the human error and the participantôs behaviour. The data was gathered though observation, video recording, questionnaires, interviews and with the aid of specific tools to gather physiological data.

In this case study only a subset of the modelôs components was considered, which were chosen to investigate the participantôs: profile; functional state; situation awareness; and the task tool usability and, objective indicators linked to the individualôs performance. Therefore only a subset of the variables present in the Human Behaviour Model was explored. The variables of interest were: (1) the experiment participant profile, gathered with POCUS; (2) the participant functional state, considering the levels of alertness and anxiety as measured with AD-ACL and the heart rate variance; (3) the situation awareness based on the workload (measured with NASA-TLX and variance electrodermal); (4) the participant emotional state (measured with GEW, GAQ and EmotAix); (5) the subjective participant satisfaction with the working tool assessed through product usability and situation characteristics, assessed with Webquest; and (6) the performance variables during the experiment, obtained through direct observation and video analysis. Details of the experiment will be given in the following subsections.

4.1 The Experiment

Place, Time and Sessions: The experiment was performed in 2011, at the Research Centre for Knowledge Psychology, Language and Emotion[1], throughout seven sessions including a pilot one.

Test participants basic profile: the group of seven participants consisted of three men: P1, P5, P6 and four women: P2, P3, P4 and P7. These spanned a large age group: P2, P3 and P6 aged between 18 and 24; P1and P7 aged between 25 and 35, and P4 and P5 aged above 35. The participantsô academic background was: P4 and P5 PhDs; P2, P3 and P6 undergraduate students; P1 a Master student and; P7 a doctorate student. The participantsôacademic domain was risk analysis, computer studies and human interface studies. Their level of expertise on the domains: product (GENEPI) use; task to be performed during the experiment (generating contingency plans) and; on the working context (risk management), is displayed on Table 1.

The task was organized into two scenarios, which differed on the level of details given to the participants, and on the time allocated for task completion. The guided scenario was given to participants with lower levels of expertise, to be completed within 30 minutes (estimated time). The open scenario, with less guidance to the participants, had an estimated completion time of 40 minutes. The difference in completion time aimed to balance the cognitive demands on both groups of participants.

Table 1: Knowledge and Expertise on: Product use**,** Task and, Working Context

		Participant						
		P1	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**
	Product	Low	Low	Low	Low	High	High	Medium

[1] *Le Centre de Recherche en Psychologie de la Connaissance, du Langage et de l'Émotion (Centre PsyCLÉ*

Knowledge and Expertise	Low	Low	Low	Low	High	High	Medium
Task Context	Medium	Low	Medium	Low	High	Low	Low
Test Scenario	Open	Guided	Guided	Guided	Open	Open	Guided

Product: *GENEPI* is a decision making support system conceived to assist crisis management, which supports the elaboration of contingency plans for maritime accidents. It is one module of a tool developed for crisis management related to marine pollution. Its objective is to facilitate and accelerate the establishment of exclusion zones, and to mobilize the appropriate means to handle these critical situations.

Participants training: The participants had an introductory presentation on the use of GENEPIôs functionalities and user interface given by the development team, which lasted 1 hour and 30 minutes.

Task and interaction Context: The experiment emulated a crisis scenario which was inspired on maritime accident reports from real situations. During this scenario the participant was asked to use the tool GENEPI to generate a contingency plan for the described maritime accident, also known as an intervention plan. The work context in which the participants were immersed was prepared in order to emulate that at the *Centre Opérationnel de Surveillance du Littoral* (in the French Mediterranean region), which is the first organism contacted after an accident happens. The work scenario begun with the participant receiving a simulated phone call from the *Préfecture Maritime de la Méditerranée*, informing about the accident. The severity of the simulated accident required that the intervention plan should have been generated within time restrictions imposed by another simulated phone call from the *Préfecture*. The participant was advised in the task description text to get additional information on the accident situation by contacting specialized services through phone calls using a given phone directory. The phone contacts were simulated with the aid of supporting participants who performed predefined roles.

Experiment team: consisted of a group of six people with multidisciplinary skills who performed multiples roles, amidst interacting with the participant during the experiment. Two of the team members were product usability experts; one was an expert in cognitive psychology and, two undergraduate students in computer science.

4.2 Data Gathering and Analysis results

The data was gathered **in accordance with the** experimental protocol PEOI. It was collected during the experiment was classed and analysed for each model variable. In order to evaluate the behaviour model, four hypotheses were investigated in this study:

H1: Individuals who were more dissatisfied with product usability (Webquest) tend to express more negative emotions (EmotAix);

H2: Individuals tend to display higher alert levels (AD-ACL) when they perceive a situation of work overload (Electrodermal peaks);

H3: Individuals tend to display higher anxiety when they perceive a situation of work overload (Electrodermal peaks);

H4: Individuals tend to have a higher error rate when they are under negative emotions (EmotAix).

The analysis was performed in two steps. The first step consisted in visually investigating a correlation between the model variables, which was performed by the domain specialists in usability and cognitive psychology. This was followed by a exploratory data analysis performed with the same purpose.

The exploratory data analysis, initially investigated the normality of the data distribution using the Shapiro-Wilk test (Shapiro and Wilk 1965). A significance level of 5% (0.05) was adopted, with a level of confidence of 95% that variables with the p-value higher than 0.05 follow a normal distribution. This was followed by a correlation investigation between sets of variables, using the Spearman's rank correlation coefficient (Zar 1972). Dispersion graphs were then built in order to propose an appropriate regression model for the relationship between predicting variable and response variables (linear or non-linear), for each hypothesis. As a result it was determined the influence of the predictor variable over the response one. It follows the analysis results.

Situation characteristics: GENEPI usability was indirectly measured through a user satisfaction Index. Table 2 presents this index calculated for all the experiment participants, according to their individual subjective satisfaction level. The opinion of the majority of participants (5/7) resulted in a negative index, which represents a moderate dissatisfaction (P7, P2, P3, P4 and P1). The other 2 participants (P5 and P6) displayed a high satisfaction level. This contrast may have resulted from the difference in participantsô familiarity levels with GENEPI; with the highly satisfied users being the most familiar ones.

Table 2: User Satisfaction levels with GENEPI

	Participant						
Satisfaction level	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**
Index	-0,068	-0,175	-0,171	-0,126	0,488	0,412	-0,240
Moderate dissatisfaction	✓	✓	✓	✓			✓
High satisfaction					✓	✓	

Individual's characteristics: these were self-assessed by the participants who answered a questionnaire. According to their answers, three participants: P5, P6, P7 considered their abilities for learning and knowledge application as high, three participants: P1, P3 and P4 self rated as average, and one participant: P2 did not answer this question. Regarding computer literacy, except for one participant: P4 (declared as average), all the others declared as high. Regarding a set of specific psychological characteristics, it follows the distribution. Ability to analyze situations and to solve problems: P1, P2, P3, and P7 self assessed as average, whereas participants P4, P5, P6 self-assessed as high. Regarding the sense of direction, P2, P3, P4 and P7 declared as average, whereas P1, P5 and P6 self-assessed as high. The distribution of self-assessment regarding the level of abstraction was: P4, P5, P6 and P7 declared as high, whereas P1, P2 and P3 declared as average. Concerning organization and planning: P1, P3, P4 and P7 declared as high, whereas P2, P5 and P6 declared as average. Lastly, regarding the ability to have a wide panoramic view of a particular situation, participant P5 did not answer this question; participants P1, P2, P3and P4 self assigned an average level, whereas participants P6 and P7 declared a high level.

Functional state (alertness and anxiety levels): the questionnaire AD-ACL was employed to determine this state, with its four dimensions: activation general, inactivity / sleepiness, high activation and inactivity general. For each dimension, the maximum score is 20 points. All participants scored high, showing that they were alert and aroused, according to the measured *activation general*. Consistently, the *inactivity and sleepiness* levels were low for all users. Given the *high activation*, the users were considered not tense during the experiment. According to the *inactivity general*, the participants were in a medium level of tranquillity. This data was converted into levels of alert and anxiety for each participant, with the maximum score being 4 points, as shown in Figure 3. When the obtained index (shown in the abscissa of Figure 3), is above 2, then anxiety and alertness are present. Two participants were anxious (P4 and P6) and five participants (P5, P1, P3, P6 and P2) were alert.

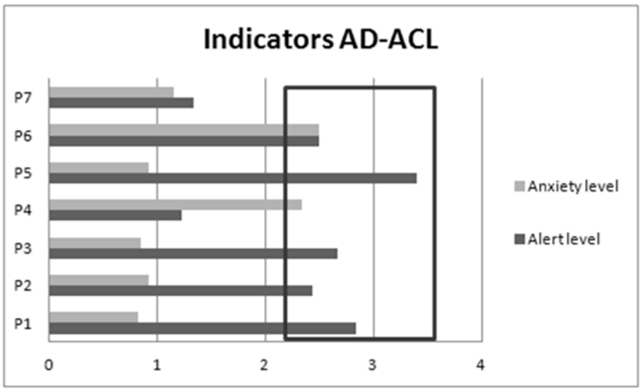


Fig 3: Anxiety and alert levels

From the analysis of the heart rate variation, with reference values in the range of 2.28 and 6.94, participants with an index greater than 6.94 were classed as having a significant anxiety level. Therefore, as shown in Figure 4, anxiety was detected using physiological measures in four participants: (P1, P3, P4 and P7), with participant P4 displaying the highest value.

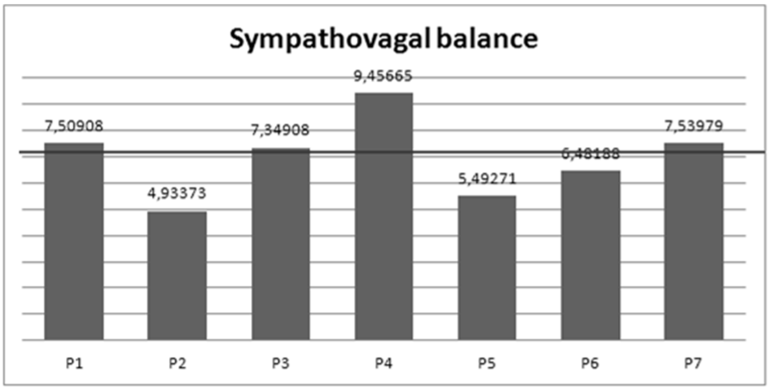


Fig 4: Participantsô Sympathovagal balance (anxiety)

Situation awareness (workload and emotional state): In the experiment, the perceived workload was measured by the questionnaire NASA-TLX. The results show that three participants scored above the line which divides the non-excessive and excessive workload perceptions. Thus three participants (P2, P4, P6) considered it excessive; whereas four participants did not consider it excessive (P1, P3, P5 and P7 ï see Figure 5).

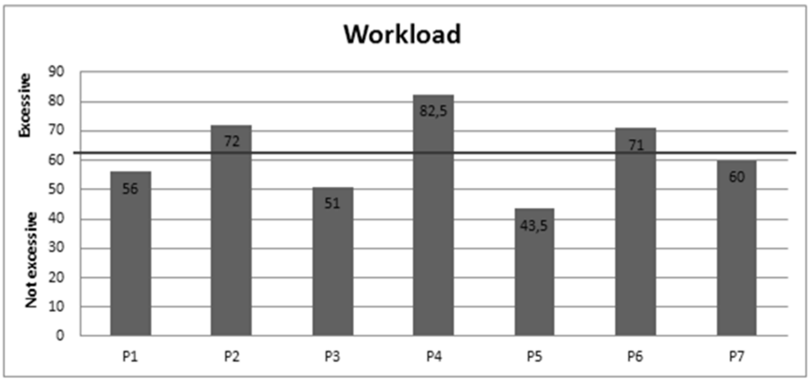


Fig 5: Workload perception

The NASA-TLX questionnaire also allows identifying the dimensions that most influenced the workload perception. The results showed, in decreasing order of relevance, the following dimensions: mental demand, frustration, cognitive effort, temporal demand, performance and physical demand. Therefore, the task represented a high cognitive workload, combined with a considerably high level of effort, leading to participant frustration.

In the experiment, the workload was also measured by electrodermal activity (EDA). The strategy adopted was to identify the number of peaks (independent of amplitude) and associate those to the situation context (events, stress factors, section of task under execution, etc). The aim was to interpret the electrodermal variation in the context of its occurrence. The number of peaks suggests a workload level to be investigated. The greater the peak occurrence the higher the workload to which the participant is subjected. The higher number of peaks was detected for participants P5 (65), P2 (62 peaks) and P1 (41 peaks), whereas lower numbers of peaks were detected for participants P7 (23 peaks), P3 (25 peaks), P4 (29 peaks) and P6 (38 peaks).

For the emotional state, measured by GAQ, it is important to note that the situation was described by the participants as: highly relevant (P1, P2, P3, P4, P5 and P6); with important implications (P2, P3 and P7); low in compatibility with norms and standards (P6); and participants (P1, P3, P4, P5, P6 and P7) declared to have low potential for coping with the situation. Therefore, one can state that the situation (interaction with GENEPI) was considered by the participants as an extreme case. Whereas the valence associated to the situation was perceived as negative by five participants (P2, P4, P5, P6 and P7), reaching the maximum intensity for two of those (P4 and P5). The emotions chosen by participants to describe their emotional state were: anxiety, irritation and sadness. The last finding with GAQ was the participantsô attempt to minimize or mask their feelings during the experience. Regardless of positive or negative valence, all participants tried to reduce the intensity and duration of their emotional episodes and tried to control or mask their feelings.

With GEW, it was possible identifying the valence and the intensity of emotions. The negative emotion was displayed by five participants (P2, P3, P5, P6 and P7), two of which with the maximum intensity (P6 and P7). Participant (P4), amidst the two who displayed positive emotions, reached the maximum intensity. It must be pointed out that the two participants with positive emotions were part of GENEPI development team, therefore they knew it well and were positively biased in their evaluation.

EmotAix was employed in the analysis of discourse recorded during the interviews after the experiment. The results showed that for all participants prevailed negative emotions (Figure 6).

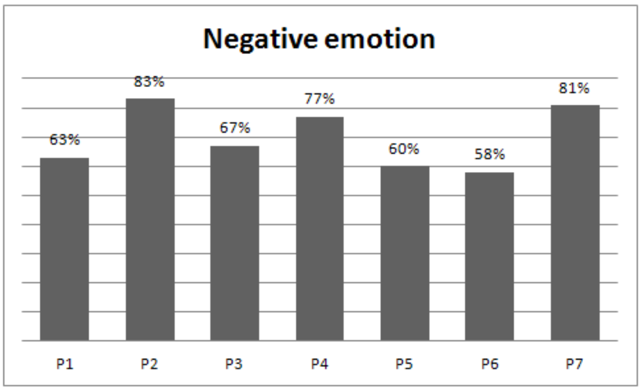


Fig 6: Participantsô negative emotion detected by EmotAix

Performance: In the experiment, performance was measured in terms of a set of dependant variables, one of which was the relationship between the estimated time to perform the task and the actual time taken to perform it. The estimated time for guided and open execution scenarios were respectively 30 and 40 minutes. This relationship is represented in Table 3. Given that only one participant (P1) did not exceed the estimated time, it was concluded that the estimation was not adequate.

Table 3: Estimated versus execution time

	Participant						
	P1	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**
Estimated time	40ô	30ô	30ô	30ô	40ô	40ô	30ô
Execution time	**29ô32ôô**	41ô24ôô	32ô58ôô	33ô01ôô	43ô34ôô	51ô08ôô	**34ô18ôô**
Over time	-	**11ô24ôô**	2ô58ôô	3ô01ôô	3ô34ôô	**11ô08ôô**	4ô18ôô

Another indicator of performance in the model is the final state achieved through performing the task, that is: *task finished with or without faults*; *unfinished due to interruption* or *abandonment*. Six users did not finish the task (P1, P2, P3, P4, P5 e P7). Half of those interrupted (P2, P5, P7), and half abandoned (P1, P3, P4). One of the reasons for abandoning the task was reaching the estimated time limit; which lead the team conducting the experiment to interrupt the task. Another reason was due to problems with the version of GENEPI, making it difficult or impossible to finish the task; and the unavailability of the necessary data to complete the task. Only participant P6 finished the task within the time limit, but this was achieved with faults.

In this study the human errors were classed in two groups. The first group was related to the interactive process, and consisted in not filling one or more fields in the GENEPI form, or filling incorrectly. The second group of errors consisted in failing to obtain the necessary information about the incident, in order to complete all GENEPIôs forms required to generate the intervention plan. This data acquisition consisted in phoning as series of organisms to obtain the required information. The authors consider that both kinds of errors were directly related to time pressure and distractions, some of which imposed by the experiment team in other to generate observable stress levels.

The failure to obtain the required information to fill the forms not only influenced the time taken to complete that task but also the task outcome and the participantôs emotional state. Given this correlation between the model variables the analyzed errors are those in the group 1, that is, errors in the interactive process. Thus, the global error rate is displayed in Table 4, with a highlight for the highest errors rates.

GENEPI consists of 68 fields organized in seven forms or tabs. To complete the task successfully, the user must fill in all fields correctly. Interaction between the user and this tool results in: (i) Omissions - fields left incomplete; (ii) Errors: fields populated with wrong data, and (iii) Success- fields correctly filled.

Given that, the contingency plan is completely generated (as opposed to partially generated) only when the GENEPI form is completed, with all 68 fields correctly filled, the error rate is expressed as a percentage of the expected 68 hits (correct data filled in the form). Table 4 ï shows the error rate in each category, as a percentage value. Participants P4 and P2 had the highest omission rates whereas P6 had the highest error rate. When the two error rates are considered, one obtains the global error rate, for which P2, P4 and P5 showed the worst performance. The table also rates all participants in terms of performance, according to their error rates, in decreasing order.

Table 4: Omission and error rates (%)

Error Classes	User						
	P1	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**
Omission	16.17	**44.1**	7.35	**72.03**	20.58	0	17.64
Error	5.88	8.82	4.41	5.88	8.82	**13.23**	8.82
Global error rate	22,05	52,92	11,76	77,91	29,4	13,23	26,46
Classification	5°	2°	7°	1°	3°	6°	4°

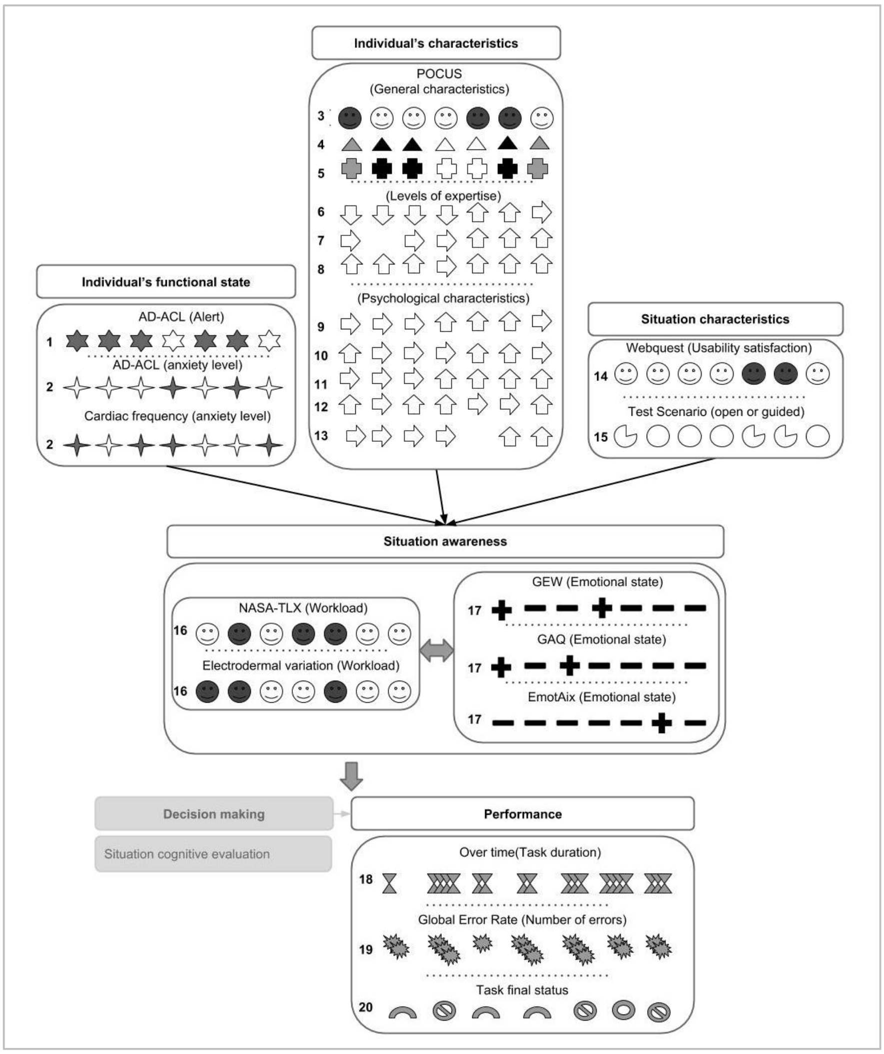
4.3 Results discussion

The previous section presented the results obtained with the protocol application when collecting the information required by the Model of the Human Behaviour in order to investigate the correlation between the model variables. The correlation could help explain the reasons for the human error occurrence, during the interaction with systems developed to support tasks under critical situations. Employing a variety of tools when gathering the same kind of data, aimed to investigate the best suited ones for the purposes of this research - that is to investigate the likely causes of human error during interaction.

A summary of the data collected in the experiment is shown in Figure 7. This Figure allows exploring the relationship between the variables investigated in the Model of the Human Behaviour, and adopts the same variables representation as the model. This global view of the data allows performing analyzes under different points of view. In the figure, the data collected is organized into categories defined in the model: individual's functional state; individual's characteristics; situation characteristics; situation awareness and performance; and grouped according to the variables of interest: alert; anxiety; participant profile; usability satisfaction; test scenario; workload perception; emotional state; task duration; global error rate and task outcome.

The variables and their respective value domains were represented in iconic form, as explained in the legend. The sequence adopted when representing the participants starts from left to right with P1 to P7. This representation allows both individual and group analysis of the investigated variables.

Figure 7 also makes reference to the instruments used to collect each variableǿs data. In some cases, more than one tool was used to collect the same variable. For instance, workload perception was captured using NASA - TLX and electrodermal variation. During this study it was intended to confront results obtained through participant opinion with stimuli involuntary responses. Differences in results can be explained through individual characteristics, such as personality traits, which influence the situation perception as positive or negative whereas physiological data is independent, since it is involuntary. Therefore, the authors recommend gathering independent data at the beginning of the experiment to use as reference values (base values) against which variable variations along the experiment can be identified.



1		Significant alert level
		Insignificant alert level
2		Significant anxiety level
		Insignificant anxiety level
3		Men
		Women
4		Aged between 18-24
		Aged between 25-35
		Aged above 35
5		Undergraduate students
		Postgraduate students
		PhDs

Low	*Medium* *High*
Levels of expertise	
6	Knowledge on: product, task and context
7	Ability for learning and knowledge
8	Computer literacy
Psychological characteristics	
9	Ability to analyze situations and to solve problems
10	Sense of direction
11	Level of abstraction
12	Level of organization and planning
13	Ability to have a wide panoramic view of a particular situation

14		High satisfaction
		Moderate dissatisfaction
15		Open Test Scenario
		Guided Test Scenario
16		Significant workload
		Insignificant workload
17		Positive emotion
		Negative emotion

Task duration		
18		On time
		Within time
		Marginally overtime
		Well above time
Number of errors		
19		Low error rate
		Average error rate
		High error rate
Task final status		
20		Completed incorrectly
		Interrupted due to time limit
		Abandoned

Fig 7: Experimentsô results mapped into the human behaviour model

It follows an example of analysis with the purpose of helping to understand how to interpret the information content of Figure 7. The analysis is simple and does not exhaust all possibilities of analysis of the data there represented. The analysis consists in exploring with the aid of the model, the potential causes behind the three cases of lowest performance (global error rate) individually presented by the participants: P2, P4 and P5.

P2 ï presented the highest global error rate in the group. The combination of low anxiety levels: declared though AD-ACL, and measured (cardiac frequency); with high declared alert level (AD-ACL), low level knowledge on the task, tool and context and moderate dissatisfaction with GENEPI usability, contributed to the high workload perception declared though (NASA-TLX), The electrodermal variation was compatible with felt and expressed negative emotions (GEW, GAQ e EmotAix). This combination of variables could explain the high error rate and the decision of interrupting the task (guided) soon after the estimated time of 30 minutes.

P4 ï presented the second highest global error rate in the group. The high values obtained with the declared anxiety (AD-ACL) and that measured physiologically (cardiac frequency) combined with the low alert level (AD-ACL), coupled with the low knowledge on the task and its context, and the moderate dissatisfaction with GENEPIõs usability, contributed to the high workload perception (NASA-TLX) leading into negative emotions (GAQ e EmotAix). This combination of variables might explain the high error rate and the decision to abandon the task (guided) immediately after reaching the estimated time of 30 minutes.

P5- with the third highest global error rate, this participant presented low anxiety levels, both declared (AD-ACL) and measured (cardiac frequency), combined with a declared high level of alertness (AD-ACL). Contrasting with participants P2 and P4, this participantõs knowledge on the task, the tool and the context was high, therefore he declared himself satisfied with the toolõs usability. His situation awareness, both declared (NASA-TLX) and measured (electrodermal variation) was a negative one ï high workload, resulting in expressed negative emotions (GEW, GAQ e EmotAix). This combination of variables resulted in a high error rate and the task outcome - interrupting the task (open) incomplete, after having exceeded the estimated time of 40 minutes. In spite of his high knowledge, high satisfaction with the tool and low level of anxiety, time pressures coupled with stressing factors introduced during the test resulted in a similar negative perception of the situation as the previously described participants.

Therefore, it could be inferred of the above analysis results, involving the three participants (P2, P4 and P5) that time pressures combined with stressing factors are very influential variables over the task outcome and the global error rate.

It must be explained that the stressing factors employed during the various sessions were different in each case and, these were chosen according to the previous knowledge of the participantsô profiles, aiming to cause a significant impact on the user attention and stress levels to be detected by the measuring instruments. An example of a stressing factor is constant interruptions by phone calls during the task. On the other hand, different stressing factors may also have caused different stress levels among the participants, interfering with the task outcome.

Although the analysis results can contribute to the understanding of aspects of the individual behaviour, it does not support the extraction of a standard human behaviour or supports inferences. This study demonstrated the complexity of trying to correlate the model variables, which remain a challenge to be further investigated. Based on calculations of the statistical power of the regression test and, aiming to define a sample size capable of reaching a high statistical power of 0,8 and an anticipated effect of ($f2 = 0,35$); it was found that experiments must be conducted with a minimum of 39 participants (system users).

5 Final considerations and future work

This work presented a methodological framework and supporting tools conceived to broaden the information available in the literature on the human behaviour during critical situations and its relation with the occurrence of the human error. Its major contribution is the Simplified Abstract Model of the Human Behaviour, supported by an experimental protocol to support the data gathering in a systematic and organized manner to support the observation of the human activity during critical situations. The simplified abstract Model of the Human Behaviour, accounts for six interrelated components: individual's characteristics, situation characteristics, functional state, situation awareness and performance.

The experiment reported on this chapter enabled to evaluate the quality and utility of the behavioural model in identifying behavioural factors which contribute to the human error during critical situations. It has also shown the viability of behavioural data gathering and analysis using the supporting protocol.

Since it was not possible to infer, a standard human behaviour which precedes the error, the authors recommend further studies, involving a larger sample of participants and broadening the investigations to include all the modelôs variables and their correlations. This approach should support inferences on which variables in the model are more influential on the user performance.

From this result it is also proposed to weight the influence of each model component on the user behaviour and to integrate this knowledge into the Method for Conceiving Ergonomic Interfaces (MCIE) (Vieira 2004) aiming to reduce the human error incidence due to human interface flaws. Finally, there are plans to build a computer system to analyze the results regarding the knowledge of the human behaviour during critical situations and issue recommendations to improve the interactive process and reduce the human error during critical situations.

References

1. AGUIAR, Y. P. C.; VIEIRA, M. F. Q; GALY, E.; MERCANTINI, J-M; SANTONI, Charles (2011). Refining a user behaviour model based on the observation of emotional states. In: Third International Conference on Advanced Cognitive Technologies and Applications - COGNITIVE, 2011, Rome. Proceedings of the Third International Conference on Advanced Cognitive Technologies and Applications, 2011.
2. AMALBERTI, R.. La conduite de systèmes à risques. Press Universitaires de France. Paris, 1996.

3. BACKS, R.W. (1995). Going beyond heart rate: modes of autonomic control in the cardiovascular assessment of mental workload. The International Journal of Aviation Psychology, 5, 25ï48.
4. BERNTSON, G.G., QUIGLEY, K.S., & LOZANO, D. (2007). Cardiovascular Psychophysiology. In J.T. Cacioppo, L.G.Tassinary, & G.G. Berntson, (Eds.). Handbook of Psychophysiology (3rd edition; pp. 182-210). Cambridge, UK: Cambridge University Press
5. BRÜNKEN, R., PLASS, J., & LEUTNER, D. (2003). Direct measurement of cognitive load in multimedia learning. Educational Psychologist, 38, 53ï61
6. CARIOU,M., GALY, E., & MÉLAN, C. (2008). Differential 24-h variations of alertness and subjective tension in process controllers: investigation of a relationship with body temperature and heart rate. Chronobiology International, 25 (4), 597ï607.
7. CHI, C. F., & LIN, F.T. (1997). A new method for describing search patterns and quantifying visual load using eye movement data. International Journal of Industrial Ergonomics,19 (3), 249ï257
8. COLLET, C., AVERTY, P., & DITTMAR, A. (2009). Autonomic nervous system and subjective ratings of strain in air-traffic control. Applied Ergonomics, 40, 23ï39
9. DEMIR, E., DESMET P.M.A., & HEKKERT, P. (2009). Appraisal Patterns of Emotions in Human-Product Interaction. International Journal of Design, 41-51
10. ENDSLEY, M. R.. Toward a theory of situation awareness in dynamic systems. Human Factors,1995, n.37, v.1, p.85-104.
11. FOLKOW, B. (2000). Perspectives on the integrative functions of the sympatho adrenomedullary system. Autonomic Neuroscience: Basic and Clinical, 83, 101ï115
12. FOURNIER, L.R., WILSON, G.F., & SWAIN, C.R. (1999). Electrophysiological, behavioural, and subjective indexes of workload when performing multiple tasks: manipulations of task difficulty and training. International Journal of Psychophysiology, 31, 129ï145
13. GALY, E., CARIOU, M., & MÉLAN, C. (2012). What is the relationship between mental workload factors and cognitive load types? International Journal of Psychophysiology, 83, 269-275.
14. HART, S.G., & STAVELAND, L.E. (1988). Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: P.A., Hancock, & N., Meshkati (Eds.), Human Mental Workload. Elsevier Science Publisher B.V.: Amsterdam, pp. 139ï183.
15. HOLLNAGEL, E. Human reliability analysis: Context and control. New Jersey: Academic Press, 1993.
16. JOHNSON, W. B.; ROUSE, W. B. (1982) Analysis and classification of human errors in troubleshooting live aircraft power plants. IEEE Transactions on Systems, Man, & Cybernetics, Vol. 12(3), May-Jun 1982, 389-393
17. KREIBIG, S. D. (2010). Autonomic nervous system activity in emotion: A review. Biological Psychology, 84, 394ï421.
18. MAYHEW, D. J. (1999) The usability engineering lifecycle: a practitioner's handbook for user interface design, San Francisco, CA: Morgan Kaufmann Publishers Inc.
19. MÉLAN, C., GALY, E., & CARIOU, M. (2007). Mnemonic processing in air traffic controllers: effects of task parameters and work organization. The International Journal of Aviation Psychology, 17 (4), 391ï409.
20. MERCANTINI J.-M., GUERRERO C.V.S, FREITAS D.D. de (2010), Designing a Software Tool to Plan Fight Action Against Marine Pollutions. The 7th International Mediterranean and Latin American Modelling Multiconference, October 13-15 2010, Fes, Morocco
21. MIYAKE, S. (2001). Multivariate workload evaluation combining physiological and subjective measures. International Journal of Psychophysiology, 40 (3), 233ï238
22. MULDER, G., & MULDER, L.J.M. (1981). Information processing and cardio-vascular control. Psychophysiology, 18, 392ï405.

23. NIELSEN, J. (1994) ñUsability Engineeringò. Morgan Kaufmann
24. NORMAN E. B., and BAUCOM D. Enhanced cognitive-behavioural therapy for couples: A contextual approach. American Psychological Association, 2002. NBR 6023
25. NORMAN, D.; DRAPER, S. User Cantered System Design in New Perspectives on Human-Computer Interaction. Hillsdale: Lawrence Erlbaum Associates, 1986
26. PAAS, F.G.W.C. (1992). Training strategies for attaining transfer of problem solving skill in statistics: a cognitive load approach. Journal of Educational Psychology, 84, 429ï 434.
27. PIOLAT, A., & BANNOUR, R. (2009). EMOTAIX: Un Scénario de Tropes pour l'identification automatisée du lexique émotionnel et affectif. L'Année Psychologique, 109, 657-700
28. RASMUSSEN, J. Skills, Rules, and Knowledge; Signals, Signs, and Symbols, and other distinctions in human performance models. IEEE: Transactions on Systems, Man, and Cybernetics, v. 13, n. 3, p. 257-266, 1983.
29. RASMUSSEN, J., PEDERSEN, O.M., MANCINI, G., CARNINO, A., GRIFFON, M., GAGNOLET, P (1981). Classification system for reporting events involving human malfunctions. In: RISO-M-2240. RISO National Laboratory. Denmark. March. 1981
30. RASMUSSEN, J.; PEJTERSEN, A.; GOODSTEIN, L.; Cognitive System Engineering. New York: John Wiley & Sons, 1994
31. REASON, J. Human error. Cambridge: Cambridge University Press, 1990.
32. REASON, J. Managing the risks of organizational accidents. London, UK: Ashgate Publishing, 1997.
33. REDISH, G (2007). Expanding Usability Testing to Evaluate Complex Systems. Journal of Usability Studies, Volume 2, Issue 3, May 2007, pp. 102-111
34. REID, G. B. & NYGREN, T. E. (1988). The Subjective Workload Assessment Technique: A scaling procedure for measuring mental workload, In: Human Mental Workload, (Eds. P. A. Hancock and N. Meshkati), Elsevier Science Publishers, North-Holland, 185-218.
35. SCHERER K. R. (2001). Appraisal Considered as a Process of Multi-Level Sequential Checking; in K. R. Scherer, A. Schorr, & T. Johnstone (Eds.) Appraisal Processes in Emotion: Theory, Methods, Research. New York and Oxford: Oxford University Press, pp. 92-120
36. SCHERER, D.; VIEIRA, M. F. Q. (2008) Accounting for the Human Error when Building the User Profile. In: Third IASTED International Conference Human-Computer Interaction, 2008, Innsbruck. Proceedings of Third IASTED International Conference Human-Computer Interaction. Zurich: ACTA Press, 2008. v. 1. p. 132-137
37. SCHERER, K. R. (2005). What are emotions? And how can they be measured? Social Science Information, 44, 695ï 729
38. SHAPIRO, S. S.; WILK, M. B. (1965) An analysis of variance test for normality (complete samples), Biometrika, vol. 52, no 3-4, 1965, p. 591ï 611
39. SZTAJZEL, J. (2004), Heart rate variability: a non-invasive electrocardiographic method to measure the autonomic nervous system. Swisse med wkly 2004; 134:514-522
40. TATTERSALL, A.J., & HOCKEY, G.R. (1995). Level of operator control and changes in heart rate variability during simulated flight maintenance. Human Factors, 37, 682ï 698.
41. THAYER, R. E. (1985). Activation (arousal): The shift from a single to multidimensional perspective. In J. Strelau, T. Gale, & F. Farley (Eds.), Biological bases of personality and behavior. Washington, D.D.: Hemisphere, pp. 115-127
42. THAYER, R. E. (1986). The Activation-Deactivation Adjective Check List: Current overview and structural analysis. Psychological Reports, 58, 607-614
43. THAYER, R.E. (1987). Problem perception, optimism, and related states as a function time of day (diurnal rhythm) and moderate exercise: Two arousal systems in interaction. Motivation and Emotion, 11 (1), 19-36

44. van EEKHOUT, J. M., & W. B. ROUSE (1981). Human Errors in Detection, Diagnosis, and Compensations for Failures in the Engine Control Room of a Supertanker. IEEE Transactions on System, Man, and Cybernetics, 1981, 12 ed., pp. 813-816
45. VIEIRA, M. F. Q. (2004). Accounting for Human Errors in a Method for the Conception of User Interfaces. In: International Mediterranean Modelling Multi Conference - I3M'04, 2004, Genoa, Italy. Proceedings of I3M'04. Bergeggi Italy. v. 1. p. 122-130
46. VIEIRA, M. F. Q.; QUEIROZ, J. E. R. de; OLIVEIRA, R. C. L. de (2005). Webquest: a configurable web tool to prospect the user profile and user subjective satisfaction. In: 11th International Conference on Human-Computer Interaction, 2005, Las Vegas - Nevada. HCII 2005. Las Vegas - Nevada: MIRA Digital Publishing, 2005
47. ZAR, J. H.. (1972) Significance testing of the Spearman rank correlation coefficient. Journal of the American Statistical Association, 1972, 67(339), 578-580.