MODELING AND SIMULATION OF OFFSHORE WORKERS' BEHAVIOR

by

© Mashrura Musharraf

A Thesis submitted to the

School of Graduate Studies

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Faculty of Engineering and Applied Science

Memorial University of Newfoundland

March, 2018

St. John's Newfoundland

ABSTRACT

The offshore oil and gas industry functions in a team work culture in which operations depend not only on individuals' competency, but also on team skills. Team skills are even more necessary when it comes to handling emergency conditions. Emergency conditions are dynamic in nature and personnel on board are challenged with evolving high-risk situations, time pressure, and uncertainty. One way to effectively handle emergencies is to train personnel to a competency level, both individually and as a part of a team. This would increase the chance of achieving safety in a timely manner using the available resources such as information, equipment, and people. Such training involves enhancing team members' understanding of human performance, in particular, the social and cognitive aspects of effective teamwork and good decision making. Post-accident analysis of offshore accidents shows that conventional training programs are often too generic, and that they are not designed to identify and tackle the human factors that are critical for evolving offshore emergency situations.

Recognition of the importance of human factors on operator performance raises the need for training that goes beyond conventional training programs and incorporates non-technical training focusing on leadership, command, decision making, communication, and teamwork. A major difficulty to design such training is that it involves practicing emergency exercises with a potentially large number of participants, each playing the appropriate role in a given scenario. Such large-scale team exercises suffer from both organizational and educational drawbacks. The amount of human and financial resources needed for such a training exercise is difficult to organize. Furthermore, it is very hard, if not impossible, to get all team members together at the same time and location. Also, the team members may have variability in the competency levels (novice versus advanced trainees) and hence different training needs. One effective and flexible solution to this problem is to use intelligent artificial agents, or ‘virtual workers’, in a virtual environment (VE) to play different roles in the team. Virtual workers are artificially intelligent agents that can reproduce behaviors that are similar to or compatible with those of a real worker. This research proposes to develop a human behavior simulation model (HBM) that can be used to create such virtual workers in the context of offshore emergency egress.

The goal of this research is to develop a human behavior model that can simulate offshore workers’ emergency response under the influence of performance influencing factors (PIFs). The first part of the work focuses on understanding human behavior during offshore emergency situations. A two level, three factor experiment was conducted in a virtual environment (VE) to investigate the relationships between the PIFs and human behavior. Influence of both internal and external PIFs were investigated. Knowledge acquisition and inference processes of individuals were also investigated in the experimental study. In the second part, a computational model was developed to capture the across-subject variability observed during the experiment. Interviews with subject matter experts (SME) were conducted at this step to ensure that the model is able to produce a realistic range of human behaviors. The final step was to validate the developed behavior model. All high-level tasks to validate the HBM were performed. Special emphasis was given on acceptability criteria

testing to ensure that the integrated HBM performs adequately under different operating conditions.

ACKNOWLEDGEMENTS

I would like to acknowledge my deepest gratitude to my supervisors, Dr. Faisal Khan, Dr. Brian Veitch and Dr. Scott MacKinnon. I have been fortunate to have supervisors whose inspiration helped me discover my passion for research. Their intriguing questions and constructive feedback have taught me to think like a researcher. Without their continuous guidance and encouragement this work would never have been completed successfully. I have thoroughly enjoyed my time working with this wonderful team.

Thanks to Dr. Syed Imtiaz for his tremendous support during the early days of my research. His advice and kind words have always been a source of motivation.

I acknowledge with gratitude the financial support of the Research & Development Corporation (RDC) of Newfoundland and Labrador, the NSERC-Husky Energy Industrial Research Chair in Safety at Sea, and Memorial University of Newfoundland.

Thanks to all the participants who volunteered for the experimental study. This work would not be possible without their interest and time.

A special thanks to Virtual Marine for collaborating on this project. Thanks to the entire Safety at Sea Team with special thanks to Jennifer and Sean. I will always cherish the wonderful memories in the lab. Thanks to the members of the Centre for Risk, Integrity and

Safety Engineering (C-RISE) for providing valuable feedbacks and suggestions on many occasions.

Thanks to my friend Norafneeza for always being there for me and reminding me the value of patience and hard work. Thanks to my St. John's family – Faten, Saad, Kashfi, Muntaser, and Milka – for their invaluable support over the years.

I can never thank enough my parents and family members for their tremendous support and patience. Their unconditional love gives me the energy to keep going. A very special thanks to my husband, Abir Zubayer, for his endless encouragements and motivational words. It would never be possible to pursue this journey without him by my side.

Table of Contents

List of Tables

List of Figures

Nomenclature

AVERT	All-hands virtual emergency response trainer
BN	Bayesian network
CES	Cognitive Environment Simulation
CGF	Computer generated force
COSIMO	Cognitive Simulation Model
CPT	Conditional Probability Table
CRM	Crew resource management
DMSO	Defense Modeling and Simulation Office
EAAGLES	Air-to-Ground Linked Environment Simulation
GPA	General Platform Alarm
HBM	Human behavior simulation model
HEP	Human error probability
HRA	Human Reliability Analysis
IDAC	Information, Decision, and Action in crew context
IPME	Integrated Performance Modeling Environment
LB	Lifeboat station
MAC	Manual alarm call point
MMA	Morale, motivation, and attitude
MOB	Man overboard
MOUT	Military Operations on Urbanized Terrain
MS	Muster station

OPSIM	Operator plant simulation
PA	Public Address
PAPA	Prepare to Abandon Platform Alarm
PIF	Performance influencing factors
PSF	Performance shaping functions
RPG	Recommended Practices Guide
SAMPLE	Situational assessment model of pilot in the loop evaluation
SME	Subject matter expert
UML	Unified Modeling Language
VE	Virtual environment

1. INTRODUCTION

1.1 Problem statement

The offshore oil and gas industry functions in a team work culture and operations usually involve a group of people working together. This makes teamwork an essential component of effective emergency responses. Members of a team need to understand their own roles and responsibilities, as well as have a clear understanding of the roles and responsibilities of the other team members. Such understanding is critical for emergency situations, as most of the members will have different roles and responsibilities than their everyday duties (Flin, 1997). Traditional training programs are often generic and are not designed to provide trainees with the understanding of social and cognitive aspects of effective team work.

O'Connor & Flin (2003) discuss the possibility of adopting the crew resource management technique, pioneered in the aviation industry, in offshore oil industries to enhance team performance. Crew resource management (CRM) is defined as “using all the available resources – information, equipment, and people – to achieve safe and efficient flight operations” (Moffat & Crichton, 2015). A significant part of the CRM training requires the trainees to participate in team training exercises using simulator flights. Organizing such team exercises for offshore industries may suffer from both organizational and educational drawbacks (Van Diggelen et al., 2010). Gathering all the team members at the same time and at the same location itself is a challenge. Even when it is possible, the financial requirement is high. Also, the members often have different training needs based on their competency levels. One solution to this problem is to develop a team training platform in a simulator where the roles of some of members are played by humans, while roles of others are played by artificial intelligent agents (Van Diggelen et al., 2010). This research aims to develop computational behavior simulation models that can be used to create such intelligent agents for an offshore emergency training simulator.

The purpose of the behavior simulation model is to reproduce the behavior of offshore workers, general personnel in particular, during offshore emergency situations. Compared to traditional human behavior models, the proposed model considers a larger fraction of the possible behavior space, which includes both correct and incorrect behaviors (Wray & Laird, 2003). To model the variability across behavior space, performance influencing factors (PIFs) are used in this research. PIFs are factors that can specifically decrement or improve human performance during a task (Blackman et al., 2008). In the first part of the work, emphasis is given to understanding human behavior variability during offshore emergency situations. A two-level, three factor experiment is conducted in a virtual environment (VE) to observe the influence of different PIFs on human emergency responses. The influence of both internal and external PIFs is investigated during the experiment. Knowledge acquisition and inference of individuals are also investigated in the experiment. In the second part of the research, a computational model is developed that capture the observed variability and are able to produce realistic human behavior. Finally, a validation experiment is designed and conducted to make sure that the model can simulate realistic human behavior in offshore emergency situations.

The rest of this chapter is organized as follows. As the experimental study and data collection in this research is centered around a VE for offshore emergency preparedness training, an introduction to the VE is presented in Section 1.2. Section 1.3 summarizes the works currently available in the behavior modeling domain, and identifies the knowledge and technological gaps. Section 1.4 defines the scope of work and objectives. Section 1.5 discusses approaches taken in this research to overcome the identified gaps. It also lists the novelty and expected contribution of the research. Section 1.6 presents the organization of the thesis.

1.2 Overview of the virtual environment (VE)

A VE is a computer aided simulation environment that allows trainees to gain artificial experience, including in dangerous scenarios. VE training can act as an enhancement to conventional training since training for emergency situations in the real world is not always ethically, logistically or financially feasible (Veitch et al., 2008). Besides facilitating emergency preparedness training, VE can also be used as a tool to observe human performance in emergency conditions and collect data for assessing human reliability (Lois et al., 2009; Bye et al., 2011; Monferini et al., 2013). The VE used in the experimental study done in this research is called the all-hands virtual emergency response trainer (AVERT) and was developed at Memorial University. AVERT was designed to enhance offshore emergency response training. The VE is modeled after an offshore oil installation platform with high levels of detail. It is capable of creating credible emergency scenarios by introducing hazards such as blackouts, fires, and explosions. For the experimental study done during this research, the offshore emergency scenarios covered a range of activities, from muster drills that required the participant to go to their primary muster station, to more complex emergency evacuation scenarios that required the participant to avoid hazards blocking their egress routes and muster at their lifeboat stations (House et al., 2014). Figure 1.1 shows a few instances of the AVERT emergency preparedness scenarios.

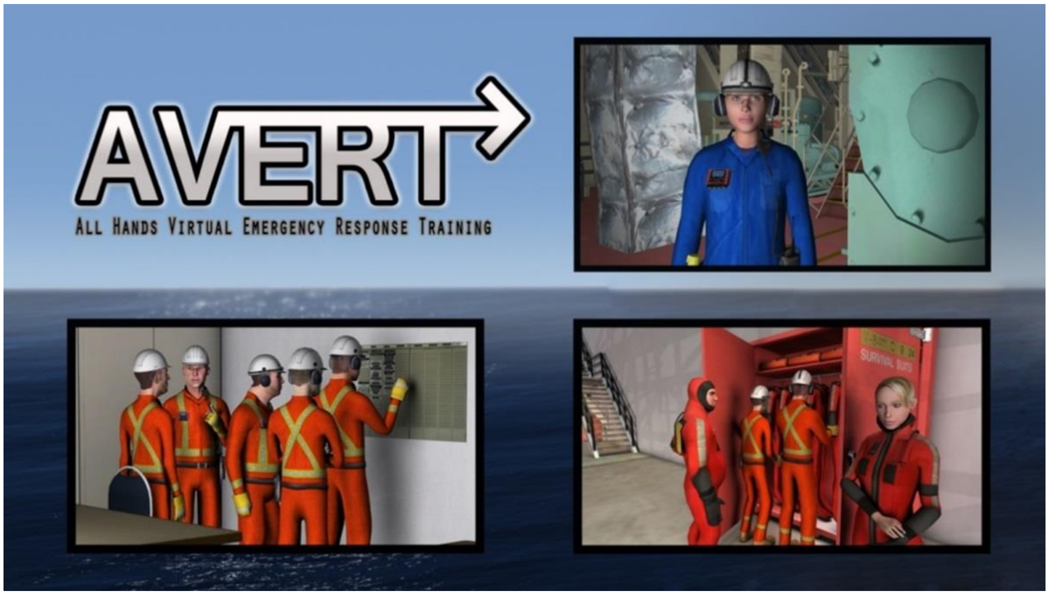


Figure 1.1: Screen capture of the virtual training environment - AVERT

1.3 Knowledge and technological gaps

Software agents, or computer generated forces (CGFs), are extensively used in a wide range of military applications, including training and rehearsal for combat situations (Karr et al., 1997). The use of virtual crew is also common in aviation and nuclear power plant simulation training (Chang & Mosleh, 2007a). Realism of agents in any platform largely depends on the sophistication of the underlying human behavior simulation models (HBM)

(Smith, 1998). This is why a significant amount of research has been done to develop computational models that can generate realistic human behavior.

Models of human behavior treat the human as a dynamic system that reacts to observed input from the environment (Huitt, 2009). Behavior simulation models can be qualitative or quantitative. Qualitative models focus on describing the evolution of the human cognition process upon receiving an external stimulus from the environment. This involves details of the cognitive functions – perception, interpretation, decision making, and execution (Thow-Yick, 1994; Trucco & Leva, 2007). Quantitative models are based on the structure of the qualitative ones, but have added computational functionalities. Quantitative models can probabilistically predict human response for a given circumstance.

Operator plant simulation (OPSIM), Cognitive Environment Simulation (CES), Cognitive Simulation Model (COSIMO), Information, Decision, and Action in Crew context (IDAC) are all examples of quantitative behavior models for nuclear power plant simulation. OPSIM models operator behavior and identifies possible human errors that might happen while following procedural instructions, but the probability of erroneous behavior is not incorporated in the model (Dang, 1996). CES and COSIMO aim to estimate operator behavior during power plant emergencies (Woods, 1987 and Cacciabue et al., 1992). CES uses a data base that represents operator knowledge. The content of the database, and the relationships between different knowledge units, are specified by knowledge engineers prior to the simulation. During simulation, the most likely crew response is calculated using artificial intelligence techniques that link various segments of the data base for a given situation. COSIMO shares the concept of using a data base to represent operator knowledge. However, the cognitive architecture is based on a stronger theoretical ground – the Fallible Machine model by Reason (1990). The cognitive architecture consists of two parts: the working memory and the knowledge base. The knowledge base is a virtually limitless repository of information that contains both declarative and procedural knowledge structures. The working memory is a limited, serial working area, and is the temporary storage of data required by the cognitive process. COSIMO focuses on the two fundamentals of cognition - similarity matching, and frequency gambling. During the similarity matching stage, attribute values of a given situation are compared to attribute values stored in the data base to find a match. If there is a conflict (i.e. more than one match found for the given situation), frequency gambling is used for conflict resolution by favoring the match that occurs most frequently. COSIMO also introduces the concept of using behavioral moderators to encode variability in the generated behavior. It does not include details on the behavioral moderator selection process, or the relationship between a moderator and human behavior. IDAC introduces the foundation of using PIFs as behavioral moderators (Chang & Mosleh, 2007b). In IDAC, operators' behaviors are probabilistically simulated under the influence of a number of explicitly modeled PIFs. Special attention has been paid to identify external, internal, static, and dynamic PIFs relevant to nuclear power plant accident scenarios. A set of rules-of-behavior is then developed that take the PIFs as input and generate behavior as output. As suggested by the authors, the rules-of-behavior used in IDAC are sufficient for demonstration of the methodology, but need further revisions for realistic and justifiable modeling (Chang & Mosleh, 2007c).

TacAir-Soar, Military Operations on Urbanized Terrain (MOUT), Air-to-Ground Linked Environment Simulation (EAAGLES), and AvatarSim are examples of military and air craft simulations that have made significant contribution to the development of realistic HBMs. TacAir-Soar is a model of expert human pilots flying tactical air mission (Jones et al., 1999). MOUT is an urban combat simulation used for building-clearing combat training (Sampson & Ripingill Jr, 2003). In a MOUT simulation, agents are used as both command team mates and opponents and are known as MOUTBots. Both TacAir-Soar and MOUTBots use the Soar architecture for cognition (Wray & Laird, 2003). The basic units of knowledge in Soar are production rules. These rules are used for defining goals and proposing, selecting, and applying actions for a given situation. Rules are collected from interviews with subject matter experts (SMEs) and are put in the knowledge base prior to simulation. TacAir-Soar focuses only on rules that generate correct behavior and ignores the possibility of erroneous behavior. MOUT offers incorporation of some erroneous behavior through behavioral moderators, but does not provide a reliable mathematical model that defines the relationship between the behavioral moderators and the choice of production rules. EAAGLES incorporate two mental models – the Situational Assessment Model of Pilot in the Loop Evaluation (SAMPLE) and Soar – to represent realistic combat behavior. Though different qualitative aspects of EAAGLES have been discussed in the literature, a reliable computational model is missing (McNally, 2005).

AvatarSim models and simulates human behavior in aircraft evacuations (Sharma, 2009). It uses psychological, environmental, and physical parameters that are natural in emergency evacuations. The psychological factors include stress, anger, and panic. Smoke, terrain, and smoothness are considered in the environmental category. Visibility, agility, and fitness are included as physical parameters. To model the uncertainty in behavior that results from the behavioral parameters, a fuzzy logic approach is used (i.e. IF Stress is of high intensity THEN speed is slow). Even though the use of behavioral parameters makes responses of AvatarSim naturalistic, it is limited in the sense that it only focuses on the effect of the parameters on agents' speed and wait time. It does not consider a broader range of behaviors that might be observed in real life emergency situations.

Once an HBM is developed, it needs to be validated to ensure that the model represents human behavior accurately. Compared to physics based simulation models, validating HBMs is much more difficult. Human behavior is complex and depends on a large number of PIFs. The PIFs can vary over many orders of magnitude and can have highly complex dependency relationships. Even small situation changes within the same system may cause different human responses. This makes HBM validation extremely difficult since the validation would require the exploration of a very large number of behavioral paths. Balancing the variability and the validation is identified as one of the most challenging problems in the domain of behavior simulation (Wray & Laird, 2003).

Because of the difficulty, so far, the most common validation technique for HBM is face validation (Goerger, 2004). In the face validation technique, an SME drives through the scenario space by issuing commands or changing the simulating situation, observes the resulting behavior, and determines, often qualitatively, whether the simulation meets a user's requirements for realism. Despite its wide application, Recommended Practices Guide on Validation of Human Behavior Representations (2001), describes face validation as the least reliable and least complete HBM validation. It discusses that, most of the time, SMEs' judgments are drawn from their own experience and can be biased. Face validation raises the possibility of conflict among multiple SMEs. It is also hard to ensure the level of consistency and accuracy of SMEs when evaluating human performance versus simulated human behavior.

Based on the literature review, following gaps between the existing methods and requirements are identified.

- Though extensive research has been done to develop HBMs for creating artificial intelligent agents in military applications, the aviation industry, and nuclear power plants, no such model is available to date for offshore emergency training simulators.
- Many HBMs focus only on the ideal human behavior and hence the success region of the total behavior space (McNally, 2005).
- HBMs that take erroneous behavior into consideration often lack a reliable modeling approach. Models often do not account for the potential dependencies among different PIFs and associated actions. Also, effects of PIFs on human behavior are often defined

using SMEs' opinions. Expert judgment can be vague and suffer from uncertainty, incomplete knowledge, and conflicts between multiple experts. Also, use of expert opinion relies on the underlying assumption that the PIFs affect all individuals in the same way (Joea & Boringa, 2014). Thus, expert opinion fails to account for the inherent variability in human nature.

- In most HBM systems, knowledge placed in the knowledge base is derived from interviews with SMEs. This fails to capture the variability in human learning and inference processes. Given the same training, people may learn and infer things differently and can have different approaches to solve the same problem.
- Though significant research is available on the development of HBMs, work done to validate the models is rare. A few attempts to validate HBMs use SMEs as referents (Harmon et al., 2002). Referent refers to a codified body of knowledge about a thing being simulated (Recommended practice guide (RPG): Special Topic - Validation of Human Behavior Representations, 2001). During validation, a referent provides the information to which the simulation outcomes are compared. As stated above, using SMEs as referents can make the validation biased and inconsistent.

1.4 Scope of work and objectives

The primary goal of this research is to develop HBMs that can simulate the behavior of offshore workers under the influence of different PIFs that are present in emergency situations. The work done toward this aim can be divided into three parts.

The first part focuses on understanding human behavior by observing people's performance in a VE. External PIFs that can influence people's performance during offshore emergency conditions were first selected. Credible emergency scenarios were then designed in the VE by varying the selected PIFs into different levels. An experiment was conducted to observe people's performance in the scenarios and collect human performance data. The collected data were divided into training and testing data sets. Figure 1.2 summarizes the purpose of the data sets.

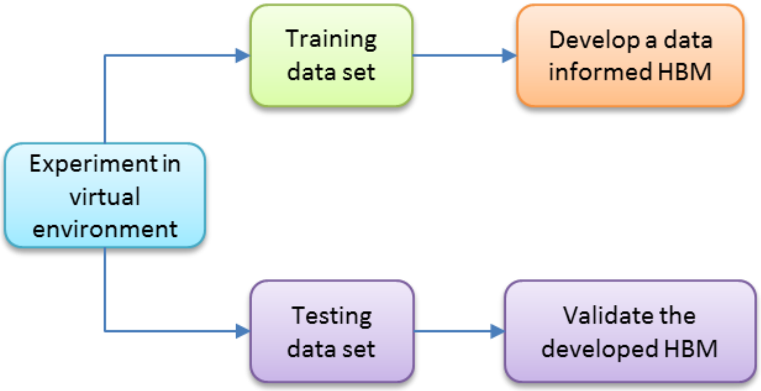

Training data set
Develop a data informed HBM
Experiment in virtual environment
Testing data set
Validate the developed HBM

Figure 1.2: Use of training and testing data sets

As shown in Figure 1.2, in the second part of the research, the training data set was used to develop an integrated HBM to reproduce the behavior of a general personnel. First, the basic task sequence of offshore general personnel was identified. Four types of cognitive tasks were considered during the task analysis – perception, interpretation, decision making, and execution (Edwards & Lees, 1974). Errors can happen while performing any of these tasks (Rasmussen, 1976). The probability of such error depends on the state of different PIFs and memorized information. A Bayesian network (BN) approach was used to model the impact of PIFs on human error. The training data set was used to quantify the BN model. Evidence collected during the experiment was also used to model the memorized information. The knowledge individuals gained from the training tutorials and scenarios was presented in the form of a knowledge matrix. An inductive reasoning algorithm – decision tree – was then used to identify the general principles or problem-solving strategies based on the individual cases in the knowledge matrix. The knowledge matrix and the decision trees together define the memorized information.

The third part of the research is focused on validating the HBMs using the testing data set. As listed in Defense Modeling and Simulation Office's (DMSO) Recommended Practices Guide (RPG), any HBM validation process needs to perform a few high-level tasks. The first task was to collect a set of requirements and acceptability criteria that set the foundation of the validation. Next, referents were to be identified to assess the credibility of the HBM. As mentioned earlier, both SMEs and empirical evidence were used as referents during the validation process. The conceptual model and the knowledge base were then validated using the referents and the defined requirements. During the validation of the conceptual model and the knowledge base, complex behavior areas of the model were identified for future validation activities. The final step was to validate the integrated HBM model using referents and requirements. This is called result validation and involves acceptability criteria testing by exercising testing scenarios to ensure that the integrated HBM performs adequately under different operating conditions. To perform this step, the HBM was integrated into AVERT to create software agents performing as general personnel. The complex areas identified in the previous step were used at the result validation step to design credible test scenarios.

Having the above scope of work, the objectives and associated tasks of this research can be listed as shown in Figure 1.3.

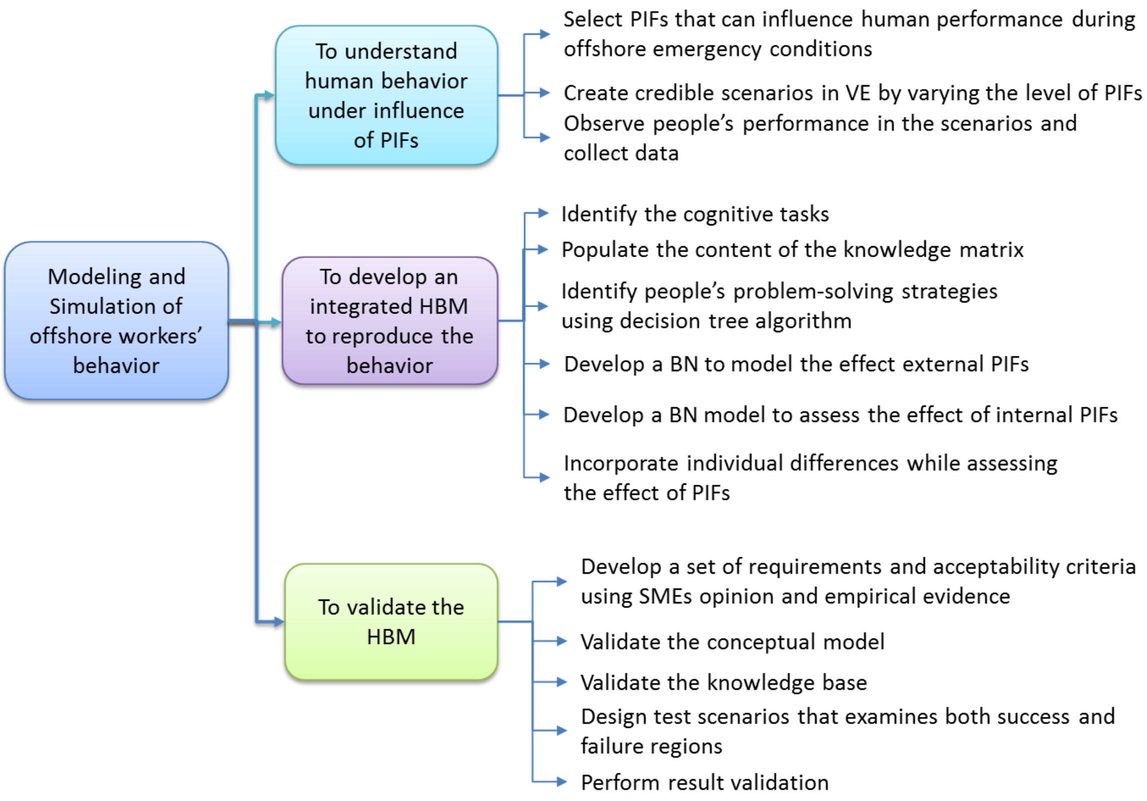

Modeling and Simulation of offshore workers' behavior
To understand human behavior under influence of PIFs
Select PIFs that can influence human performance during offshore emergency conditions
Create credible scenarios in VE by varying the level of PIFs
Observe people's performance in the scenarios and collect data
To develop an integrated HBM to reproduce the behavior
Identify the cognitive tasks
Populate the content of the knowledge matrix
Identify people's problem-solving strategies using decision tree algorithm
Develop a BN to model the effect external PIFs
Develop a BN model to assess the effect of internal PIFs
Incorporate individual differences while assessing the effect of PIFs
To validate the HBM
Develop a set of requirements and acceptability criteria using SMEs opinion and empirical evidence
Validate the conceptual model
Validate the knowledge base
Design test scenarios that examines both success and failure regions
Perform result validation

Figure 1.3: Objectives and associated tasks of this research

1.5 Novelty and contribution

This research attempts to overcome the gaps identified in Section 1.3 by taking the following steps:

- In this research, HBMs representing behavior of offshore workers during emergencies are developed. The goal of the research is to develop HBMs that can reproduce realistic human behavior for general personnel working offshore. To make the behavior naturalistic, both successful and erroneous behaviors are considered. The behavior paths generated by the HBM represent both success and failure regions of the total behavior space. Variability in behavior is encoded using internal and external PIFs.
- To model the effect of PIFs on human behavior, a BN approach is used. BNs have proven to be a powerful modeling tool due to their capability to 1) consider dependency among PIFs and associated actions, 2) quantify the impact of different PIFs on successful or erroneous behavior, and 3) update success or failure likelihood each time new evidence is available (Fenton & Neil, 2012; Podofillini & Dang, 2013; Sundaramurthi & Smidts, 2013). BNs have been widely used to model the impact of different PIFs on human performance or human error (Baraldi, et al., 2009; Dang & Stempfel, 2012). Kim & Seong (2006), Cai et al. (2013) and Martins & Maturana (2013) show examples of using the evidential reasoning aspect of BN to find the underlying causes of human error. Also, the BN model allows the incorporation of multiple sources of data into a single predictive HRA model (Groth & Mosleh, 2012). This research uses BN to model the effect of PIFs on task performance during offshore

emergency situations. Instead of using expert judgement, data required to quantify BNs are collected by conducting experiments in the virtual environment AVERT.

- This research acknowledges the fact that unlike machines, each human is different. Effects of different PIFs can vary from individual to individual. The virtual experimental data collection technique enables the consideration of individual differences while assessing and modeling people's success or failure likelihood.
- Special attention has been paid in this research to model the decision making of general personnel during an emergency. A data informed modeling approach is used. Data collected using the VE has been used to define the memorized information in the HBM. An inductive reasoning approach - decision tree - is then used to model the evolution of general understanding of emergency situations through training and experience (Han et al., 2011). Decision tree offers a visual representation of the reasoning process and has valuable diagnostic capabilities. Compared to other methods, such as artificial neural networks, or support vector machines, decision trees can be constructed relatively quickly. Another benefit of decision tree that is particularly important for this research is that it does not require any prior assumptions about the data and can work with limited data compared to other techniques (Duffy, 2008). Given a collection of training examples (condition x, action $f(x)$) the decision tree generates a hypothesis h that approximates the action $f(x)$. The aim of the reasoning process is to find a hypothesis that fits well with the training examples (Shaw et al., 1990). In this research, decision tree induction is used to generate a hypothesis based on the matrix of training

examples. Use of experimental data, rather than SMEs' opinions, allows capturing the actual observed variability in people's learning and decision making process.

- Special attention has been paid in this research to validate the developed HBM. All high-level tasks of HBM validation are performed. Special emphasis given on the acceptability criteria testing to make sure that the integrated HBM performs adequately under different operating conditions. Besides SMEs' opinions, empirical evidence has been used during the validation process. The outcomes of HBM are tested against the acceptability criteria established from the observations of human behavior in an experimental setup.

The expected contribution of the research includes:

- Primary contribution of this research is to enable offshore emergency preparedness team training. The HBM developed in this research is integrated into AVERT to create intelligent software agents that can play the role of general personnel with different levels of skill. Three types of agents – naïve, ideal, and in-between – are created to facilitate the team training process. This will give the opportunity to train personnel in a team environment to understand team roles, communicate effectively, gain assertiveness and leadership qualities, manage stress, and make group decisions. Training such non-technical skills, which are critical for successful emergency handling, will increase competency and enhance safety of the personnel working in offshore industries.

- The BN developed in this research can be used to assess people’s reliability during emergency situations. These results can be used to assess if someone is competent or reliable enough to handle emergency situations.
- Though the primary purpose of BN models developed in this research is to assess people’s response during emergency situations, they can also be used as a diagnostic tool. The BN model can quantify people’s sensitivity to different PIFs and identify their strengths and weaknesses. For example, if a participant is found to be more sensitive to a PIF, then training scenarios with different variations of that PIF can be provided to the participant until an accepted level of competency is reached. This kind of adaptive training can help individuals to obtain competency faster.
- Besides assessing the effect of external PIFs on human behavior, the research also looks into the effects of internal PIFs, such as bias, compliance, prioritization, and efficacy of information use. Conventionally, assessment of internal PIFs is done using a safety compliance questionnaire. Though questionnaires are sufficient to ensure that people have necessary knowledge about the safety procedures, they cannot ensure people will be able to apply that knowledge under the pressure of an emergency. In addition to questionnaires, this research uses virtual scenarios to assess internal PIFs. Assessment of internal PIFs using virtual scenarios can help to ensure that people not only know the safety procedures, but are also able to apply that knowledge during emergency situations.

- Sensitivity analysis done for the internal PIFs can be useful in the personnel selection process. Knowing if someone is compliant or a risk taker can help identify into which role they would best fit.
- The decision trees represent the behavioral pattern of individuals. Recognizing such patterns can be useful to predict what decision an individual might make for a given emergency situation. This can be extremely helpful in designing adaptive training so that individuals can reach competency faster.
- The decision trees also reflect the learning and inference of individuals given the training. The problem-solving strategies identified using decision trees can be used to assess the efficacy of the training curriculum and/or pedagogical approach. It is expected that a sound training process would ensure convergence amongst trainees to strategies that lead to success. A systemic exception might be an indication of weakness of the training approach itself. Identification of such weakness can help design better training curricula or pedagogy.
- The research demonstrates how use of empirical evidence along with SMEs' opinion can facilitate the HBM validation process.

1.6 Organization of the thesis

The thesis is written in manuscript format, including six journal papers as chapters. Table 1.1 shows the papers written during the course of this research and establishes their connection to the overall objectives and associated tasks listed in Figure 1.3.

Table 1.1: Papers and connection to the research objectives and associated tasks

Papers as chapters	*Research objectives*	*Associated tasks*
Chapter 2: Incorporating individual differences in human reliability analysis: an extension to the virtual experimental technique	• To understand human behavior under influence of PIFs • To develop an integrated HBM to reproduce the behavior	• Identify the cognitive tasks • Select appropriate external PIFs for offshore emergency situations • Create credible scenarios in VE by varying the level of PIFs • Observe people's performance in the scenarios and collect data • Develop a BN to model the effect *external* PIFs on human performance • Incorporate individual differences while assessing the effect of *external* PIFs
Chapter 3: Assessing offshore emergency evacuation behavior in a virtual environment using a Bayesian Network approach	• To understand human behavior under influence of PIFs • To develop an integrated HBM to reproduce the behavior	• Select appropriate *internal* PIFs • Develop a BN model to assess the effect of *internal PIFs* on human performance • Incorporate individual differences while assessing the effect of *internal PIFs*
Chapter 4: Identifying route selection strategies in offshore emergency situations using Decision Trees: A step towards adaptive training	• To understand human behavior under influence of PIFs • To develop an integrated HBM to reproduce the behavior	• Populate content of knowledge matrix • Identify people's problem-solving strategies using a reasoning algorithm (i.e. decision tree algorithm)
Chapter 5: Modeling and simulation of personnel response during offshore emergency situations	• To develop an integrated HBM to reproduce the behavior	• Integrate the BN model and reasoning structure to develop an HBM to reproduce the behavior of general personnel

Papers as chapters	***Research objectives***	***Associated tasks***
Chapter 6: Validating human behavior representation model of general personnel during offshore emergency situations	• To validate the HBM	• Develop a set of requirements and acceptability criteria using SMEs opinion and empirical evidence • Validate the conceptual model • Validate the Knowledge base • Design test scenarios that examines both success and failure regions • Perform result validation
Chapter 7: Human performance data collected in a virtual environment	• To provide data availability and direction towards future work	• Share the data collected during this research • Describe the data to facilitate reproduction if necessary • Discuss value of the data to help identify opportunities of future research collaboration

An outline of each chapter is presented below.

Chapter 2 describes the experimental study done in this research to collect human performance data. The PIFs varied to create virtual emergency scenarios are described in detail. The chapter then discusses the integration of the collected data into a BN to assess reliability of individuals. The chapter also demonstrates how use of the virtual experimental technique allows one to account for individual differences during reliability assessment.

Chapter 3 provides some more details on the experimental study. The focus of Chapter 3 is to investigate the effect of internal PIFs. The chapter shows how evidence collected from a virtual environment can be used to assess the internal PIFs.

Chapter 4 focuses on the decision making process of the general personnel during offshore emergency evacuation. Data collected during the experimental study is used to populate the knowledge matrix of the participants. An inductive reasoning technique – decision tree – is then used to identify the problem-solving strategies of the participants. The paper shows that given the same training, people may learn and develop the general understanding of emergency situations differently. This results in different problem-solving strategies (i.e. route selection strategies) across participants.

Chapter 5 describes how works done in Chapter 2 to 4 can be integrated to develop an HBM that reproduces the behavior of general personnel. The dynamic response model presented in this chapter consists of four component models - an environment model, an operator model, a performance shaping model, and a task network model. Understanding from Chapter 2 has been used to develop the environment model, understanding from Chapter 3 is used in the development of the operator model. The performance model uses the understanding from Chapter 2 to 4. The task network model was primarily based on (DiMattia, Khan, & Amyotte, 2005) with additional modification done according to SMEs' suggestions.

Chapter 6 focuses on the validation of the developed HBM. The chapter discusses the high-level tasks performed during the validation of an HBM. It starts by listing a set of requirements and acceptability criteria. It then discusses the conceptual model validation and knowledge base validation in detail. The complex behavior regions identified during conceptual model and knowledge base validation are used to design test scenarios for the result validation. Performance of the HBM is then tested in the designed scenarios and compared to the acceptability criteria established earlier using empirical evidence.

The aim of Chapter 7 is to make the data collected during the research publicly available for potential reuse. The data are described in detail to facilitate reproduction if necessary. The value of the data is discussed to help identify opportunities for future research collaboration.

Chapter 8 summarizes and concludes the thesis. It discusses the challenges faced during the research and provides recommendations to overcome them. The chapter also discusses potential future works.

A co-authorship statement is provided at the beginning of each chapter. The statement describes the contribution of each author in different stages of the research.

References

Baraldi, P., Conti, M., Librizzi, M., Zio, E., Podofillini, L., & Dang, V. (2009). A Bayesian network model for dependence assessment in human reliability analysis. *Proceedings of the Annual European Safety and Reliability Conference (ESREL),* (pp. 223-230). Prague.

Blackman, H. S., Gertman, D. I., & Boring, R. L. (2008). Human error quantification using performance shaping factors in the SPAR-H method. *Proceedings of the human factors and ergonomics society annual meeting* (pp. 1733-1737). Sage CA: Los Angeles: CA: SAGE Publications.

Bye, A., Lois, E., Dang, V., Parry, G., Forester, J., Massaiu, S., Boring, R., Braarud, P.Ø., Broberg, H., Julius, J. and Männistö, I. (2011). *International HRA Empirical Study - Phase 2 Report.* Washington: U.S. Nuclear Regulatory Commission.

Cacciabue, P. C., Decortis, F., Drozdowicz, B., Masson, M., & Nordvik, J. P. (1992). COSIMO: a cognitive simulation model of human decision making and behavior in accident management of complex plants. *Systems, Man and Cybernetics, IEEE Transactions, 22*(5), 1058-1074.

Cai, B., Liu, Y., Zhang, Y., Fan, Q., Liu, Z., & Tian, X. (2013). A dynamic Bayesian networks modeling of human factors on offshore blowouts. *Journal of Loss Prevention in the Process Industries, 26*(4), 639-649.

Chang, Y. H., & Mosleh, A. (2007a). Cognitive modeling and dynamic probabilistic simulation of operating crew response to complex system accidents: Part 1: Overview of the IDAC Model. *Reliability Engineering & System Safety, 92*(8), 997-1013.

Chang, Y. H., & Mosleh, A. (2007b). Cognitive modeling and dynamic probabilistic simulation of operating crew response to complex system accidents. Part 2: IDAC performance influencing factors model. *Reliability Engineering & System Safety*, *92*(8), 1014-1040.

Chang, Y. H., & Mosleh, A. (2007c). Cognitive modeling and dynamic probabilistic simulation of operating crew response to complex system accidents: Part 5: Dynamic probabilistic simulation of the IDAC model. *Reliability Engineering and System Safety, 92*(8), 1076-1101.

Dang, V. N. (1996). *Modeling operator cognition for accident sequence analysis: development of an operator-plant simulation.* Doctoral dissertation, Massachusetts Institute of Technology.

Dang, V., & Stempfel, Y. (2012). Evaluating the Bayesian belief network as a human reliability model - the effect of unreliable data. *Proceedings of the international conference on probabilistic safety assessment and management and the European safety and reliability conference PSAM 11 & ESREL 2012.* Helsinki, Finland.

DiMattia, D. G., Khan, F. I., & Amyotte, P. R. (2005). Determination of human error probabilities for offshore platform musters. *Journal of loss prevention in the process industries, 18*(4), 488-501.

Duffy, V. G. (2008). *Handbook of digital human modeling: research for applied ergonomics and human factors engineering.* CRC press Taylor & Francis Group.

Edwards, E., & Lees, F. P. (1974). *The human operator in process control.* London: Taylor & Francis.

Fenton, N., & Neil, M. (2012). *Risk assessment and decision analysis with Bayesian networks.* CRC Press.

Flin, R. (1997). Crew resource management for teams in the offshore oil industry. *Team Performance Management, 3*(2), 121-129.

Goerger, S. R. (2004). *Validating human behavioral models for combat simulations using techniques for the evaluation of human performance.* MONTEREY, CA: NAVAL POSTGRADUATE SCHOOL.

Groth, K. M., & Mosleh, A. (2012). Deriving causal Bayesian networks from human reliability analysis data: A methodology and example model. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, 226*(4), 361-379.

Han, J., Kamber, M., & Pei, J. (2011). *Data mining: concepts and techniques.* Elsevier.

Harmon, S. Y., Hoffman, C. W., Gonzalez, A. J., Knauf, R., & Barr, V. B. (2002). *Validation of human behavior representations.* Foundations for V&V in the 21st Century Workshop.

House, A. W., Smith, J., MacKinnon, S., & Veitch, B. (2014). Interactive simulation for training offshore workers. *Oceans'14 MTS/IEEE Conference* (pp. 1-6). St. John's, NL: IEEE.

Huitt, W. (2009). Humanism and open education. *Educational psychology interactive*.

Joea, J. C., & Boringa, R. L. (2014). Individual Differences in Human Reliability Analysis. *12th Bi-Annual International Meeting of the Probabilistic Safety Assessment and Management (PSAM) Conference.*

Jones, R. M., Laird, J. E., Nielsen, P. E., Coulter, K. J., Kenny, P., & Koss, F. V. (1999). Automated intelligent pilots for combat flight simulation. AI magazine. *AI magazine, 20*(1), 27.

Karr, C. R., Reece, D., & Franceschini, R. (1997). Synthetic soldiers [military training simulators. *IEEE spectrum, 34*(3), 39-45.

Lois, E., Dang, V.N., Forester, J., Broberg, H., Massaiu, S., Hildebrandt, M., Braarud, P., Parry, G., Julius, J., Boring, R. and Mannisto, I. (2009). *International HRA Empirical Study - Phase 1 Report.* Washington: U.S. Nuclear Regulatory Commission.

Martins, M. R., & Maturana, M. C. (2013). Application of Bayesian Belief networks to the human reliability analysis of an oil tanker operation focusing on collision accidents. *Reliability Engineering & System Safety, 110,* 89-109.

McNally, B. H. (2005). An approach to human behavior modeling in an air force simulation. *Proceedings of the 37th conference on Winter simulation* (pp. 1118-1122). Winter Simulation Conference.

Moffat, S., & Crichton, M. (2015). Investigating non-technical skills through team behavioral markers in oil and gas simulation-based exercises. *Procedia Manufacturing, 3*, 1241-1247.

Monferini, A., Konstandinidou, M., Nivolianitou, Z., Weber, S., Kontogiannis, T., Kafka, P., Kay, A.M., Leva, M.C. and Demichela, M. (2013). A compound methodology to assess the impact of human and organizational factors impact on the risk level of hazardous industrial plants. *Reliability Engineering & System Safety, 119,* 280-289.

O'Connor, P., & Flin, R. (2003). Crew resource management training for offshore oil production teams. *Safety Science, 41*(7), 591-609.

Podofillini, L., & Dang, V. N. (2013). A Bayesian approach to treat expert-elicited probabilities in human reliability analysis model construction. *Reliability Engineering & System Safety, 117,* 52-64.

Rasmussen, J. (1976). Outlines of a hybrid model of the process plant operator. In *Monitoring behavior and supervisory control* (pp. 371-383). Springer US.

Reason, J. (1990). *Human Error.* New York: Cambridge University Press.

Sampson, S. R., & Ripingill Jr, A. E. (2003). *System and method for training in military operations in urban terrain.* Washington, DC: U.S. Patent and Trademark Office.

Sharma, S. (2009). Avatarsim: A multi-agent system for emergency evacuation simulation. *Journal of Computational Methods in Sciences and Engineering, 9*(1, 2S1), 13-22.

Smith, R. D. (1998). Essential techniques for military modeling and simulation. *Proceedings of the 30th conference on winter simulation* (pp. 805-812). IEEE Computer Society Press.

Sundaramurthi, R., & Smidts, C. (2013). Human reliability modeling for the Next Generation System Code. *Annals of Nuclear Energy, 52*, 137-156.

Thow-Yick, L. (1994). The basic entity model: A fundamental theoretical model of information and information processing. *Information Processing & Management, 30*(5), 647-661.

Trucco, P., & Leva, M. C. (2007). A probabilistic cognitive simulator for HRA studies (PROCOS). *Reliability Engineering & System Safety, 92*(8), 1117-1130.

Van Diggelen, J., Muller, T., & Van den Bosch, K. (2010). Using artificial team members for team training in virtual environments. In *Intelligent Virtual Agents* (pp. 28-34). Springer Berlin Heidelberg.

Veitch, B., Billard, R., & Patterson, A. (2008). Emergency Response Training Using Simulators. *Offshore Technology Conference.*

Verification, Validation, and Accreditation (VV&A) Recommended Practices Guide (RPG): Special Topic - Validation of Human Behavior Representations (2001). Department of Defense Modeling and Simulation Office (DMSO). Retrieved from http://www.msiac.dmso.mil/vva/Special_topics/hbr-Validation/default.htm

Woods, D. D. (1987). Cognitive environment simulation: an artificial intelligence system for human performance assessment. *NUREG/CR-4862, 1-3*.

Wray, R. E., & Laird, J. E. (2003). Variability in human behavior modeling for military simulations. *Proceedings of Behavior Representation in Modeling and Simulation Conference (BRIMS).*

2. INCORPORATING INDIVIDUAL DIFFERENCES IN HUMAN RELIABILITY ANALYSIS: AN EXTENSION TO THE VIRTUAL EXPERIMENTAL TECHNIQUE

Mashrura Musharraf, Jennifer Smith, Faisal Khan**, Brian Veitch, Scott MacKinnon*

Centre for Risk, Integrity and Safety Engineering (C-RISE),

Faculty of Engineering & Applied Science,

Memorial University of Newfoundland,

St John’s, Newfoundland and Labrador, Canada A1B 3X5

*Department of Mechanics and Maritime Sciences,

Chalmers University, Gothenburg, Sweden

** Correspondence author: Tel: + 1 709 864 8939; Email: fikhan@mun.ca

Co-authorship statement

A version of this manuscript has been accepted for publication in the Journal of Safety Science. Authors Mashrura Musharraf and Jennifer Smith designed the experiment, conducted the experiment, and performed necessary data collection. The lead author Mashrura Musharraf performed the literature review, developed the Bayesian network for human reliability assessment, performed the data integration, generated the results, and prepared the draft of the manuscript. Co-authors Faisal Khan, Brian Veitch, and Scott MacKinnon supervised the experimental study. Faisal khan and Brian Veitch reviewed and corrected the Bayesian network model and results. All co-authors reviewed and provided feedback on the manuscript. Mashrura Musharraf revised the manuscript based on the co-authors' feedback and during the peer review process.

Abstract

Predicting human behavior and assessing human performance in offshore emergency conditions is a challenge. There are many human reliability analysis (HRA) methods available today, however none of these methods are applicable in the context of offshore emergencies. The data required to perform HRA for emergency conditions are not readily available and are difficult to retrieve from accident investigations. In the absence of emergency conditions data, the conventional approach of gathering data for HRA is using expert judgment. Expert judgment often suffers from uncertainty, subjectivity, and incompleteness which makes the reliability of this data collection technique questionable. Moreover, the technique has an underlying assumption that the influence of different factors on human performance is the same for all individuals. A more recent approach is to collect data by conducting experiment in virtual environments with human subjects. Though virtual experimental technique addresses the issues of uncertainty, subjectivity, and incompleteness, it still does not consider individual differences while assigning the influence of different factors on human performance. Unlike machines, each human is different and the influence of factors on performance may vary from individual to individual. This paper proposes to advance the virtual experimental technique by enabling the consideration of individual differences. An experiment using virtual environment was done to observe performances of 36 individuals during offshore emergency evacuation. By integrating the data collected from the virtual environment into an HRA model, the

reliability of each individual was assessed. Sensitivity analysis was then performed to identify the most influential factors that contributed to failure in emergency conditions. This analysis can help identify specific weaknesses that a participant might have. For example, if a participant is found to be more sensitive to a particular factor, then training scenarios with different variations of the factor can be provided to the participant until an accepted level of competency is reached. Identification of a weakness can be combined with adaptive human factor training so that each individual can obtain competence more quickly.

2.1 Introduction

Human reliability is defined as the probability that a person correctly performs system-required activities in a designated time period (Swain & Guttmann, 1983). There are many human reliability quantification techniques available today to assess how reliable humans are in different contexts. Examples include: Success Likelihood Index Methodology (SLIM), Technique for Human Error Rate Prediction (THERP), and A Technique for Human Error Analysis (ATHENA) (Kirwan, 1994; Cooper et al., 1996). The Bayesian network (BN) approach has also been applied to human reliability analysis (HRA) (Baraldi et al., 2009). Most of the human reliability quantification techniques involve the calculation of human error probability (HEP), which is the probability that a person will fail to carry out a task as required (Kirwan, 1994). Performance influencing factors (PIFs) are often used to calculate HEP (Blackman et al., 2008). Human performance, and hence error, is influenced by PIFs, and therefore the relationship between PIFs and human errors must be defined to calculate HEP. Due to lack of real or ecologically-valid data, the majority of the human error prediction techniques (i.e. SLIM, THERP, BN) often use expert judgment to define this relationship. Though expert judgement is a valuable technique, it can suffer from uncertainty, subjectivity, and incompleteness. Significant conflict among judgements may also arise when collected from multiple experts. Recent works (Musharraf et al., 2014) have proposed the use of virtual experimental technique as an alternative to expert judgement. This technique collects empirical evidences required to perform a human reliablity assessment by conducting experiments in virtual environments with human subjects. However, this work does not account for individual differences when it comes to the influence or importance of PIFs on human errors. Humans are inherently different and therefore the role that different PIFs play on performance may vary from individual to individual. For example, consider a case where complexity and visibility are two different PIFs that can influence one's performance during an evacuation. While complexity can play a more important role than visibility for one individual, it can be the other way around for another individual. This paper proposes an expansion of the virtual experimental technique to account for individual differences during the HRA process. In this paper, the term individual difference refers to the difference between the sensitivity of two individuals to external PIFs. It does not cover the more general aspects that might differ between individuals such as gender, education, and physical characteristics.

The HRA technique used in this paper is the BN approach. BNs have proven to be a powerful tool for HRA for the following reasons: 1) this approach can consider the dependencies among PIFs and the associated actions, 2) it can incorporate new evidence and update the HEP, and 3) it can support the root-cause analysis of human error (Podofillini & Dang, 2013; Sundaramurthi & Smidts, 2013). BNs have been widely used to model the impact of different PIFs on human performance or human error (Baraldi, et al., 2009; Dang & Stempfel, 2012). Kim & Seong (2006), Cai et al. (2013) and Martins & Maturana (2013) show examples of using the evidential reasoning aspect of BN to find the underlying causes of human error. Also, the BN model allows the incorporation of multiple sources of data into a single predictive HRA model (Groth & Mosleh, 2012b). A more comprehensive list of the demonstrated benefits of BN for HRA in different domains can be found in Groth & Swiler (2013) and Mkrtchyan et al. (2015).

In this paper, a BN model is developed to observe the impact of two PIFs (complexity and visibility) on human error during an offshore emergency evacuation. In this model, PIFs and errors are all random variables, and the probability of an error occurring is conditionally dependent on the PIFs. To define conditional dependencies in the BN, necessary data were collected from a study conducted in a virtual environment with 36 participants. At the beginning of the study each participant was assigned to one of two training groups: 1) G1: high level training and 2) G2: low level training. The training level assigned to each participant remained unchanged for the rest of the experiment. Virtual emergency scenarios were created with different levels of visibility (clearly visible versus blackout conditions) and complexity (low complexity, such as a muster drill vs. high complexity, such as a dynamic emergency situation). Participants' performance in the series of virtual emergency scenarios were observed. By integrating the performance data into the BN, the reliability of each subject was assessed. Next, sensitivity analysis was performed to find the relative contribution of the PIFs to failure.

Section 2.2 gives an overview of the BN approach to HRA and the virtual environment used in the experiment. Section 2.3 describes the methodology, data collection and integration using a case study of offshore emergency evacuation. Section 2.4 presents and explains the results. The limitation of the study and future works are discussed in Section 2.5. Section 2.6 summarizes and concludes the paper.

2.2 Background

2.2.1 Bayesian network (BN) approach to HRA

A BN approach was used to calculate the HEP. According to Pearl (1988), BNs are acyclic directed graphical models that represent conditional dependencies among a set of random variables. While performing a task or exercise, errors can occur at different steps of the process. Each error is regarded as the outcome of the joint influence of different PIFs (as depicted in Figure 2.1). In the BN approach to HRA, error is the critical node which depends on several PIFs that can influence the occurrence of the error. For example, in an offshore emergency evacuation situation, interacting with hazards (e.g. smoke or fire) is an error that may occur because the visibility is compromised (PIF_1), or the operator is not familiar with the complexity of the situation (PIF_2), or both. Figure 2.1 shows the relationship between human error and PIFs. This paper investigates the impact of only two

PIFs (visibility and complexity) on human error. A comprehensive list of PIFs can be found in Groth & Mosleh (2012a) and Mearns et al. (2001).

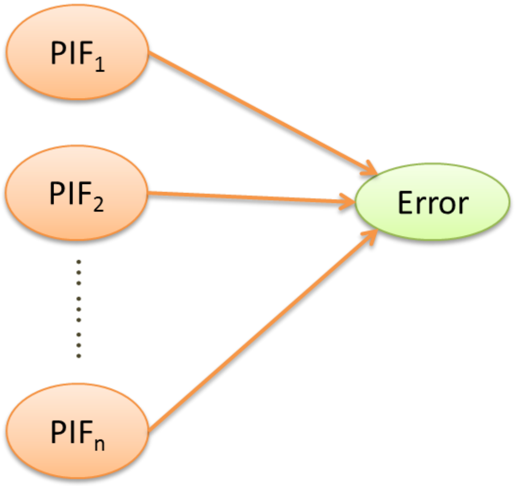


Figure 2.1: Relationship between PIFs and human error. Error is the outcome of joint influence of PIF$_1$ to PIF$_n$.

To define the relationship between a human error and PIFs, two parameters are needed: 1) the prior belief (in terms of probabilities) of the PIFs and 2) the conditional belief (in terms of probability distribution) of the human error. In this case, prior probabilities of all possible states of a PIF are assumed equal (50% if the PIF is binary). The difficult part is to define the conditional probabilities, which represent the conditional dependency of human error on PIFs. This paper uses data collected in a virtual environment to define these conditional dependencies. Conditional dependencies are defined separately for each individual to reflect the fact that influence of PIFs on error may vary from individual to individual. Section 2.3 illustrates the approach in detail.

Once the probabilities of different errors during a task are calculated, they can be combined using the definitional/synthesis idiom, rather than a causal relationship, to achieve an overall failure probability for the task (Fenton & Neil, 2012). For example, in an offshore emergency evacuation situation, if an operator is interacting with a smoke hazard *($Error_1$)* while keeping all fire doors open throughout the evacuation process *($Error_2$)*, then these errors can be combined to get an overall failure probability of the operator for the task evacuation. To reduce the computational complexity, errors ($Error_{1\text{-}n}$) are first classified into categories *($CT_{1\text{-}m}$)* and then combined to get an overall failure *(F)* probability. The different categories of error considered in this paper are as follows: perception error, recognition error, procedural error, and lack of situational awareness. Each error can be classified into one or more categories. For example, interaction with a hazard can be categorized as a failure to perceive the severity of the hazard (perception error) and keeping fire doors open can be categorized as a procedural error. Figure 2.2 shows how error probabilities in different categories can be combined to quantify the overall failure probability.

As shown in Figure 2.2, there are two relationships that need to be defined: 1) the relationship between the errors *($Error_{1\text{-}n}$)* and different categories *($CT_{1\text{-}n}$)* and 2) the relationship between different categories ($CT_{1\text{-}n}$) and overall failure *(F)*. Two parameters are needed to define these relationships: 1) the conditional belief (in terms of probability distribution) of the categories *(CT)*, and 2) the conditional belief (in terms of probability distribution) of the overall failure *(F)*.

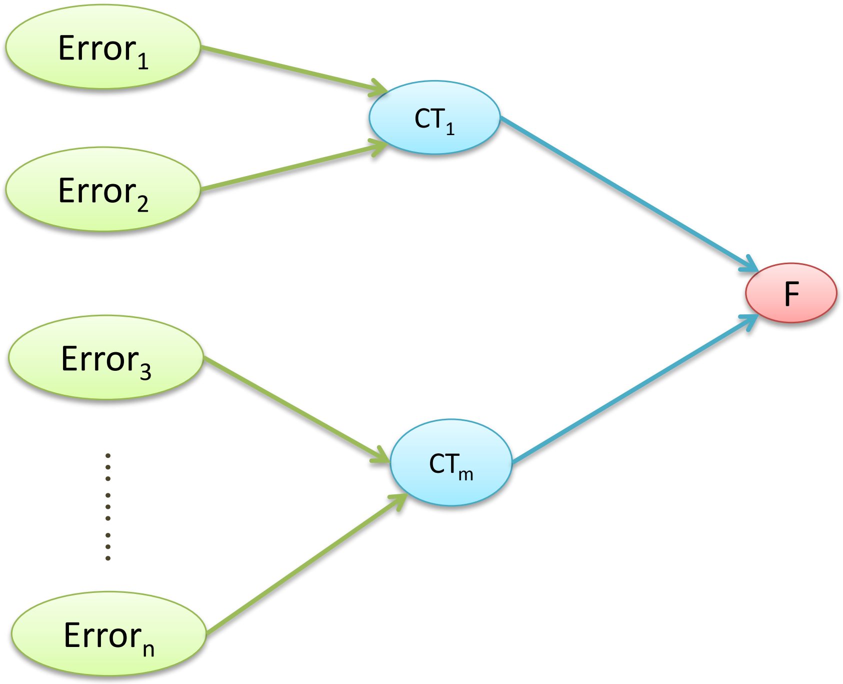


Figure 2.2: Combining Error$_1$ to Error$_n$ to get an overall failure probability. Error$_1$ to Error$_n$ are first combined according to categories (CTs), the categories are then combined to get overall failure (F) probability.

To demonstrate how conditional probability distribution of *CTs* can be defined, a simple case is considered where the category variable CT_1 is binary and can have two possible states: *acceptable* and *not acceptable*. CT_1 is assumed to be dependent on $Error_1$ and $Error_2$. Table 2.1 shows the conditional probability table for CT_1. As shown in the table, P(CT_1=*Acceptable*) becomes zero if either $Error_1$ or $Error_2$ occurs. The only case when P(CT_1=*Acceptable*) becomes one is when none of the errors have occurred.

The conditional probability table for the failure node *F* can be defined in the same way. A simple case can be considered where *F* is binary and can have two possible states: *Yes* and

No. Table 2.2 shows the conditional probability table for *F* when it is dependent on CT_1 and CT_2. As shown in Table 2.2, if either CT_1 or CT_2 is not acceptable, P(*F*=*Yes*) becomes one. P(*F*=*Yes*) becomes zero when both CT_1 and CT_2 are acceptable.

Table 2.1: Conditional probability table for category (CT_1)

$Error_1$	$Error_2$	$P(CT_1\text{=}Acceptable \mid Error_1,Error_2)$
No	No	1
Yes	No	0
No	Yes	0
Yes	Yes	0

Table 2.2: Conditional probability table for failure (*F*)

CT_1	CT_2	$P(F\text{=}Yes \mid CT_1, CT_2)$
Acceptable	Acceptable	0
Not acceptable	Acceptable	1
Acceptable	Not acceptable	1
Not acceptable	Not acceptable	1

It has to be noted that, the relationships shown in Table 2.1 & 2.2 are defined by the analyst and are not dependent on the collected data. These relationships are context sensitive and may need to be redefined by the analyst for a given situation. Also, category variables are considered binary in this example only to simplify the illustration. In reality, the category variables can have two or more possible states depending on the context.

Using the relationships shown in Figure 2.1 & 2.2, the final network can be developed (as shown in Figure 2.3).

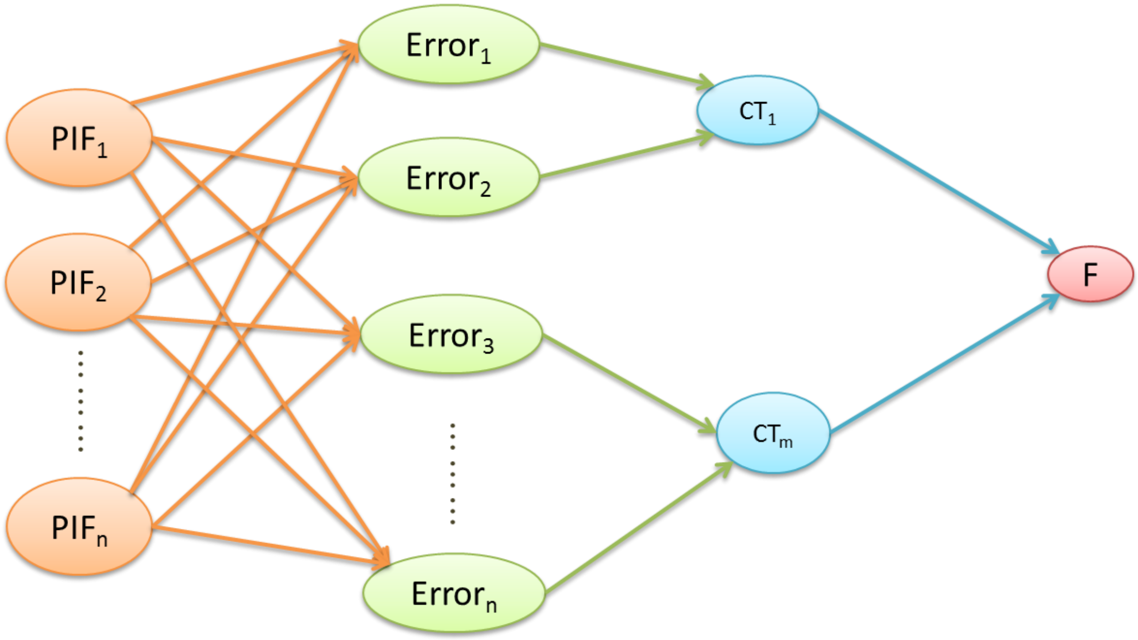


Figure 2.3: BN to show causal dependency between PIFs, errors, and overall failure.

2.2.2 Overview of virtual environment

A virtual environment is a computer aided simulation environment that allows trainees to gain artificial experience including performing in dangerous scenarios. Virtual environment training can act as an enhancement to conventional training since training for emergency situations in the real world is ethically, logistically or financially unfeasible (Veitch et al., 2008). Besides facilitating emergency preparedness training, virtual environments can also be used as a tool to observe human performance in emergency conditions and collect data for HRA (Lois et al., 2009; Bye et al., 2011; Monferini et al., 2013). The virtual environment used in the case study is called the all-hands virtual emergency response trainer (AVERT) and was developed at Memorial University. AVERT was designed to enhance offshore emergency response training. The virtual environment is modeled after an offshore oil installation platform with high levels of detail. It is capable of creating credible emergency scenarios by introducing hazards such as blackouts, fires and explosions. For the case study, the offshore emergency scenarios covered a range of activities, from muster drills that required the participant to go to their primary muster station, to more complex emergency evacuation scenarios that required the participant to avoid hazards blocking their egress routes and muster at their lifeboat stations (House et al., 2014). The scenarios in the case study were designed using AVERT to observe the effect of the PIFs visibility and complexity on individuals' performance during offshore emergency evacuation. Details of the case study are presented in the next section.

2.3 Case study: Offshore emergency evacuation in a virtual environment

2.3.1 Experimental setup

The data used in this paper were originally collected during an experimental study presented in Smith (2015) and Musharraf et al. (2016). This paper uses the data collected during the study to demonstrate the incorporation of individual differences in HRA.

A total of 36 participants took part in the study with a goal to learn how to perform a successful offshore emergency evacuation. The participants were naïve concerning any detail of the experimental design, they were not employed in the offshore oil and gas industry, and therefore they were not familiar with the offshore platform. Each participant was assigned to one of two groups for training: 1) G1: high level training and 2) G2: low level training. Participants in both groups attended 3 sessions. The content of each session was different between the two groups. In the first session, both groups received a basic offshore emergency preparedness tutorial. G1 then received 4 training scenarios, a multiple choice test and 4 testing scenarios. G2 only received the multiple choice test and 4 testing scenarios after the tutorial. In both Session 2 and Session 3, G1 received an advanced training tutorial about alarms and hazards respectively, 4 additional training scenarios, a multiple choice test, and 4 testing scenarios. G2 received no advanced training tutorial and only received a multiple choice test and 4 testing scenarios in Sessions 2 and 3. Both groups were provided with feedback on their performance in the multiple choice test and virtual environment testing scenarios in each session.

2.3.2 Design of the experiment

Once a participant was assigned to a group, his/her training level remained static (either low or high) for the rest of the study. The other two PIFs: visibility and complexity, on the other hand, were set to different levels to investigate how these PIFs influence each participant.

Visibility refers to the amount of ambient light available while performing a specific task. This PIFs was varied at two different levels: *clearly visible* and *blackout*. In clearly visible conditions, there was enough ambient light to perform the assigned task. In the blackout conditions, the visibility was reduced by reducing the available ambient light. However, the participants were allowed to use a virtual flashlight in the blackout conditions. The flashlight allowed participants to have a limited but functional visibility.

Complexity in this context refers to the difficulty of any given situation and the degree of responsibilities of an individual in that situation. The more difficult the situation is, the higher is the chances of human error. Similarly, higher responsibilities also imply higher chances of human error. Two levels of complexity were considered in this experiment: *low* and *high*. In low complexity conditions, there was no obstacle in the egress route, and the responsibility assigned to the participant was minimal. High complexity situations were created by blocking the escape routes with hazards (i.e. smoke, fire, and explosion), and assigning more responsibilities to the participants.

Training and testing scenarios were designed with different levels of visibility and complexity. Several performance metrics of the participants were recorded during each scenario. The following are the performance metrics that are most relevant to this paper: time to muster, time spent running, interaction with fire doors and watertight doors, interaction with hazards, and reporting at muster stations. Replay videos of participants' performance in the scenarios were also recorded for further analysis. For HRA purposes, only the performance metrics collected during the testing scenarios were used. Figure 2.4 presents a schematic diagram of the experimental design.

There were 4 testing scenarios in each session. For demonstration purposes, only the testing scenarios in Session 3 will be used in this paper. Table 2.3 gives an overview of the 4 testing scenarios in session 3.

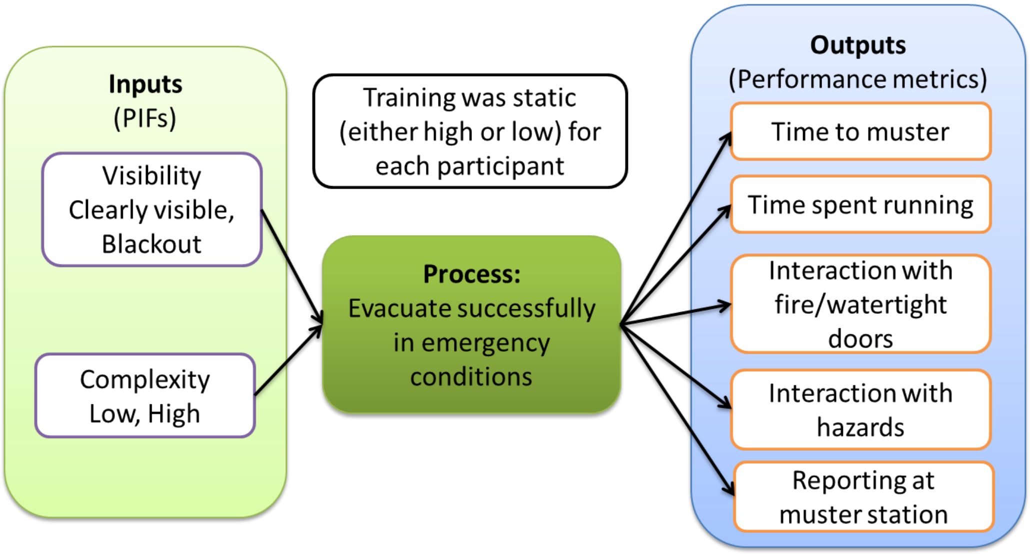


Figure 2.4: Schematic diagram of the experimental design

Table 2.3: Testing scenarios created in AVERT varying the state of the PIFs (for session 3)

Scenario Name	*PIF1: Visibility*	*PIF2: Complexity*	*Context*
Scn1	Normal	Low	A fire and explosion on the helideck signal a GPA. High winds cause the smoke to engulf a portion of the platform exterior. The participant must go to muster station, but re-route to lifeboat station due to the increase in emergency severity and the alarm change to PAPA. Complexity is low as the hazards do not block the primary route through the main stairwell.
Scn2	Normal	High	Fire erupts in the galley signaling a GPA. The participant must go to the muster station but re-route to the lifeboat station due to the fire and smoke spreading to the adjacent muster station. Complexity is high as the primary egress route and the muster station are compromised by the hazards.
Scn3	Blackout	Low	An electrical fire and dense smoke fill a portion of the engine room. The GPA sounds. The participant must go to muster station but re-route to lifeboat station due to the increase in situation severity and alarm change to PAPA. Complexity is low as the participant was assigned to only one task: evacuate successfully.

Scenario Name	*PIF1: Visibility*	*PIF2: Complexity*	*Context*
Scn4	Blackout	High	A fire and explosion occur in the main engine and result in a vessel-wide blackout. The alarm is not immediately triggered. The fire blocks access to the secondary egress routes. The participant must raise the alarm and go to the muster station but re-route to lifeboat station due to the increase in situation severity and alarm change to PAPA. Complexity is high as the participant had an additional responsibility of raising the alarm before evacuation.

2.3.3 Bayesian Network (BN) for the case study: Data collection and integration

The primary interest of the case study was to account for the difference between individuals while defining a relationship between the PIFs (visibility and complexity) and human error during offshore emergency evacuation. For this purpose, a set of possible errors during an offshore emergency evacuation was defined. Table 2.4 shows a list of possible errors and different error categories during an offshore emergency evacuation. For a BN approach to HRA, all listed errors are critical nodes and depend on the states of the PIFs. The prior probabilities of each state of the PIFs (visibility and complexity) are assumed 50%. The next step is to define the conditional probability distribution of each error. Table 2.5 shows an example conditional probability table for the error *"Wrong muster station"*.

Table 2.4: List of possible errors during offshore emergency evacuation and their corresponding category

Error Category	*Possible errors*
Perception of hazard	Interacting with fire
	Interacting with smoke
Situational awareness	Taking more time to muster than necessary
	Going to wrong muster location
Recognition of alarm	Failing to follow alarm and going to wrong muster location
Compliance with basic safety procedures	Running on the platform
	Leaving fire doors and/or watertight doors open

Table 2.5: Conditional probability table for error *"wrong muster station"*

Visibility	*Complexity*	*P(Wrong muster station = Yes)*	*P(Wrong muster station = No)*
Normal	Low	These conditional probabilities were defined using the data collected during the experiment.	
Normal	High		
Blackout condition	Low		
Blackout condition	High		

As shown in Table 2.5, there are eight conditional probabilities that need to be defined. Data collected from virtual environment scenarios were used in this paper to obtain these probabilities. Each participant was tested in 4 testing scenarios during each session: Scn1 (visibility=normal, complexity=low), Scn2 (visibility=normal, complexity=high), Scn3 (visibility=blackout condition, complexity=low) and Scn4 (visibility=blackout condition, complexity= high). For instance, in Scn1, if the participant went to the wrong muster station, P(wrong muster station=Yes| visibility=normal, complexity=low) = 1 and hence P(wrong muster station=No| visibility=normal, complexity=low) = 0. The other conditional probabilities were defined the same way. Table 2.6 shows the conditional probability table for the error *"wrong muster station"* after the data were collected for one participant. It should to be noted that, unlike the conventional approaches, these conditional probabilities are defined for each individual and may vary from participant to participant. Figure 2.5 shows the total probability of *"wrong muster station"* after the collected data were integrated.

Table 2.6: Conditional probability table for the error *"wrong muster station"* after collecting data for one participant

Visibility	*Complexity*	*P(Wrong muster station = Yes)*	*P(Wrong muster station = No)*
Normal	Low	0	1
Normal	High	0	1
Blackout condition	Low	0	1
Blackout condition	High	0	1

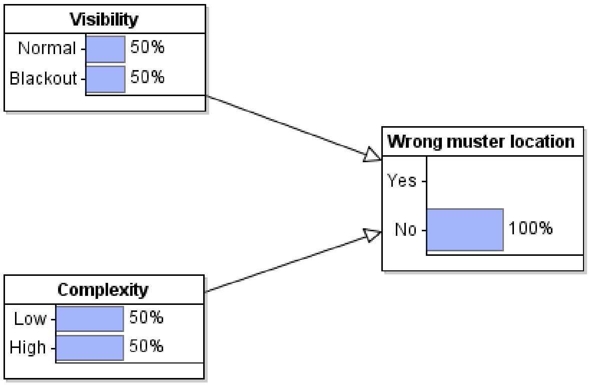


Figure 2.5: Total probability of "wrong muster location" after integrating the data collected for the participant in Table 2.6

The data were collected and integrated similarly for all the errors listed in Table 2.4. The errors were combined as presented in Figure 2.3, to get probabilities for the different error categories. Finally, the different categories of error were combined to get the overall failure probability for one participant. The conditional probabilities of the error categories and failure were defined using the same approach shown in Table 2.1 and Table 2.2. Figure 2.6 shows the final BN for the participant.

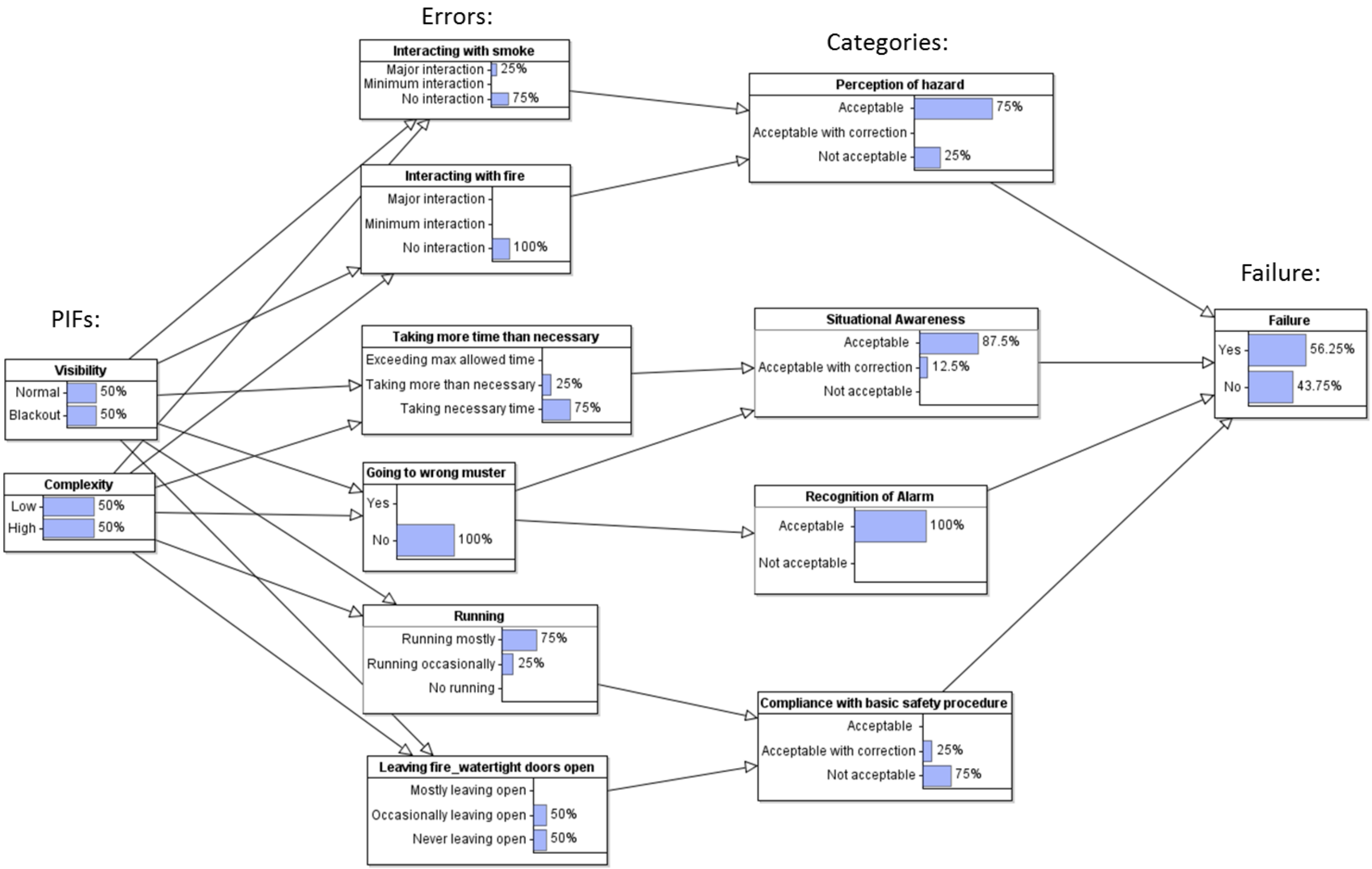


Figure 2.6: Final BN for one sample participant

2.4 Results and discussion

2.4.1 Results of complete study

Section 2.3.3 explained the failure probability calculation for one sample participant in detail. The failure probability was calculated for all 36 participants in the same way. Figure 2.7 shows the histogram of failure probability for all participants. As shown in the figure, almost 83% of the participants had a failure probability of 50% or higher.

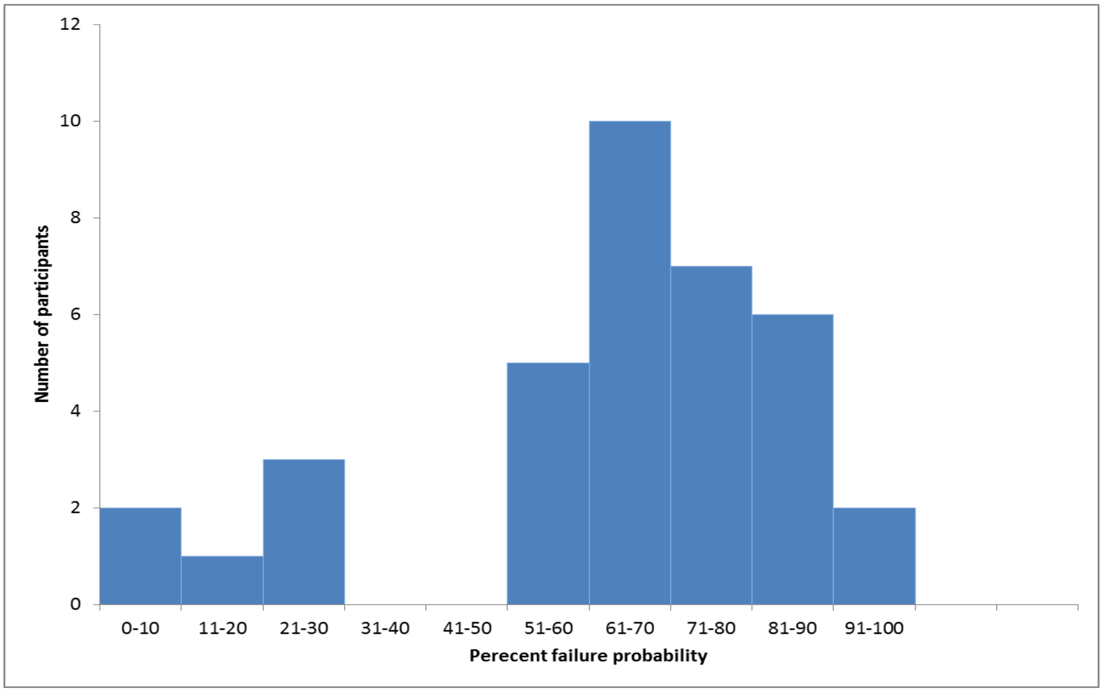


Figure 2.7: Histogram of percent failure probability for 36 participants in Session 3

As stated in Section 2.3.1, the participants were divided into two groups and G1 had a more advanced level of training than G2. A comparison between the failure probability of G1 and G2 shows that the average failure probability of G2 was much higher (63.5%) compared to G1 (43.2%). This is consistent with the expectation that advanced training can reduce the likelihood of failure in emergency conditions.

2.4.2 Sensitivity analysis: which PIF contributes most to failure?

Once the failure probability of a participant was calculated, sensitivity analysis was performed to determine which PIF (visibility or complexity) contributed most to failure for the given participant. Figure 2.8 shows the tornado graph (Fenton & Neil, 2012) of sensitivity analysis for the same participant as in section 2.3.2. As shown in Figure 2.8, the probability of failure given complexity went from 0.375 to 0.75 (when changing complexity from low to high). Similarly, probability of failure given visibility went from 0.5 to 0.625 (when changing visibility from a blackout to normal conditions). For the participant under consideration, complexity is the node that has the highest contribution to failure.

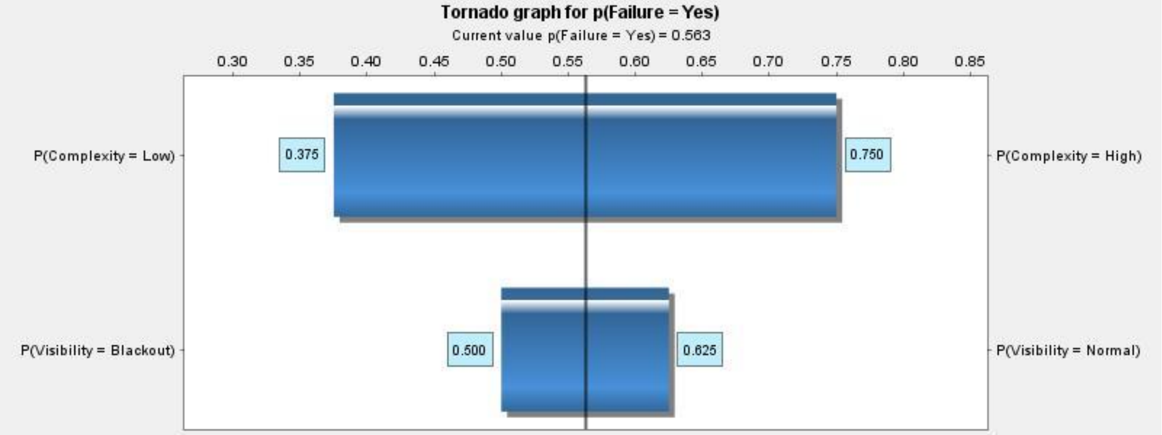


Figure 2.8: Tornado graph showing which node most impact failure (for one sample participant)

Sensitivity analysis was done similarly for all participants, and Figure 2.9 summarizes the results of this analysis. 19% of the participants were found to be more sensitive to complexity and 11% were found to be more sensitive to visibility. The rest of the participants were equally sensitive to both complexity and visibility.

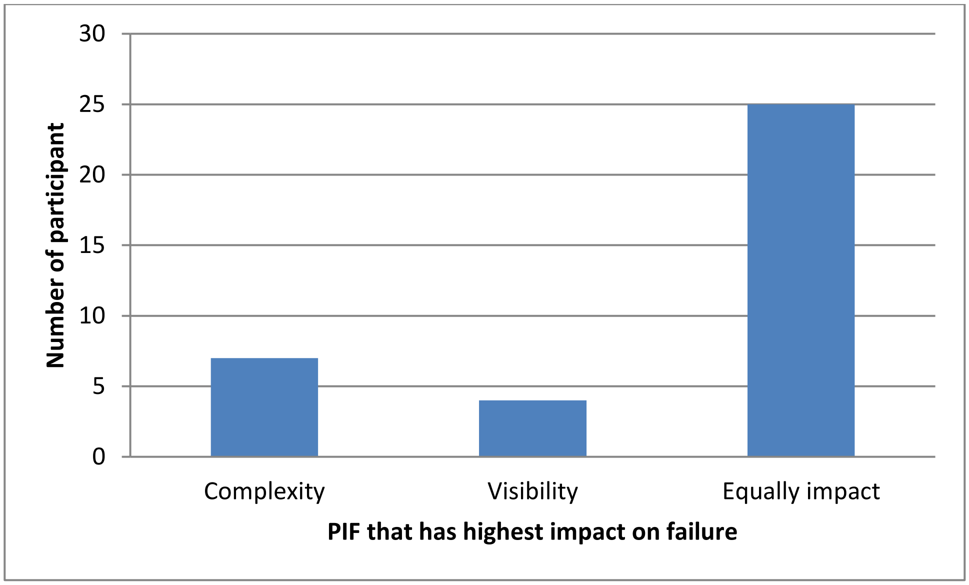


Figure 2.9: Sensitivity of 36 participants

In a comparison between G1 and G2, both of the PIFs were equally important for 70% of the participants in G1 and 68% of the participants in G2. Among the remaining participants in G1, 25% were more sensitive to complexity, and only 5% were more sensitive to visibility. In G2 16% of the remaining participants were found to be more sensitive to complexity, and 16% were more sensitive to visibility.

The results support the fact that sensitivity to PIFs may vary from participant to participant. Sensitivity analysis can be extremely helpful in personalizing training. For example, if a participant is found to be more sensitive to high complexity, training exercises with high complexity situations can be provided to better prepare for those situations and reduce the probability of failure. Such adaptive training will help to reach competency faster than with conventional training.

2.5 Limitations of the study

There are a few limitations with the current study that need to be considered. First of all, it has to be considered that virtual environments can provide a certain degree of realism and should not be expected to be an exact counterpart to real life emergency situations. Testing the validity of the achieved outcomes in a real world operational environment is out of the scope of this paper and is considered as a future research study. Secondly, since the work presented in this paper was done retrospectively, the experimental settings were not ideal for this particular work. For example, testing scenarios in each session were not randomized as they should have been for the purpose of this paper. Finally, to keep the experiment feasible in a laboratory setting, the effect of only two PIFs (complexity and visibility) were examined in the presented work. A more comprehensive set of PIFs will be used in future studies based on the context and associated priorities.

It must be noted that incorporation of individual differences presents new challenges in the conventional verification and validation paradigm. Since conditional probabilities in the BN can be different for each individual, quantification of parameterization confidence suggested in a conventional validity framework is nearly impossible (Pitchforth & Mengersen, 2013). However, BNs developed in this paper are integrated into human behavior models (HBMs) in later work (Musharraf et al., 2017). The authors are currently working on the validation of the HBM with the underlying belief that the uncertainty involved in the BN structure is negligible.

2.6 Conclusion

Reliability analysis of operators during offshore emergency conditions has always been a challenge due to the lack of data. The virtual experimental technique proposes to use virtual environments as a data source for the reliability analysis. This paper proposes an extension of the virtual experimental technique by incorporating individual differences. Performance data for each individual is first collected by conducting an experiment in a virtual environment. By integrating the collected data into a BN model, the reliability of any individual during an offshore emergency evacuation can be assessed. The model can also be used to perform a sensitivity analysis to determine if the individual is sensitive to any specific PIF. Though the case study presented in this paper suffers from a few limitations, it successfully demonstrates how individual differences can be taken into account while calculating human reliability. It also presents the way of identifying individuals' sensitivity to any external PIF. Future work involves using the results of the sensitivity analysis to help designing adaptive training for individuals. Adaptive training applied to virtual environments can help overcome any weakness an individual might have and assist them in achieving competence more quickly. Authors also plan to use a more comprehensive and informative list of PIFs in future studies. Though validation of the BN models presented in this paper will not be done separately, validation of the integrated HBM model is considered as a future work.

Acknowledgments

The authors acknowledge with gratitude the support of the Research & Development Corporation of Newfoundland and Labrador (RDCNL) and the NSERC-Husky Energy Industrial Research Chair in Safety at Sea.

References

Baraldi, P., Conti, M., Librizzi, M., Zio, E., Podofillini, L., & Dang, V. (2009). A Bayesian network model for dependence assessment in human reliability analysis. *Proceedings of the Annual European Safety and Reliability Conference (ESREL),* (pp. 223-230). Prague.

Blackman, H. S., Gertman, D. I., & Boring, R. L. (2008). Human Error Quantification Using Performance Shaping Factors in the SPAR-H Method. *52nd Annual Meeting of the Human Factors and Ergonomics Society. 52*, pp. 1733-1737. SAGE Publications.

Bye, A., Lois, E., Dang, V.N., Parry, G., Forester, J., Massaiu, S., Boring, R., Braarud, P.Ø., Broberg, H., Julius, J. and Männistö, I. (2011). *International HRA Empirical Study - Phase 2 Report.* Washington: U.S. Nuclear Regulatory Commission.

Cai, B., Liu, Y., Zhang, Y., Fan, Q., Liu, Z., & Tian, X. (2013). A dynamic Bayesian networks modeling of human factors on offshore blowouts. *Journal of Loss Prevention in the Process Industries, 26*(4), 639-649.

Cooper, S., Ramey-Smith, A., & Wreathall, J. (1996). *A Technique for Human Error Analysis (ATHEANA).* US Nulcear Regulatory Commission.

Dang, V., & Stempfel, Y. (2012). Evaluating the Bayesian belief network as a human reliability model - the effect of unreliable data. *Proceedings of the international conference on probabilistic safety assessment and management and the European safety and reliability conference PSAM 11 & ESREL 2012.* Helsinki, Finland.

Fenton, N., & Neil, M. (2012). *Risk assessment and decision analysis with Bayesian networks.* CRC Press.

Groth, K. M., & Mosleh, A. (2012a). A data-informed PIF hierarchy for model-based Human Reliability Analysis. *Reliability Engineering & System Safety, 108*, 154-174.

Groth, K. M., & Mosleh, A. (2012b). Deriving causal Bayesian networks from human reliability analysis data: A methodology and example model. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, 226*(4), 361-379.

Groth, K. M., & Swiler, L. P. (2013). Bridging the gap between HRA research and HRA practice: A Bayesian network version of SPAR-H. *Reliability Engineering & System Safety, 115,* 33-42.

House, A. W., Smith, J., MacKinnon, S., & Veitch, B. (2014). Interactive simulation for training offshore workers. *Oceans'14 MTS/IEEE Conference* (pp. 1-6). St. John's, NL: IEEE.

Kim, M. C., & Seong, P. H. (2006). An analytic model for situation assessment of nuclear power plant operators based on Bayesian inference. *Reliability Engineering & System Safety, 91*(3), 270-282.

Kirwan, B. (1994). *A Guide to practical human reliability assessment.* Taylor & Francis.

Lois, E., Dang, V.N., Forester, J., Broberg, H., Massaiu, S., Hildebrandt, M., Braarud, P., Parry, G., Julius, J., Boring, R. and Mannisto, I. (2009). *International HRA Empirical Study - Phase 1 Report.* Washington: U.S. Nuclear Regulatory Commission.

Martins, M. R., & Maturana, M. C. (2013). Application of Bayesian Belief networks to the human reliability analysis of an oil tanker operation focusing on collision accidents. *Reliability Engineering & System Safety, 110,* 89-109.

Mearns, K., Flin, R., Gordon, R., & Fleming, M. (2001). Human and organizational factors in offshore safety. *Work & Stress*, *15*(2), 144-160.

Mkrtchyan, L., Podofillini, L., & Dang, V. N. (2015). Bayesian belief networks for human reliability analysis: A review of applications and gaps. *Reliability engineering & system safety, 139*, 1-16.

Monferini, A., Konstandinidou, M., Nivolianitou, Z., Weber, S., Kontogiannis, T., Kafka, P., . . . Demichela, M. (2013). A compound methodology to assess the impact of human and organizational factors impact on the risk level of hazardous industrial plants. *Reliability Engineering & System Safety, 119*, 280-289.

Musharraf, M., Bradbury-Squires, D., Khan, F., Veitch, B., MacKinnon, S., & Imtiaz, S. (2014). A virtual experimental technique for data collection for a Bayesian network approach to human reliability analysis. *Reliability Engineering & System Safety, 132*, 1-8.

Musharraf, M., Khan, F., & Veitch, B. (2017). Modeling and simulation of personnel response during offshore emergency situations. *Proceedings of the 3rd Workshop and Symposium on Safety and Integrity management of operations in harsh environments (CRISE-3).*

Musharraf, M., Smith, J., Khan, F., Veitch, B., & MacKinnon, S. (2016). Assessing offshore emergency evacuation behavior in a virtual environment using a Bayesian Network approach. *Reliability Engineering & System Safety, 152,* 28-37.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference.* Morgan Kaufmann.

Pitchforth, J., & Mengersen, K. (2013). A proposed validation framework for expert elicited Bayesian Networks. *Expert Systems with Applications, 40*(1), 162-167.

Podofillini, L., & Dang, V. N. (2013). A Bayesian approach to treat expert-elicited probabilities in human reliability analysis model construction. *Reliability Engineering & System Safety, 117*, 52-64.

Smith, J. (2015). *The effect of virtual environment training on participant competence and learning in offshore emergency egress scenarios.* St. John's, NL: Faculty of Engineering and Applied Science, Memorial University on Newfoundland.

Sundaramurthi, R., & Smidts, C. (2013). Human reliability modeling for the Next Generation System Code. *Annals of Nuclear Energy, 52*, 137-156.

Swain, A. D., & Guttmann, H. E. (1983). *Handbook of Human Reliability Analysis with Emphasis on Nuclear Power*. Washington, US Nuclear Regulatory.

Veitch, B., Billard, R., & Patterson, A. (2008). Emergency Response Training Using Simulators. *Offshore Technology Conference.*

3. ASSESSING OFFSHORE EMERGENCY EVACUATION BEHAVIOR IN A VIRTUAL ENVIRONMENT USING A BAYESIAN NETWORK APPROACH

Mashrura Musharraf, Jennifer Smith, Faisal Khan**, Brian Veitch, Scott MacKinnon*
Faculty of Engineering & Applied Science,
* School of Human Kinetics and Recreation,
Memorial University of Newfoundland,
St John’s, Newfoundland and Labrador, Canada A1B 3X5
** Correspondence author: Tel: + 1 709 864 8939; Email: fikhan@mun.ca

Co-authorship statement

A version of this manuscript has been published in the Journal of Reliability Engineering & System Safety. Authors Mashrura Musharraf and Jennifer Smith designed the experiment, conducted the experiment, and performed necessary data collection. The lead author Mashrura Musharraf performed the literature review, developed the Bayesian network for assessing internal factors, performed the data analysis, and prepared the draft of the manuscript. Co-authors Faisal Khan, Brian Veitch, and Scott MacKinnon supervised the experimental study, reviewed and corrected the model and results. All co-authors reviewed and provided feedback on the manuscript. Mashrura Musharraf revised the manuscript based on the co-authors’ feedback and during the peer review process.

Abstract

In the performance influencing factor (PIF) hierarchy, person-based influencing factors reside in the top level along with machine-based, team-based, organization-based and situation/stressor-based factors. Though person-based PIFs like morale, motivation, and attitude (MMA) play an important role in shaping performance, it is nearly impossible to assess such PIFs directly. However, it is possible to measure behavioral indicators (e.g. compliance, use of information) that can provide insight regarding the state of the unobservable person-based PIFs. One common approach to measuring these indicators is to carry out a self-reported questionnaire survey. Significant work has been done to make such questionnaires reliable, but the potential validity problem associated with any questionnaire is that the data are subjective and thus may bear a limited relationship to reality. This paper describes the use of a virtual environment to measure behavioral indicators, which in turn can be used as proxies to assess otherwise unobservable PIFs like MMA. A Bayesian Network (BN) model is first developed to define the relationship between person-based PIFs and measurable behavioral indicators. The paper then shows how these indicators can be measured using evidence collected from a virtual environment of an offshore petroleum installation. A study that focused on emergency evacuation scenarios was done with 36 participants. The participants were first assessed using a multiple choice test. They were then assessed based on their observed performance during simulated offshore emergency evacuation conditions. A comparison of the two assessments demonstrates the potential benefits and challenges of using virtual environments to assess behavioral indicators, and thus the person-based PIFs.

3.1 Introduction

Since its introduction in 1960, more than a dozen Human Reliability Analysis (HRA) methods have been proposed to identify, model and quantify the probability of human errors. Most HRA methods involve the use of performance influencing factors (PIFs) to qualify and quantify human error probability (HEP). To ensure consistency across different HRA methods, Groth & Mosleh (2012) presented a standard set of PIFs and a PIF hierarchy. According to this hierarchy, all PIFs can be categorized in five categories: organization-based, team-based, person-based, situation/stressor-based, and machine-based. PIFs in these five categories can be observable, partially observable, or unobservable depending on how the states of the PIFs are assessed. If the state of a PIF can be assessed through direct measurement, then it is considered observable. *Tool availability* is an organization-based PIF that is directly observable: either the tool is available or it is not. An example of a partially observable PIF is the situation/stressor-based PIF *complexity*. The level of complexity depends on the perception of the individual and it cannot be directly measured. Complexity can be partially observed in terms of the number of assigned tasks at a given time: more tasks indicate at least nominally higher complexity. Finally, there are PIFs that are nearly impossible to measure and hence are called unobservable PIFs. Examples include person-based PIFs like *moral, motivation, and attitude (MMA)*. Though most of the HRA methods provide guidelines about how to assess the state of observable and partially observable PIFs, there is a lack of specific guidelines regarding assessing the state of unobservable PIFs. One possible solution is to associate the unobservable PIFs with specific indicators or metrics that are measurable and indicate the state of the unobservable PIFs. For example, a person’s attitude towards safety is unobservable, but it is possible to measure if the person complies with safety rules, so compliance is an indicator that can be used to assess safety attitude.

Studies have been done to measure these indicators by a subjective analysis (Rundmo et al., 1998; Rundmo, 2000; Adie et al., 2005). In these studies, a self-assessment questionnaire survey is conducted among personnel on offshore installations to gain insight into unobservable PIFs like safety attitude. Though significant work has been done to make self-assessment questionnaires reliable (Mearns & Flin, 1995; Flin et al., 2000), it is still questionable if self-assessment is a true reflection of the way an operator will behave in real emergencies. Questionnaires have the inherent risk of representing one’s knowledge about safety and/or one’s willingness to behave safely, rather than representing one’s actual behavior in emergency situations (Breitsprecher et al., 2007). This paper describes the use of a virtual environment to measure indicators that can provide insight into unobservable PIFs like MMA. A Bayesian network (BN) was developed to define relationships among unobservable PIFs and associated indicators. The network was then extended by associating the measurable indicators with evidence that can be collected using a virtual environment. An experimental study of offshore emergency evacuation in a virtual environment was done with 36 human subjects. Behavioral indicators of the participants were assessed using both a multiple choice test and the performance evidence collected from the virtual environment. A comparison of the two approaches demonstrates the potential benefits and challenges of using virtual environments to assess behavioral indicators.

Section 3.2 gives an overview of the virtual environment used in the paper and explains the fundamentals of BN. Section 3.3 describes a BN approach to quantify unobservable PIFs. Sections 3.4 and 3.5 demonstrate the application of the proposed approach to a case study of offshore emergency evacuation. Results are presented in Section 3.6. Section 3.7 lists the limitations of the study. Section 3.8 summarizes and concludes the paper.

3.2 Background

3.2.1 Overview of virtual environment

A virtual environment is a computer aided simulation environment that allows trainees to gain artificial experience, including performing in dangerous scenarios. Virtual environment training can act as an enhancement to conventional training since training for emergency situations in the real world is ethically, logistically or financially unfeasible (Veitch et al., 2008). Besides facilitating emergency preparedness training, virtual environments can also be used as a tool to observe human performance in emergency conditions (Lois et al., 2009; Bye et al., 2011; Monferini et al., 2013). The virtual environment used in the case study is called the all-hands virtual emergency response trainer (AVERT) and was developed at Memorial University. AVERT was designed to enhance offshore emergency response training. The virtual environment is modeled after an offshore oil installation platform with high levels of detail. It is capable of creating credible emergency scenarios by introducing hazards such as blackouts, fires and explosions. For the case study, the offshore emergency scenarios covered a range of activities, from muster drills that required the participant to go to their primary muster station, to more complex emergency evacuation scenarios that required the participant to avoid hazards blocking their egress routes and muster at their lifeboat stations (House et al., 2014).

3.2.2 Bayesian network fundamentals

BNs are probabilistic models representing interaction of parameters through directed acyclic graph and Conditional Probability Tables (CPTs) (Pearl, 1988). The networks are composed of nodes and links. Nodes represent the variables of interest whereas links joining the nodes represent causal relations among the variables. Nodes and links together define the qualitative part of the network. The quantitative part is constituted by the conditional probabilities associated with the variables. Conditional probabilities specify the probability of each dependent variable (also called child node) for every possible combination of the states of the variables it is directly dependent on (also called parent node). The probabilities of the independent variables, i.e., nodes with no predecessor (also called root nodes) are also given. Given the probabilities associated with each root node and the conditional probability table associated with each child node, the probabilities of child node can be calculated (Fenton & Neil, 2012). If there are n variables $X_1, X_2, \dots, X_n$ in the network and $Pa(X_i)$ represents the set of parents of each X_i, then the joint probability distribution for the entire network can be defined as:

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^{n} P\big(X_i \big| Pa(X_i)\big) \tag{3.1}$$

where $P\big(X_i \big| Pa(X_i)\big)$ is the discrete conditional probability distributions of X_i given its parents.

Therefore, the following need to be specified to define a BN:

1) the set of variables (nodes): $X_1, X_2, \dots, X_n$,
2) the interaction (links) between variables, and
3) the conditional probability distribution $P\big(X_i \big| Pa(X_i)\big)$ for each variable X_i.

This paper presents a BN model to quantify unobservable PIFs. Section 3.3 illustrates how the BN model is defined.

3.3 Quantifying unobservable PIFs: A Bayesian network (BN) approach

This section presents the BN model to quantify unobservable PIFs. First, a set of necessary variables is defined. Having defined the variables, the relationship between variables (both links and conditional dependency) are specified.

3.3.1 Variables

Two types of variables compose the proposed BN model: variables to measure the unobservable PIFs, and variables to collect evidence.

3.3.1.1 Variables to measure unobservable PIFs

Unobservable PIFs are impossible to measure directly. There are several PIFs in the standard set that are unobservable, but the focus of this paper is on the person-based PIFs *bias* and *MMA*. Both *bias* and *MMA* are internal characteristics of an individual and are not directly observable, but internal characteristics of individuals manifest themselves in the way they behave, and behaviors are observable (Groth & Mosleh, 2012). Hence the unobservable PIFs have been associated with measurable behavioral indicators in this paper.

Groth et al. (2012) define bias as "the tendency of a human to make conclusions based on selected pieces of information while excluding information that does not agree with the conclusion." It is impossible to directly measure if an individual has a bias and the degree to which the bias is present. In this paper, *bias* is associated with behavioral indicators *inclination to previous experience* and *information use,* which are measurable and can help to define the state of *bias* at a given time. It can be tested if in any given situation an individual disregards valuable information in order to come to a conclusion that has worked well for him/her on previous occasions. Thus, *inclination to previous experience* and *information use* are indicators of bias. An expanded list of biases and mechanisms can be found in (Brewer, 2005).

Morale, motivation, and attitude together refer to the "willingness to complete tasks, the amount of effort a person devotes to tasks, and the state of mind of the worker" (Steers & Porter, 1979; Triandis, 1971). MMA plays a significant role in shaping the performance of an individual, but it is extremely difficult to measure. There are measurable behavioral indicators that are associated with MMA: *information use*, *prioritization*, and *compliance*. The *information use* behavior measures an individual's effectiveness in using information presented to him/her. Individuals may favor some information over others due to bias. *Prioritization* is how an individual orders the tasks assigned to them, or the goals that are to be achieved. *Compliance* refers to an individual's commitment to follow directions and policies established by the organization or the industry. *Information use, prioritization*, and *compliance* are behaviors shaped by the MMA of an individual. Additional behaviors can be included depending on the context.

Table 3.1 lists the unobservable PIF variables and associated behavior indicator variables used in this paper.

Table 3.1: List of unobserved PIF variables and associated behavior indicator variables

Unobservable PIF variables	***Associated behavior indicator variables***
Bias	Inclination to previous experience Information use
MMA	Information use Compliance Prioritization

All unobservable PIF variables and associated behavior indicator variables are considered to be binary in this paper (i.e. the participant is either compliant or not).

3.3.1.2 Variables to collect evidence

Once behavioral indicators associated with each unobservable PIF are identified, the next step is to measure these indicators. For this purpose, each behavioral indicator variable is associated with evidential variables that are used to collect information relevant to behavioral indicators. One possible source of evidence is multiple choice test items in a self-assessment questionnaire (Rundmo, 2000). Using this source, behavioral indicators are associated with multiple choice test items and are assessed based on the answers given by the participant. The fundamental problem of using multiple choice test items as evidence is that the answers individuals choose in the questionnaire often represent their knowledge about the safety regulations and/or their desire to act safely, rather than how they will actually behave under high risk, time pressure, and complexity of emergency conditions. Another problem with multiple choice questionnaires is the high guessing factor, which represents the possibility that an individual will guess the right answer to a question by chance even when he/she does not know the answer.

In addition to multiple choice questionnaires, this paper uses the performance of participants in a virtual environment to collect evidence regarding behavioral indicators. The behavioral indicators to be assessed are associated with different tasks and exercises that an individual will perform in the virtual emergency scenarios. The indicators can then be measured based on how the individual performs the assigned tasks. Unlike questionnaires, a virtual environment is capable of simulating the dynamism and urgency of emergency scenarios and is expected to be a closer representation of an individual's performance in real life emergency. As performing in a virtual environment scenario is an open ended problem, the guessing factor is much lower than in the multiple choice questionnaire.

Thus, evidence variables of the proposed BN model are either multiple choice items (in case of self-assessment multiple choice questionnaire), or tasks/exercises (in case of virtual environment scenarios). They are considered to be binary in both cases (i.e. the participant either answers a question correctly or not).

3.3.2 Relationships between variables

Once variables are defined, the next step is to define the relationship between variables. This requires that both the links and parameters be specified for each relationship.

3.3.2.1 Relationships between unobservable PIFs and behavioral indicators

As discussed in Section 3.3.1.1, unobservable internal characteristics like bias and MMA have a causal influence on the way an individual behaves. Adding these dependencies between unobservable PIFs and behavioral indicators gives a BN shown in Figure 3.1.

The parameters of the network shown in Figure 3.1 are: 1) the prior belief (in terms of probabilities) of the unobservable PIF variables $P(UV)$, and 2) conditional belief (in terms of probability distribution) of indicator variables $P(IV|UV_i,\ i = 1,2,\dots,n)$. The prior probabilities of the possible states of each unobservable PIF are assumed to be equal (50%). The conditional probabilities are approximated by a canonical interaction model: the binary Noisy-OR gate (Pearl, 1988). Two assumptions of the Noisy-OR model are: 1) each of the unobservable PIFs is sufficient to shape a behavior with a probability of p_i in the absence of all other causes, and 2) the ability of each unobservable PIF being sufficient is independent of the presence of other causes. If p_i represents the probability that a behavior is formed by the unobservable PIF UV_i when all other causes $UV_j, j \neq i,$ are absent, then the conditional probability distribution of the behavioral indicator variables can be defined as:

$$P(IV = Positive|\{UV_1, UV_2, \dots, UV_n\}) = 1 - \prod_{i \in S}(1 - p_i) \qquad (3.2)$$

where S is a subset of the $UV_i s$ that are present.

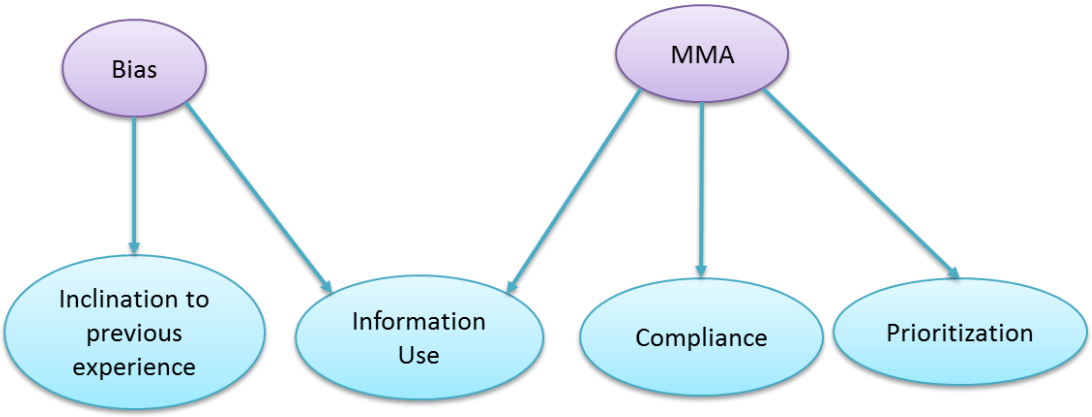


Figure 3.1: Causal dependency between the unobservable PIFs and the associated behavioral indicators

3.3.2.2 Relationships between behavioral indicators and evidential variables

Relationships between behavioral indicators and evidential variables are based on the causality that behaviors have an influence on how a question will be answered or a situation will be solved by an individual (Millán & Pérez-De-La-Cruz, 2002). Figure 3.2 illustrates the causal dependency between the behavioral indicators and collected evidence (EV_1 to EV_n). The probability assignment of the behavioral indicator variables $P(IV)$ is already described in Section 3.3.2.1. The additional parameter for the network shown in Figure 3.2 is the conditional belief (in terms of probability distribution) of the evidential nodes $P(EV|IV_i,\ i = 1,2, \dots, n)$. To approximate the conditional probability, again the binary Noisy-OR gate is used. If p_i represents the probability that a behavior IV_i will influence an individual to choose a correct answer to a question, or take a correct action in the virtual scenarios when all other causes $IV_j, j \neq i,$ are absent, then the conditional probability distribution of the evidential variables can be defined as:

$$P(EV = Right|\{IV_1, IV_2, \dots, IV_n\}) = 1 - \prod_{i \in S}(1 - p_i) \qquad (3.3)$$

where S is a subset of the $IV_i s$ that are present.

Combining the causal dependencies shown in Figures 3.1 and 3.2, a complete causal model can be developed as shown in Figure 3.3. Using this model, we can infer what we cannot see (unobservable PIF variables) from what we can see (evidence variables).

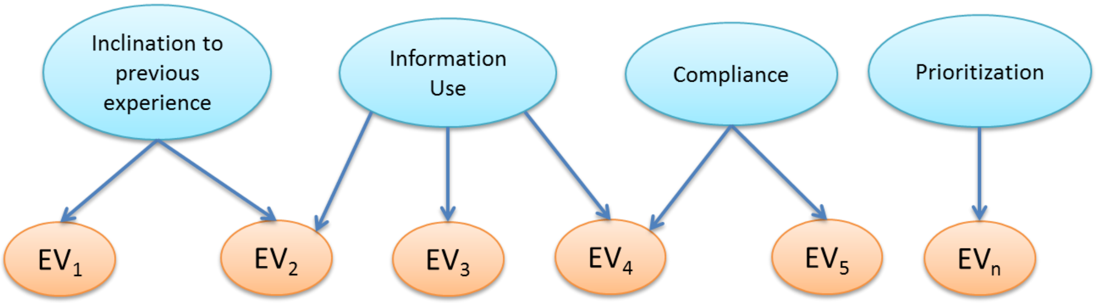


Figure 3.2: Causal dependency model of the behavioral indicators and collected evidence. EV_1 to EV_n represent collected evidence: either a multiple choice test item or a task in a virtual environment scenario. The causal dependency is the same in both cases.

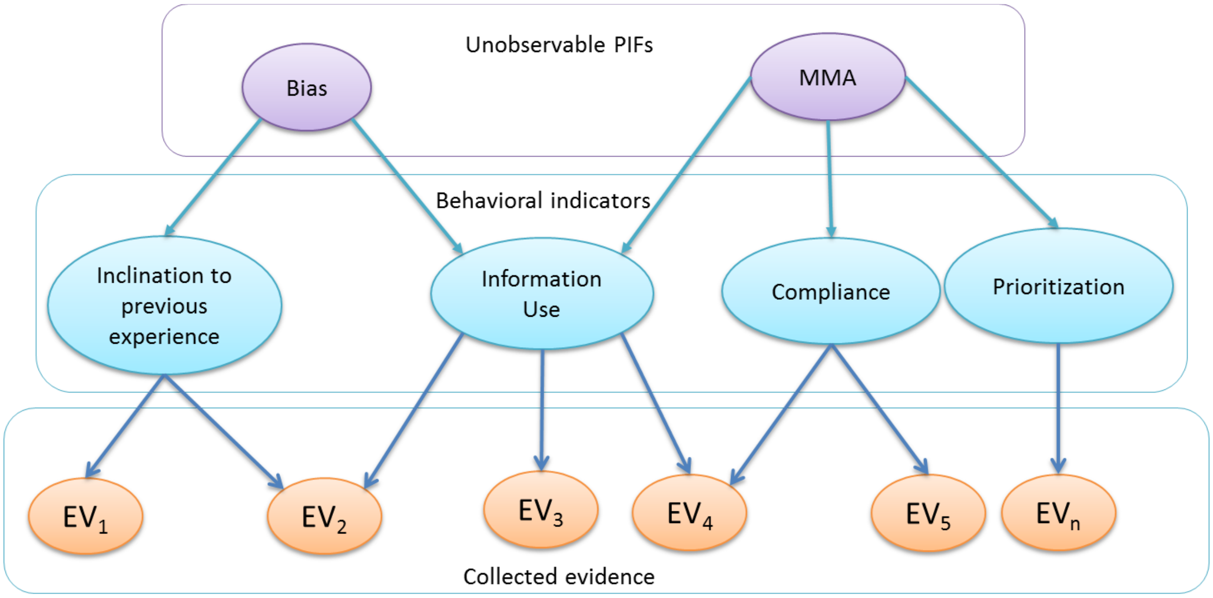


Figure 3.3: Causal dependencies among unobservable PIFs, behavioral indicators and collected evidence

3.4 Case study: Offshore emergency evacuation

This paper assesses the behavioral indicators of individuals by using both a multiple choice test and virtual environment scenarios. As stated in Section 3.2.1, a virtual environment called AVERT was used in this study. An experimental study was designed using AVERT with multiple research objectives: 1) assess competency in offshore emergency evacuation using virtual environments (Smith et al., 2015), 2) collect data for human reliability assessment using virtual environments, and 3) assess behavioral indicators of individuals' during offshore emergency evacuation using virtual environments. The focus of this paper is to demonstrate the use of virtual environments to assess behavioral indicators and in turn assess unobservable PIFs.

A total of 36 participants took part in the study with a goal to learn how to perform a successful offshore emergency evacuation. The participants were naïve concerning any detail of the experimental design, they were not employed in the offshore oil and gas industry, and therefore they were not familiar with the offshore platform. Each participant was assigned to one of two groups: 1) G1: high level training and 2) G2: low level training. Participants in both groups attended 3 sessions. The content of each session was different between the two groups. In the first session, both groups received a basic offshore emergency preparedness tutorial. G1 then received 4 training scenarios, a multiple choice test and 4 testing scenarios. G2 only received the multiple choice test and 4 testing scenarios after the tutorial. In both Session 2 and Session 3, G1 received an advanced training tutorial about alarms and hazards respectively, 4 additional training scenarios, a multiple choice test, and 4 testing scenarios. G2 received no advanced training tutorial and only received a multiple choice test and 4 testing scenarios in Sessions 2 and 3. Both groups were provided with feedback on their performance in the multiple choice test and virtual environment testing scenarios in each session. Figures 3.4 and 3.5 summarize the design of the experiment.

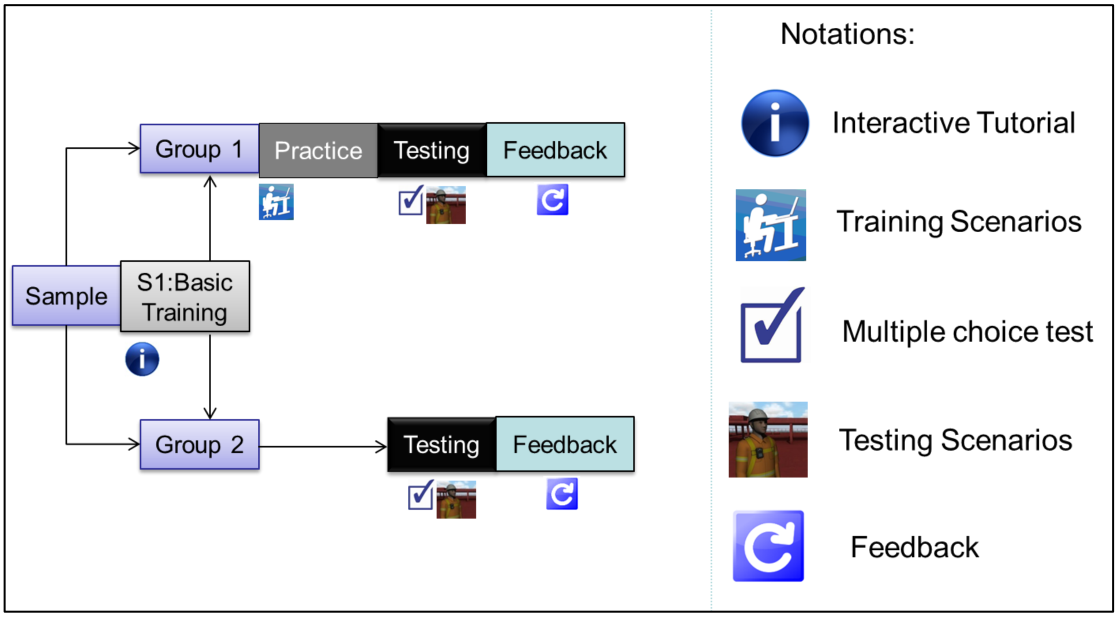


Figure 3.4: Experimental design of Session 1

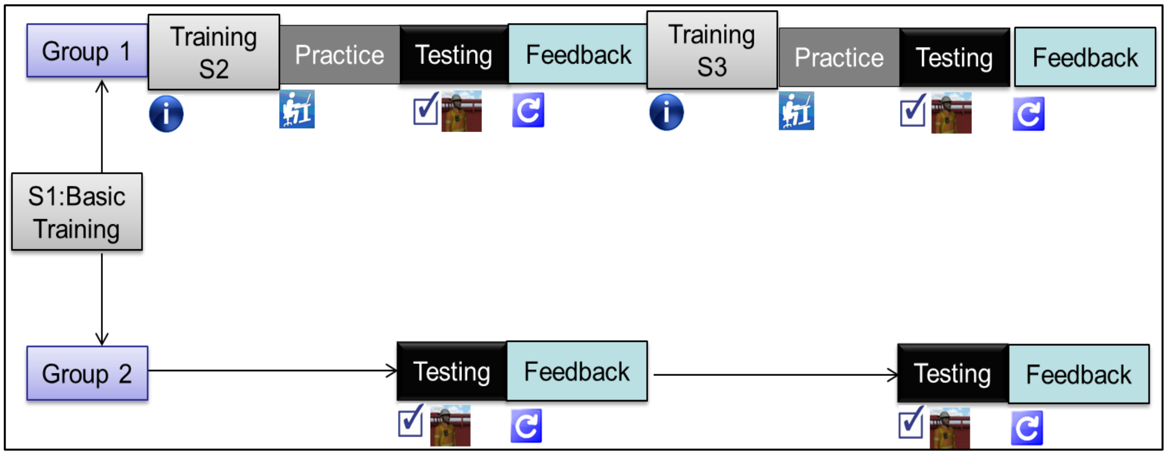


Figure 3.5: Experimental design of Session 2 & 3

The training and testing scenarios were designed with varying levels of visibility (clearly visible or blackout conditions) and complexity (low complexity with no obstacles on the primary evacuation route, high complexity with obstacles on the escape route and increased responsibility). Several performance metrics of the participants were recorded during each scenario. The following are the performance metrics that are most relevant to this case study: route selected for evacuation, time spent running, interaction with fire doors and watertight doors, interaction with hazards, reporting at muster stations, and interaction with manual alarm. Replay videos of participants' performance in scenarios were also recorded for further analysis. Performance and behavior of the participants were assessed in the multiple choice test and virtual environment testing scenarios. As stated above, there was only 1 multiple choice test and 4 virtual environment testing scenarios in each of the sessions. For demonstration purposes, only the multiple choice test and virtual environment testing scenarios of the last session (Session 3) have been included in this paper. Table 3.2 shows how the different questions (EV) in the multiple choice test for Session 3 were used to assess behavioral indicators (IV). Each question listed in Table 3.2 had multiple options to choose from. Table 3.3 gives an overview of the 4 testing scenarios and shows how evidence (EV) regarding behavioral indicators (IV) was collected in these scenarios.

Table 3.2: Multiple choice questions in Session 3 used to assess the behavioral indicators

Question number	***Question to collect evidence (EV)***	***Behavioral indicators assessed (IV)***
S3_Q2	The Station Bill provides what information?	Information Use
S3_Q3	What do you do in the event that your primary muster station is compromised?	Compliance
S3_Q4	If you can't remember how to get to your muster station what should you do?	Information Use
S3_Q6	What do you do in the event of a minor incident?	Prioritization
S3_Q7	What would you do in the event of an alarm that wasn't followed by a PA announcement?	Compliance
S3_Q8	What is the safest exit to take given where the hazard is located? [A diagram of the situation was given that depicted an explosion and fire in the engine room, blocking access to the secondary and tertiary egress routes.]	Compliance
S3_Q14	What do you do when your primary muster route has been blocked?	Compliance
S3_Q17	What is the safest exit to take given where the hazard is located? [A diagram of the situation was given that depicted a hallway filling with smoke outside the cabin, blocking access to the primary egress route.]	Compliance
S3_Q18	What is the safest exit to take given where the hazard is located? [A diagram of the situation was given that depicted fire and smoke in the engine room, blocking access to the primary egress route.]	Compliance

Table 3.3: Overview of the virtual environment testing scenarios in Session 3 used to assess behavioral indicators

Scenario name	***Context***	***Task to collect evidence (EV)***	***Behavioral indicators assessed (IV)***
S3_Scn1	Fire erupts in the gally signaling a General Platform Alarm (GPA). The participant must go to the muster station but re-route to the lifeboat station due to the fire and smoke spreading to the adjacent muster station.	Follow PA announcement	Compliance, Information use
		Follow alarm	Compliance, Information use
		Avoid running in the platform	Compliance
		Keep fire doors and watertight doors closed	Compliance, Information use
		Avoid interaction with hazard	Compliance
		Review feedback carefully and learn about correct muster station so that mistake is not repeated	Information use
		Avoid previously explored route if not safe	Inclination to previous experience
S3_Scn2	A fire and explosion on the helideck signal a GPA. High winds cause the smoke to engulf a portion of the platform exterior. The participant must go to muster station but re-route to lifeboat station due to the increase in emergency severity and the alarm change to Prepare to Abandon Platform (PAPA).	Same as S3_Scn1	Same as S3_Scn1
S3_Scn3	An electrical fire and dense smoke fill a portion of the engine room, blocking access to the primary egress route. The GPA sounds. The participant must go to muster station but re-route to lifeboat station due to the increase in situation severity and alarm change to PAPA.	Same as S3_Scn1	Same as S3_Scn1

Scenario name	***Context***	***Task to collect evidence (EV)***	***Behavioral indicators assessed (IV)***
S3_Scn4	A fire and explosion in the main engine result in a vessel-wide blackout. The alarm is not immediately triggered. The fire blocks access to the secondary egress routes. The participant must raise the alarm and go to the muster station but re-route to lifeboat station due to the increase in situation severity and alarm change to PAPA.	Same as S3_Scn1 with the added following task: Raise the alarm before evacuating if first observer of the hazard	Same as S3_Scn1 with the added following behavioral indicator: Prioritization

Association of evidence variables (EV) with behavioral indicators (IV) shown in Tables 3.2 & 3.3 are defined by the analyst based on the understanding of the context. The next section shows how evidence collected from the multiple choice tests and the virtual environment scenarios were integrated in the BN developed in Section 3.3 to assess behavioral indicators and unobservable PIFs.

3.5 Integrating evidence in Bayesian network (BN)

3.5.1 Integrating evidence collected using multiple choice test

Test questions were the evidence variables in the case of the multiple choice test. Integration of the evidence variables (as shown in Table 3.2) in the BN proposed in Section 3.3 provided the final network shown in Figure 3.6. As stated in Section 3.3, the prior probabilities of each possible state of the unobservable PIF variables were assumed to be equal (50%). The conditional probability distributions of indicator variables and evidence variables were calculated using equations 3.1 and 3.2, respectively. It has to be noted that

p_is used in equations 3.1 and 3.2 are defined by the analyst and are not dependent on the collected data. Values of p_is are context sensitive and may need to be redefined for a given situation. As discussed in Section 3.3, an individual can guess the right answer to a question by chance even when he/she does not know the answer. To address this issue, the conditional probability $P(EV = Right|\{IV_1, IV_2, \dots, IV_n\})$ is considered to be equal to a guess factor when all the IV_is are absent. If a multiple choice question had n possible options to choose from, the guess factor would be considered to be $1/n$ (Millán et al., 2013).

The state of the evidence variables was defined based on the answers the participants chose in the multiple choice test. Figure 3.7 presents the state of the network after all the questions in the multiple choice test were answered (using one participant's results as an example).

3.5.2 Integrating evidence collected using virtual environment testing scenarios

The tasks performed in the virtual environment were the evidence variables in the virtual testing scenarios. Integration of the evidence variables in the testing scenarios (as shown in Table 3.3) in the BN proposed in Section 3.3 yielded the network shown in Figure 3.8. Similar to the multiple choice test, the prior probabilities of each possible state of the unobservable PIF variables were assumed to be equal (50%) and the conditional probability distributions of indicator variables and evidence variables were calculated using equations 3.2 and 3.3 respectively. Again the p_is are defined by the analyst and may need to be redefined for a different context. However, unlike the multiple choice test, the guess factor was considered to be as low as 5% because of the fact that the test scenarios are open ended problems and the chances that an individual will make a correct decision or take a correct action by chance, without knowing, was considered to be low.

The state of the evidence variables was defined based on the decisions and actions of the participants in the virtual environment testing scenario. Figure 3.9 shows the state of the network as an example after one participant finished the last scenario in Session 3 (S3_Scn4).

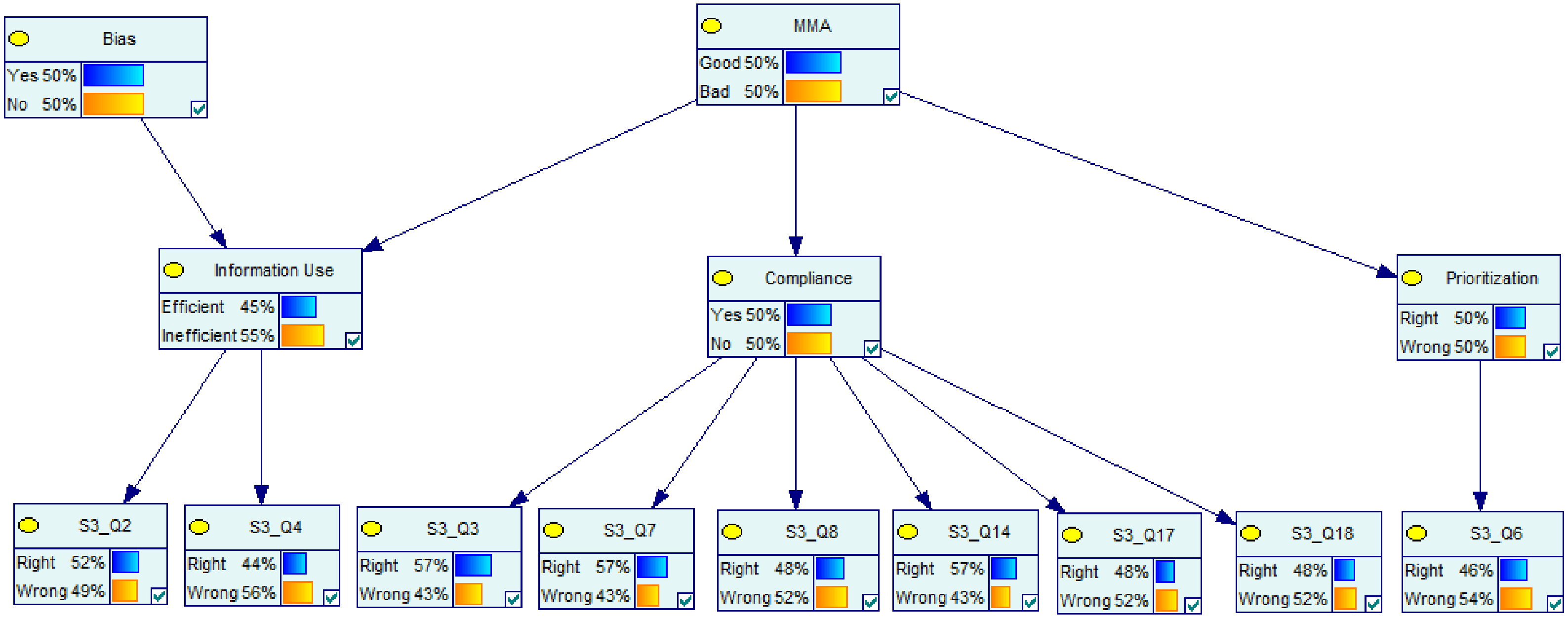


Figure 3.6: BN model to assess behavioral indicators and unobservable PIFs using multiple choice test. From bottom to top, the first level contains the evidence variables, the second level contains indicator variables and the third level contains unobservable PIF variables.

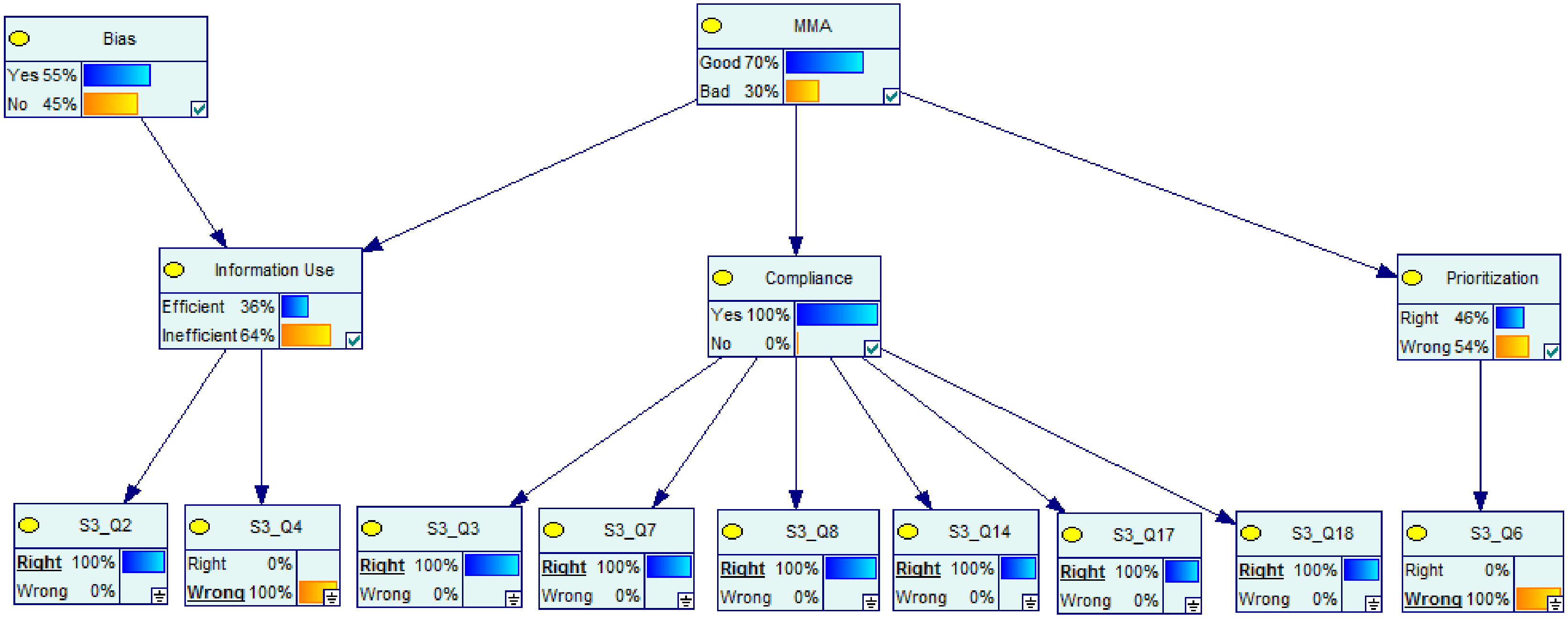


Figure 3.7: State of the network after all questions in the multiple choice test in Session 3 have been answered by a participant.

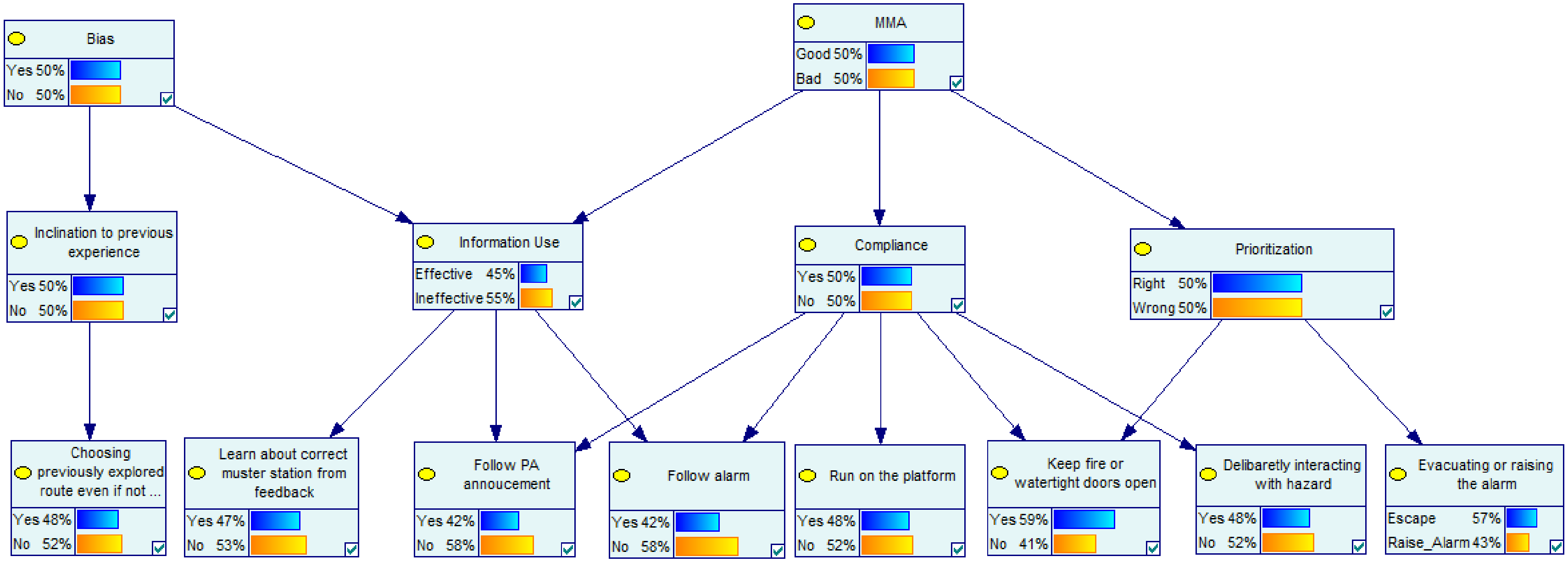


Figure 3.8: BN model to assess behavioral indicators and unobservable PIFs using virtual environment test scenario. From bottom to top, the first level contains the evidence variables, the second level contains indicator variables and the third level contains unobservable PIF variables.

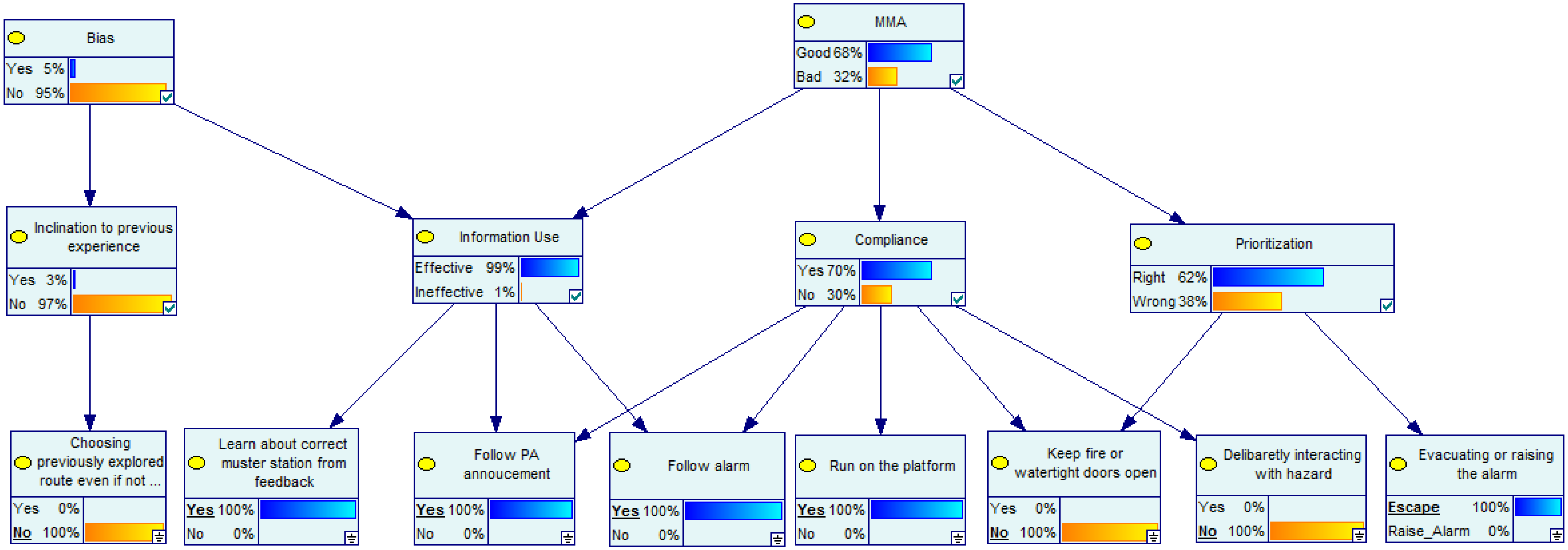


Figure 3.9: State of the network after a participant finishes the last scenario in Session 3 - S3_Scn4.

3.6 Result and discussion

Figures 3.7 and 3.9 show the scores of the same sample participant for the multiple choice test and in a virtual environment testing scenario, respectively. The performance of all 36 participants was assessed using multiple choice tests and virtual environment testing scenarios for all 3 sessions, and scores were calculated similarly. Table 3.4 shows a comparison between average scores of all participants for the multiple choice test in Session 3, and in the most complex testing scenario in Session 3 (denoted as S3_Scn4). For both *compliance* and *prioritization,* the average score for the testing scenario S3_Scn4 was below the multiple choice test (57% vs. 97% for compliance, and 57% vs. 94% for prioritization). However, participants were able to *use information* more effectively in the virtual environment scenario than the multiple choice test (average score 90% vs. 80% for information use). The differences in the scores shows that the participants behaved differently than anticipated based on the multiple choice test when put in a highly complex virtual emergency situation. Many of the participants who were able to successfully answer multiple choice questions related to *prioritization* and *compliance* were unable to demonstrate these in the virtual environment testing scenario. This is an indication that the multiple choice test can be used to diagnose an individual's knowledge about safety regulations and/or their willingness to behave safely. However, the sole use of a multiple choice test cannot predict if the individuals will be able to put their knowledge and willingness into practice.

Table 3.4: Comparison between average scores in the multiple choice test and virtual environment testing scenario S3_Scn4 in Session 3

Behavioral indicators	*Multiple choice test (%)*	*Testing scenario (%)*
Information Use (Effective)	80	90
Compliance (Yes)	97	57
Prioritization (Right)	94	57
MMA (Good)	96	61

As stated in Section 3.4, participants were divided into two groups: 1) G1: high level training or 2) G2: low level training. Table 3.5 shows a comparison between the average scores of the two groups. For *compliance* and *prioritization*, both of the groups performed better in the multiple choice test than in the virtual environment testing scenario. Both groups demonstrated better use of information in the testing scenario compared to the multiple choice test. In the multiple choice test, G1 demonstrated superior performance in terms of *information use* and *compliance* (86% vs. 75% for information use, and 99% vs. 95% for compliance). G1 & G2 showed similar performance in prioritizing tasks (94% vs. 95% for task prioritization), in the multiple choice test. Similar results were found for the virtual environment testing scenario. G1 demonstrated better *information using* capabilities and *compliance* compared to G2 (93% vs 87% for information use, and 62% vs. 53% for compliance) in the virtual environment testing scenario. G1 showed slightly better performance in prioritizing for the virtual environment test scenario (58% vs. 56% for task prioritization).

Table 3.5: Comparison between average scores of G1 and G2 in multiple choice test and virtual environment testing scenario S3_Scn4 in Session 3

Behavioral indicators	***Multiple choice test G1 (%)***	***Multiple choice test G2 (%)***	***Testing scenario G1 (%)***	***Testing scenario G2 (%)***
Information Use (Effective)	86	75	93	87
Compliance (Yes)	99	95	62	53
Prioritization (Right)	94	95	58	56
MMA (Good)	97	95	64	58

Figures 3.10 to 3.12 show a one to one comparison of the multiple choice test scores and the virtual environment testing scenario scores for all participants for *information use, compliance,* and *prioritization,* respectively. The percentages of participants (in the total sample size) who achieved good scores in both the multiple choice test and the virtual environment testing scenario are 69%, 28% and 14% for *information use, compliance* and *prioritization*, respectively. There was a significant difference between the scores of the multiple choice test and the virtual environment testing scenario for the remaining participants. For *compliance,* 69% of the participants achieved a good score in the multiple choice test but failed to demonstrate so in the virtual environment testing scenario. For *prioritization,* this percentage was 78%. For *information use,* the result was quite the contrary. Only 6% of participants achieved a better score in the multiple choice test than in the virtual environment testing scenario. However, 22% of the participants achieved a better score for *information use* in the virtual environment testing scenario than in the multiple choice test.

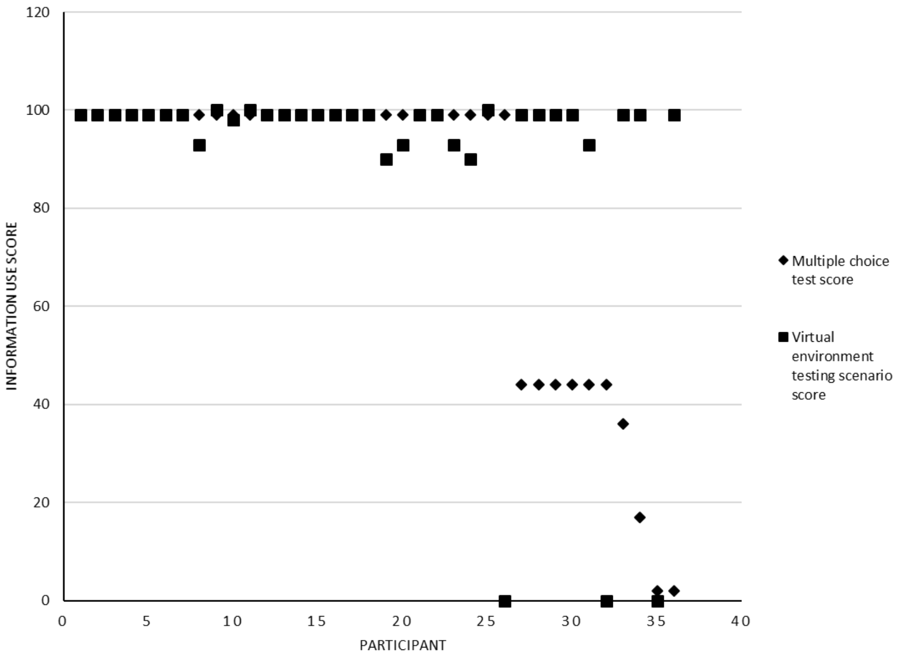


Figure 3.10: One to one comparison of each participant's virtual environment testing scenario score and their multiple choice test score for the behavioral indicator *information use*

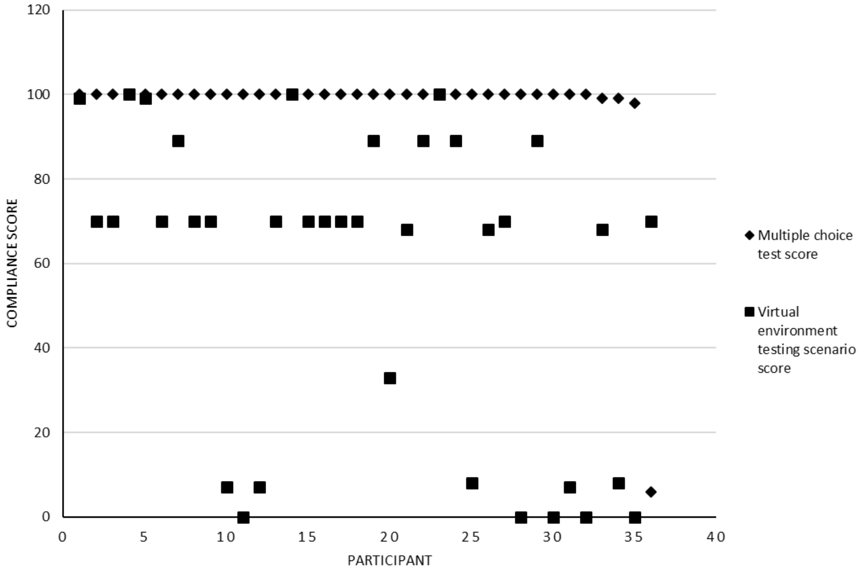


Figure 3.11: One to one comparison of each participant's virtual environment testing scenario score and their multiple choice test score for the behavioral indicator *compliance*

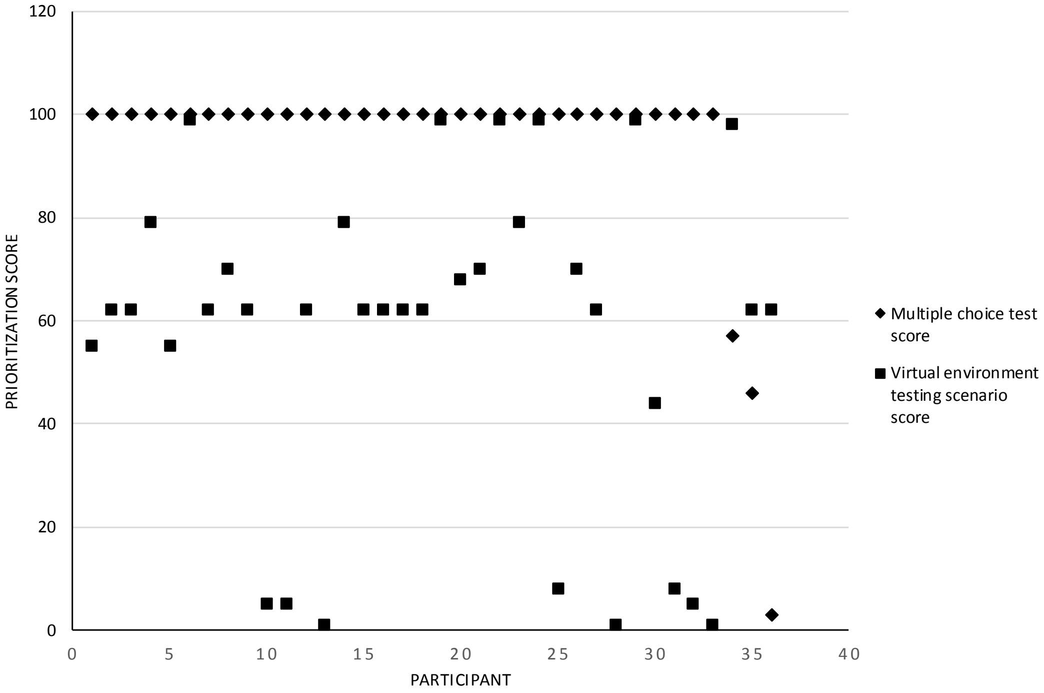


Figure 3.12: One to one comparison of each participant's virtual environment testing scenario score and their multiple choice test score for the behavioral indicator *prioritization*

The fact that a substantial number of participants who achieved a high score in the multiple choice test failed to demonstrate so in the virtual environment testing scenario indicates that a multiple choice test can be used to assess participants' knowledge and/or their willingness to follow instructions, but that it cannot assess whether the participants will be able to apply the knowledge and willingness in emergency conditions. Similarly, a participant who received a good score in the virtual environment testing scenario and a poor score in the multiple choice test may have a sound judgement in a particular situation but may lack the knowledge and/or willingness. Rather than using a multiple choice test or a virtual environment test as a standalone, a combination of the two techniques can provide a better understanding of individuals' behavior in emergency conditions and help ensure they are better prepared for emergency situations. The multiple choice test can be used initially to provide an assessment of individuals' knowledge about safety regulations and their willingness to behave safely. The virtual environment testing scenarios can be used next to assess if individuals are able to apply their knowledge and willingness into practice. For example, individuals achieving good scores in *compliance* for the multiple choice test are believed to have sufficient knowledge about the rules and regulations. A virtual environment can then be used to assess how compliant these individuals are in following the rules and regulations in emergency conditions. Once the other behavioral indicators are assessed in the same way, unobservable PIFs can be assessed using the behavioral indicators.

3.7 Limitation of the study

There are a few limitations of the study that should be noted. First of all, it has to be considered that the virtual environment can provide a certain degree of realism and should not be expected to be an exact counterpart of real life emergency situations. Secondly, the evidence (i.e. questions or tasks) was not evenly distributed across behavioral indicators and this may have biased some of the results. For example, only one question (S3_Q6) was used to assess *prioritization* in the multiple choice test, whereas there were six questions to assess *compliance*. So, the assessment of *prioritization* may not be as robust as it should be. Having sufficient evidence for each indicator will increase the accuracy of the assessment. Thirdly, the study was designed to achieve three different research objectives and as a result there were a few constraints that needed to be maintained. The management or controlling of constraints in the study may have conditioned the performance of the participants in particular ways. An example could be the order of the multiple choice test and the virtual environment testing scenarios in each session. The participants had to complete the multiple choice test in advance of the testing scenarios. This order of testing meant that the feedback provided after multiple choice test may have influenced the score of the virtual environment testing scenario, which was not considered in the experimental design. An in-depth analysis was not performed to determine the effect of an individual's prior video gaming experience on their testing scenario performance. The groups were balanced in terms of their self-reported video gaming experience, but it was not considered if the video gaming experience helped individuals get a better score in the virtual environment testing scenarios. Finally, the limited sample size might have imposed some constraints on the results and should be taken into consideration in future studies.

3.8 Conclusion

Though unobservable PIFs play an important role in shaping human performance, they are nearly impossible to measure. However, behavioral indicators associated with the unobservable PIFs are measurable and can help define the state of the unobservable PIFs. In this paper, a BN was first developed using the causal relationship between behavioral indicators and unobservable PIFs. The network was then extended by connecting each behavioral indicator with evidence variables. Conventional approaches for collecting this form of evidence involve using self-assessment multiple choice questionnaires. However, as reliable as these questionnaires may be, there is always the risk that the collected responses will only represent an individual's knowledge and/or willingness to behave safely, instead of their actual behavior in emergency situations. This paper proposes the use of virtual environments along with the questionnaires to overcome this problem.

Evidence was collected using both multiple choice tests and virtual environment testing scenarios in the experimental study presented in this paper. The comparison of outcomes of the two tests shows the difference between the individuals' expected behavior and their actual behavior when placed in an emergency situation. A substantial number of participants who achieved a high score in the multiple choice test failed to demonstrate so in the virtual environment testing scenario. On the other hand, some participants who did not do well in the multiple choice test managed to demonstrate acceptable performance in the virtual environment testing scenarios. A combined testing approach including both multiple choice test and virtual environment test can help to ensure that participants have the required knowledge and the skill to apply the knowledge in emergency situations. A comparison between the two groups (G1: high level training and G2: low level training) shows that the highly trained group performed either better or equal in both multiple choice test and virtual environment testing scenarios. This indicates the benefit of advanced training in improving participants' performance.

Virtual environment scenarios are a closer representation of real life emergency situations when compared to a multiple choice test and can provide a better understanding of individuals' behavior. This can be extremely helpful when delivering personalized training to offshore personnel and as a result can ensure better preparedness for personnel in real life emergency situations.

References

Adie, W., Cairns, J., Macdiarmid, J., Ross, J., Watt, S., Taylor, C. L., & Osman, L. M. (2005). Safety culture and accident risk control: Perceptions of professional divers and offshore workers. *Safety Science, 43*(2), 131-145.

Breitsprecher, K., Lang, K. A., & McGrath, T. S. (2007). Use of a computer simulation to assess Hazard and Risk perception. *SPE Asia Pacific Health Safety and Security Environment Conference and Exhibition.* Bangkok: Society of Petroleum Engineers.

Brewer, J. D. (2005). *Risk perception and strategic decision making: General insights, a new framework, and specific application to electricity generation using nuclear energy*. Albuquerque: NM: Sandia National Laboratories.

Fenton, N., & Neil, M. (2012). *Risk assessment and decision analysis with Bayesian networks.* CRC Press.

Flin, R., Mearns, K., O'Connor, P., & Bryden, R. (2000). Measuring safety climate: identifying the common features. *Safety science, 34*(1), 177-192.

Groth, K. M., & Mosleh, A. (2012). A data-informed PIF hierarchy for model-based Human Reliability Analysis. *Reliability Engineering & System Safety, 108*, 154-174.

Mearns, K., & Flin, R. (1995). Risk perception and attitudes to safety by personnel in the offshore oil and gas industry: a review. *Journal of Loss Prevention in the Process Industries, 8*(5), 299-305.

Millán, E., & Pérez-De-La-Cruz, J. L. (2002). A Bayesian diagnostic algorithm for student modeling and its evaluation. *User Modeling and User-Adapted Interaction, 12*(2-3), 281-330.

Millán, E., DescalçO, L., Castillo, G., Oliveira, P., & Diogo, S. (2013). Using Bayesian networks to improve knowledge assessment. *Computers & Education, 60*(1), 436-447.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference.* San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Rundmo, T. (2000). Safety climate, attitudes and risk perception in Norsk Hydro. *Safety science, 34*(1), 47-59.

Rundmo, T., Hestad, H., & Ulleberg, P. (1998). Organisational factors, safety attitudes and workload among offshore oil personnel. *Safety science, 29*(2), 75-87.

Smith, J., Veitch, B., & MacKinnon, S. (2015). Achieving competence in offshore emergency egress using virtual environment training. *Proceedings, Offshore Mechanics and Arctic Engineering.* St. John’s.

Steers, R., & Porter, L. (1979). *Motivation and work behavior.* New York: McGraw-Hill Professional Publishing.

Triandis, H. C. (1971). *Attitude and attitude change.* New York: Wiley.

Veitch, B., Billard, R., & Patterson, A. (2008). Emergency Response Training Using Simulators. *Proceedings, Offshore Technology Conference.* Houston.

4. IDENTIFYING ROUTE SELECTION STRATEGIES IN OFFSHORE EMERGENCY SITUATIONS USING DECISION TREES: A STEP TOWARDS ADAPTIVE TRAINING

Mashrura Musharraf, Jennifer Smith, Faisal Khan**, Brian Veitch,
Faculty of Engineering & Applied Science,
Memorial University of Newfoundland,
St John’s, Newfoundland and Labrador, Canada A1B 3X5
** Correspondence author: Tel: + 1 709 864 8939; Email: fikhan@mun.ca

Co-authorship statement

A version of this manuscript has been submitted for review in the Journal of Reliability Engineering & System Safety. Part of this work has been published and presented in the 15th International Conference on Cognitive Modelling (ICCM, 2017). Authors Mashrura Musharraf and Jennifer Smith designed the experiment, conducted the experiment, and performed necessary data collection. The lead author Mashrura Musharraf performed the literature review, developed the framework, performed the data analysis, and prepared the draft of the manuscript. Co-authors Faisal Khan, Brian Veitch, and Scott MacKinnon supervised the experimental study. Faisal Khan reviewed and revised the framework. Brian Veitch reviewed and corrected the interpretation of the result and suggested further possible applications of the method. All co-authors reviewed and provided feedback on the manuscript. Mashrura Musharraf revised the manuscript based on the co-authors’ feedback.

Abstract

Offshore emergency conditions are dynamic in nature and personnel on board are challenged with high risk, time pressure, uncertainty, and the complexity of the situation. This paper investigates how different attributes of emergency scenarios influence people's choice of egress route subsequent to training. An empirical study was carried out in a virtual environment (VE) with 17 naïve participants. The participants were trained to muster during emergencies using a lecture based training (LBT) approach. Training sessions in LBT consisted of computer based training tutorials and simulated training scenarios. Participants' performance was then tested in simulated testing scenarios. It was observed that given the same training, people used different sets of attributes to make decisions on the egress route. This can help to diagnose causes of poor performance and to design adaptive training lessons. Such identification can also help in the assessment of the efficacy of the training curriculum, or the pedagogical approach.

To evaluate the prediction accuracy of the decision trees, the outcomes were compared to the actual observed outcomes of the participants in scenarios in the testing data set. Results show an average of 95% prediction accuracy of the decision trees.

4.1 Introduction

Post-accident analyses of disasters like Piper Alpha and BP Deepwater Horizon show that the crises might have been managed more effectively if the personnel on board could take proper decisions and actions immediately (Flin, 1997). Being able to handle the remoteness of the installation, deal with dynamically evolving situations, and effectively use information coming from different sources demand a high level of competency (Flin, Slaven, & Stewart, 1996). Virtual environments (VEs) in the offshore and maritime industries can help people gain such competency. VEs can be used to create artificial emergency scenarios with hazards such as fire, explosion, and blackouts, and to train individuals about their roles and responsibilities during an emergency (Veitch et al., 2008). During an emergency, the role of general personnel is to choose a tenable route to egress and muster at their designated muster stations. The focus of this paper is to discover egress strategies of individuals during emergency conditions after they have been trained in a VE (Smith et al., 2017).

To this aim, an experimental study was conducted in a VE called All-hands Virtual Emergency Response Trainer (AVERT). AVERT is modeled after an offshore oil platform with high levels of detail and can create credible emergency scenarios ranging from muster drills to more complex scenarios where selected egress routes are blocked with hazards (House et al., 2014). 17 participants took part in the study with a goal to learn how to successfully muster during an offshore emergency situation. Participants were trained using a lecture based training (LBT) approach consisting of interactive video tutorials and simulated training scenarios. After training, participants performance' was tested in multiple simulated testing scenarios. Behavior of the participants were observed during both training and testing scenarios, and human performance data were collected. The collected data were divided into training and testing data sets. After feeding the training data to a decision tree algorithm, a set of decision rules were obtained that describes how people use different attributes of emergency scenarios to choose an egress route. It was observed that even though the participants were exposed to the same training scenarios, on many occasions, they evidently comprehended the information provided in the scenarios differently. Thus, the characteristics of the attributes in a scenario can vary from individual to individual, and so can their decision trees (Joea & Boringa, 2014).

Identifying route selection strategies can be useful to:

- predict whether the participant will be able to successfully egress in a given context.
- identify holes in the strategies that lead to poor performance (Elkind et al., 2014). Comparison of successful versus unsuccessful strategies may help to identify weaknesses of certain strategies and uncover ways to improve performance by adaptive training.
- identify weaknesses of different pedagogical approaches and suggest possible improvements. Given proper training and repeated exposure to emergency scenarios, it is expected that the problem-solving strategies of individuals will converge and lead to success. If not, this can be an indication of weaknesses in the training curriculum or pedagogy.
- train software agents or virtual operators so that they can reproduce similar or compatible problem-solving strategies (Massaguer et al., 2006).

The decision trees were used to predict people's performance in scenarios in the testing data set. Outcomes of the decision trees were compared to the observed outcome of the participants, thereby providing a basis to calculate the prediction accuracy of the trees.

As the decision tree is at the core of the work presented in the paper, an overview of decision tree induction is presented in Section 4.2. Section 4.3 presents the methodology and covers the major work done in the paper. Section 4.4 presents and discusses the results. Section 4.5 summarizes and concludes the findings.

4.2 Overview of Decision tree induction

Induction refers to the process involved in creating generalizations from the observed phenomenon (Badino, 2004). In inductive reasoning, inference leads from individual cases to general principles. Given a collection of training examples (condition x, action $f(x)$) a hypothesis h is generated that approximates the action $f(x)$ (Shaw et al., 1990). Among different induction techniques, decision tree induction is used in this paper. Decision tree offers a visual representation of the reasoning process and has valuable diagnostic capabilities. Compared to other methods, such as artificial neural networks or support vector machines, decision trees can be constructed relatively quickly. Another benefit of decision tree, which is particularly important for this paper, is that it does not require any prior assumptions about the data and can work with limited data compared to other techniques (Duffy, 2008).

The process of induction involves dividing the data cases into certain groups based on the value of a selected attribute, with the goal that the examples in any particular group will belong to the same class. One of the critical tasks of developing a tree is to select the best attribute to branch. Different decision tree algorithms (i.e. ID3, C4.5, CART) use different attribute selection measures such as *information gain*, *gain ratio*, and *Gini index* (Rokach & Maimon, 2014). As all attributes are categorical and there is no concern of missing data points, the ID3 decision tree algorithm is used in this paper. ID3 uses *information gain* for attribute selection (Han et al., 2011).

Information gain is an attribute selection measure that is based on the concept of "information content" or the entropy of a message. The entropy of a random variable X measures the amount of uncertainty of X. A small entropy implies low uncertainty. The idea is to partition data cases into groups such that entropy, and hence uncertainty, is minimized.

Suppose there are m distinct class labels, $L_1, L_2, \dots, L_m$. A random variable $X = L_i$ if a randomly selected object from the entire population has label L_i. Given S is the training set, and S_i is the subset of objects in S with the label L_i, $P(X = L_i)$ can be calculated using equation 4.1.

$$P(X = L_i) \approx \frac{|S_i|}{|S|} \tag{4.1}$$

Entropy of X can then be computed as:

$$Entropy\ (X) = -\sum_{i=1}^{m} P(X = L_i)\ log_2\ P(X = L_i) = -\sum_{i=1}^{m} \frac{|S_i|}{|S|}\ log_2\ \frac{|S_i|}{|S|} \qquad (4.2)$$

Now, suppose a data set S is being partitioned using attribute A. A is discrete and has k distinct values $a_1, a_2, \dots, a_k$. Partitioning the data set on A will result into k data subsets, $S_1, S_2, \dots, S_k$ where each S_j contains data cases that have $A = a_j$. The weighted average entropy across all S_j can be calculated using equation 4.3.

$$Entropy\ (A) = \sum_{j=1}^{k} \frac{|S_j|}{|S|}\ Entropy\ (S_j) \qquad (4.3)$$

Entropy of all attributes can be measured in the same way. Entropy provides a ranking of the attributes given the training data cases. At any time, the attribute with the lowest entropy is chosen for partitioning. Or equivalently, the information gain of the attribute can be calculated as $Gain\ (A) = Entropy(S) - Entropy\ (A)$, and the attribute with the highest gain can be selected for partition.

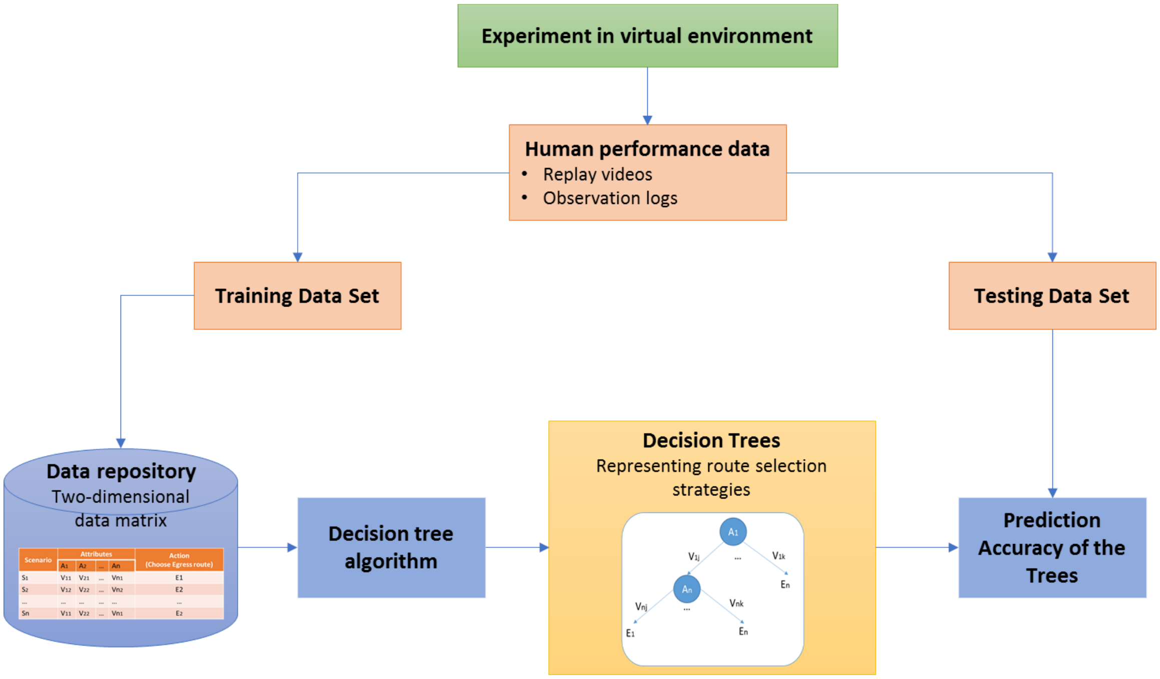
Algorithm 4.1 summarizes the steps of inducing a decision tree from the training data set (Han et al., 2011).

Algorithm 4.1: Basic algorithm for inducing a decision tree from a training data set

```
Input: Training data set, Attribute list
Output: A decision tree

Method:
Begin
    1. Create a node.
    2. If all examples at the current node are of the same class, then label the node with
       the class   and stop.
    3. If the data subset at the current node is empty, then label the node with the
       majority class label in its parent data set.
    4. If no attributes are left for further classification, then label the node with the
       majority class in the current data subset and stop.
    5. For each remaining attribute $A_i$, compute the value of information gain $Gain(A_i)$.
    6. Choose the attribute with the highest gain $Gain(A_i)$ to branch the current node.
    7. For each branch node, go to step 2.
End
```

4.3 Methodology

Figure 4.1 summarizes the steps followed to discover the route selection strategies of individuals in emergency situations. First, an experimental study was conducted in the VE with 17 participants. Participants were trained using a LBT approach and their performance data were collected during the experiment. The collected data were divided into training and testing data sets. The training data were stored in a data repository in the form of a two-dimensional matrix. This data matrix was used as input to the decision tree algorithm to identify the behavioral patterns of route selection. Section 4.3.1 describes the experimental design and data collection in detail. Section 4.3.2 discusses the development of the data matrices. Section 4.3.3 illustrates the development of the decision trees using the data matrices. The testing data set was used to assess the prediction accuracy of the decision trees. This is discussed separately in Section 4.4.



Scenario	Attributes				Action (Choose Egress route)
	A1	A2	...	An	
S1	V11	V21	...	Vn1	E1
S2	V12	V22	...	Vn2	E2
...	...	...	...	...	...
Sn	V11	V22	...	Vn1	E2

Figure 4.1: Steps to identify route selection strategies using experimental data

4.3.1 Experimental design

The data used in this paper were originally collected during an experimental study presented in Smith (2015). For clarity, only the part of the experimental study that is relevant to this paper is discussed here. More details can be found in Smith (2015) and Musharraf et al. (2016).

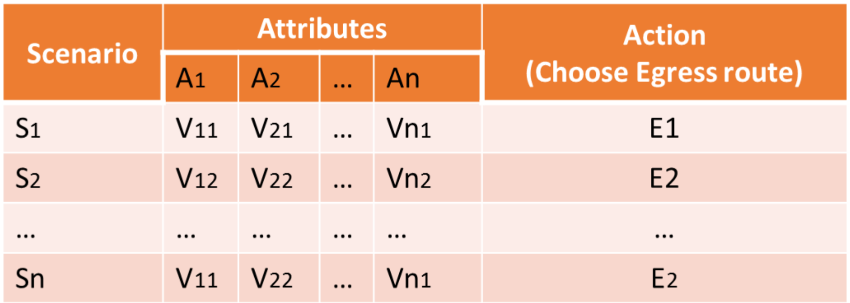
A total of 36 participants took part in the study with the goal to learn how to muster successfully during offshore emergency situations. The participants were naïve concerning any detail of the experimental design, they were not employed in the offshore oil and gas industry, and therefore they were not familiar with the offshore platform. 17 randomly selected participants were given a higher level of training than the others, and only data collected from these participants are used in this paper. The participants were trained using a LBT approach consisting of 3 sessions. In the first session, they received a basic offshore emergency preparedness tutorial, 4 training scenarios, a multiple choice test, and 4 testing scenarios. In both Session 2 and Session 3, participants received an advanced training tutorial about alarms and hazards respectively, 4 additional training scenarios, a multiple choice test, and 4 testing scenarios. So, the participants had to perform in 24 scenarios in total. The purpose of the training scenarios was to provide participants exposure to the VE and emergency situations so that they could learn to choose an egress route and successfully muster at their designated muster stations. After the training phase, their performance was assessed in the testing scenarios. Though active feedback was not provided during the scenarios, an automated review of the performance was presented to the participants after each scenario. The review included several performance metrics, such as route selected for evacuation, time spent running, interaction with fire doors and watertight doors, interaction with hazards, and reporting at the muster stations. A total of 12 scenarios are analyzed in this paper for which *the starting location* was the cabin in the accommodation block.

Videos of participants' performance in the scenarios were recorded during the study for later data analysis. Observation logs were also kept by the investigators to record any observations about the participants' behavior that was not captured by the video. The log also recorded any questions that participants asked during the study.

4.3.2 Developing the data matrices

Among the 12 scenarios, 8 scenarios were used to train the decision tree algorithm. Among the 4 remaining scenarios, 3 were used to test the prediction accuracy of the decision tree. The other scenario had to be excluded from the testing data set as the attribute values in this scenario were very different than the training scenarios and decision trees cannot make a prediction for attribute values they have not seen before.

The training data were stored in a repository in the form of a matrix. Figure 4.2 shows an instance of the matrix. As shown in the figure, the matrix is two-dimensional, with scenarios as row heads $(S_1, S_2, \ldots, S_n)$, and attributes $(A_1, A_2, \ldots, A_n)$ and associated actions (E) as column heads. In general, attributes can be discrete or continuous. All attributes considered in this paper are discrete.



Scenario	Attributes				Action (Choose Egress route)
	A1	A2	…	An	
S1	V11	V21	…	Vn1	E1
S2	V12	V22	…	Vn2	E2
…	…	…	…	…	…
Sn	V11	V22	…	Vn1	E2

Figure 4.2: Data stored in the repository in the form of a two-dimensional matrix. $S_1, S_2, \dots, S_n$ represent the scenarios, $E_1, E_2, \dots, E_n$ represent the associated actions, $A_1, A_2, \dots, A_n$ represent the attributes, V_{ij} represents the j^{th} value of the i^{th} attribute (each attribute can take k possible values)

The attributes of different scenarios in this study were: *Final destination*, *Lights*, *Presence of hazard*, *Alarm*, *Route direction in the PA*, and *Obstructed route*. An additional attribute, *Route taken in previous scenario*, was considered to represent the participant's chosen route in the preceding scenario. Table 4.1 shows the different possible values of each attribute.

Table 4.1: Possible values of each attribute

Attribute	***Possible values***
Final destination	Muster station (MS), Lifeboat station (LB)
Lights	On, Off
Presence of hazard	Yes, No
Alarm	None, General platform alarm (GPA), Prepare to abandon platform alarm (PAPA)
Route direction in the public address (PA) announcement	None, Primary route, Secondary route
Obstructed route	None, Primary route, Secondary route
Route taken in previous scenario	None, Primary route, Secondary route

In all of the scenarios, the participants started in their cabin and were asked to muster at either the muster station or lifeboat station. There were two routes (primary route and secondary route) to get to the final destination from the cabin. Depending on the values of the attributes, participants had to choose a route to egress (i.e. IF obstructed route = primary route, THEN the route of choice should be secondary route).

It has to be noted that even though the participants went through the same training and were exposed to the same scenarios, their understanding of the situations could be different. This means that even for the same scenario, characteristics of scenario attributes may vary from individual to individual. For example, suppose in a scenario the PA announcement is suggesting to take the primary route. For a participant who understood the PA, the value of the attribute *Route direction in PA* would be *Primary route*. However, for a participant who failed to understand the PA, the value of the same attribute could be *None*. The video files and information recorded in the observation logs were used by the investigators to interpret participants' understanding of a scenario and assign the attribute values. This resulted in several different data matrices. Table 4.2 shows the data matrix for one sample participant. It includes both the training and testing scenarios from Session 1, and the training scenarios from Session 2 for a total of 6 data cases.

Data matrices of other participants can be generated in the same way.

Table 4.2: Data matrix at the end of training scenarios in Session 2

Attributes								*Action*
Scenario	***Final Destination***	***Lights***	***Presence of hazard***	***Alarm***	***Route direction in PA***	***Obstructed route***	***Previous route taken***	***Choose route to egress***
LE2	MS	On	No	None	Primary	None	N/A	Primary
LE3	LB	Off	No	None	Secondary	None	Primary	Secondary
TE1	LB	On	No	None	None	None	Secondary	Primary
TE3	MS	Off	No	None	None	None	Primary	Primary
LA2	MS	Off	No	GPA	None	None	Primary	Primary
LA3	LB	On	No	PAPA	None	None	Primary	Primary

4.3.3 Decision tree development

Decision tree induction generates a decision tree from the data cases of known classes described in terms of a fixed set of attributes (Shaw et al., 1990). Given the data cases shown in Figure 4.2, the goal was to classify the cases into groups such that all examples in a group have the same *choice of egress route*. As discussed in Section 4.2, at any time *'t'* the classification is done using the attribute with highest information gain. Figure 4.3 summarizes the process.

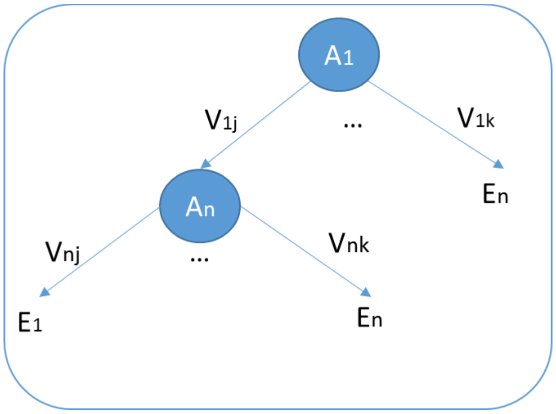


Figure 4.3: Classifying data cases shown in Figure 4.2 based on the characteristics of the attributes

The data matrices generated in Section 4.3.2 were used as inputs to the decision tree algorithm. Figure 4.4 shows the resulting decision tree for the data matrix presented in Table 4.2.

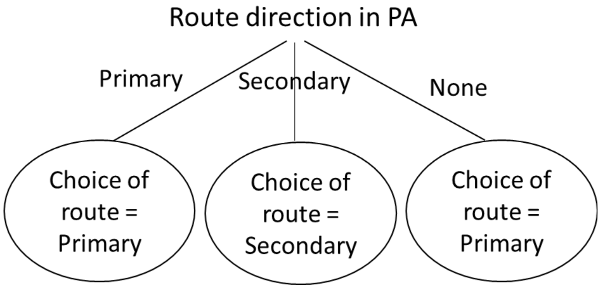

Route direction in PA
Primary
Secondary
None
Choice of route = Primary
Choice of route = Secondary
Choice of route = Primary

Figure 4.4: Decision tree developed from the data matrix shown in Table 4.2

As shown in Figure 4.4, the strategy of the participant is to listen to the PA announcement for route direction and choose a route accordingly. When no route direction is provided in the PA, the participant follows his/her preferred route which is the *primary route*. Strategies of the other participants can be discovered in the same way.

The tree shown in Figure 4.4 can be used to predict the participant's performance in the testing scenarios in Session 2. Once a participant finishes performing in the testing scenarios, experiences from the scenarios are added to the data matrix for re-training a new tree to test the next session’s scenarios.

As the participant moves into Session 3 and finishes performing the training scenarios, more content is added to the data matrix. This changes the decision tree as well. Table 4.3 shows the state of the data matrix for the same participant after training in Session 3. Figure 4.5 shows the corresponding decision tree. The tree shown in Figure 4.5 can be used to predict the participant’s performance in the testing scenarios in Session 3. More on the prediction accuracy will be discussed in Section 4.4.2.

Table 4.3: Data matrix at the end of training scenarios in Session 3

Attributes								*Action*
Scenario	***Final Destination***	***Lights***	***Presence of hazard***	***Alarm***	***Route direction in PA***	***Obstructed route***	***Previous route taken***	***Choose route to egress***
LE2	MS	On	No	None	Primary	None	N/A	Primary
LE3	LB	Off	No	None	Secondary	None	Primary	Secondary
TE1	LB	On	No	None	None	None	Secondary	Primary
TE3	MS	Off	No	None	None	None	Primary	Primary
LA2	MS	Off	No	GPA	None	None	Primary	Primary
LA3	LB	On	No	PAPA	None	None	Primary	Primary
TA1	MS	On	No	GPA	None	None	Primary	Primary
TA3	LB	Off	No	PAPA	None	None	Primary	Primary
LH3 (Frame 1)	MS	On	No	GPA	None	None	Primary	Primary
LH3 (Frame 2)	MS	On	Yes	GPA	Secondary	Primary	Primary	Secondary
LH4 (Frame 1)	MS	On	No	GPA	None	None	Secondary	Primary
LH4 (Frame 2)	LB	On	Yes	GPA	None	None	Secondary	Secondary
LH4 (Frame 3)	LB	Off	Yes	PAPA	None	None	Secondary	Secondary

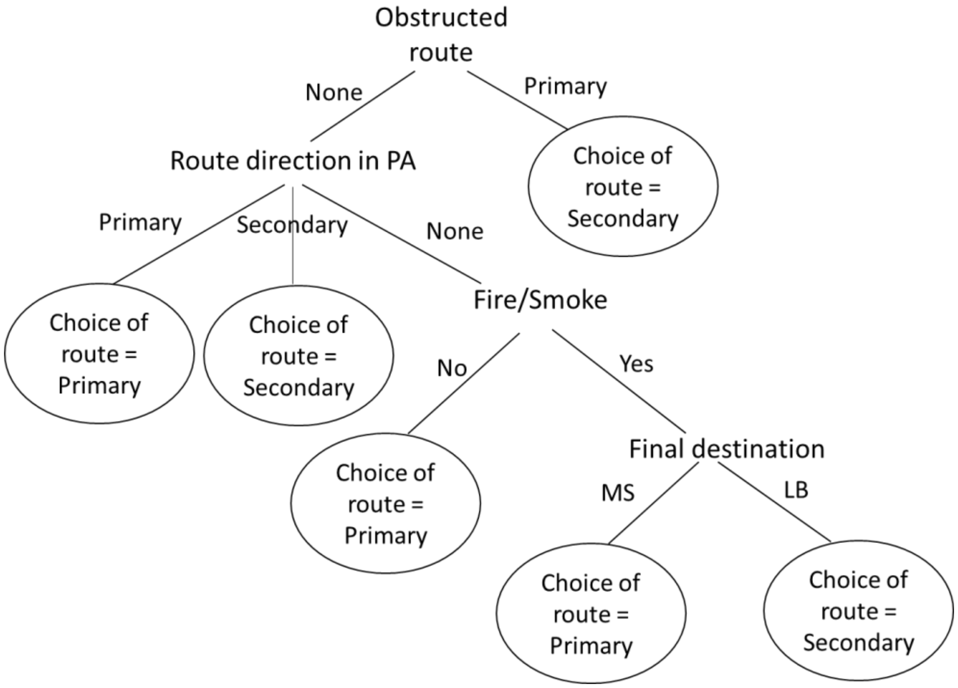

Obstructed route
None
Primary
Route direction in PA
Choice of route = Secondary
Primary
Secondary
None
Choice of route = Primary
Choice of route = Secondary
Fire/Smoke
No
Yes
Choice of route = Primary
Final destination
MS
LB
Choice of route = Primary
Choice of route = Secondary

Figure 4.5: Decision tree developed from the data matrix shown in Table 4.3

It has to be noted that scenarios in Session 3 of the experiment were dynamic, and the value of the attributes changed during the scenarios. As shown in Table 4.3, these scenarios (i.e. LH3, LH4) were divided into frames such that in each frame the characteristics of the attributes remains static. The participant had to choose an egress route based on the value of the attributes in each frame. For example, in LH3, the participant started in his/her cabin when a GPA sounded and the PA announced that the platform is in alarm status. There was no indication of hazard or obstruction of any route in the PA. Based on the value of the attributes, the participant chose the primary route to egress at this time. This is the first frame of LH3. As the scenario progressed, the PA announced that there is a fire on the upper deck and smoke in the main stairwell of the accommodation block. This confirmed the presence of a hazard and indicated that the primary route was obstructed. Based on the changed values of the attributes, in Frame 2, the participant re-routed and took the secondary route to egress.

This section discussed the data matrix and decision tree generation of one sample participant in detail. As discussed earlier in this section, the values of the attributes of the data matrix can vary from individual to individual and so can the decision trees. The following section shows the decision trees for all 17 participants.

4.4 Results and discussion

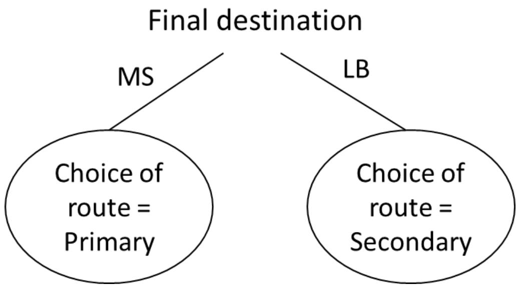
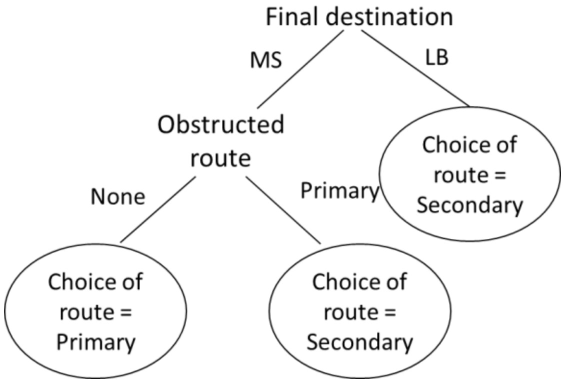
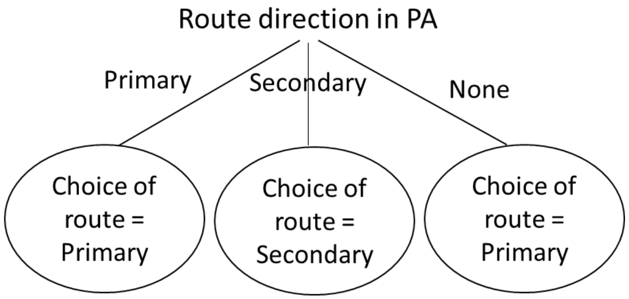
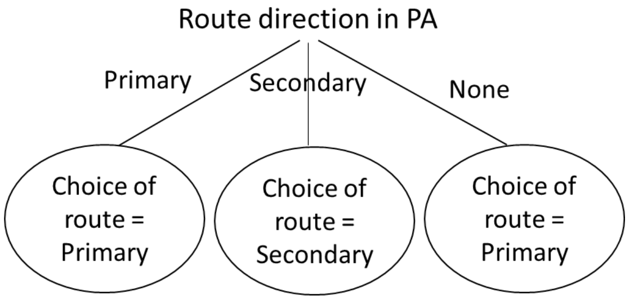
4.4.1 Analysis of results

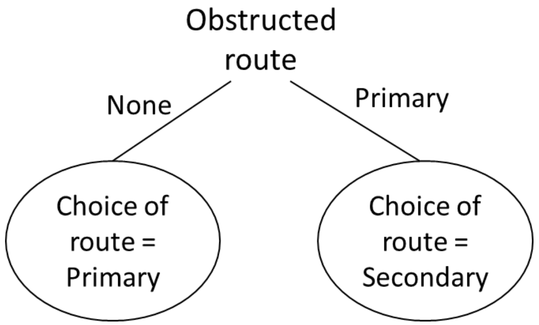
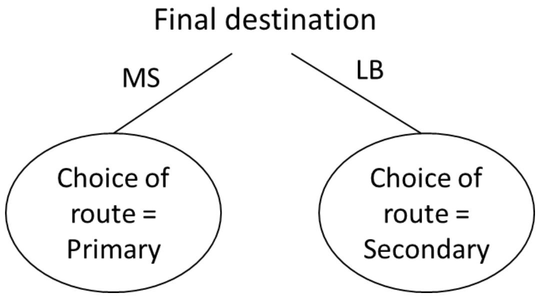
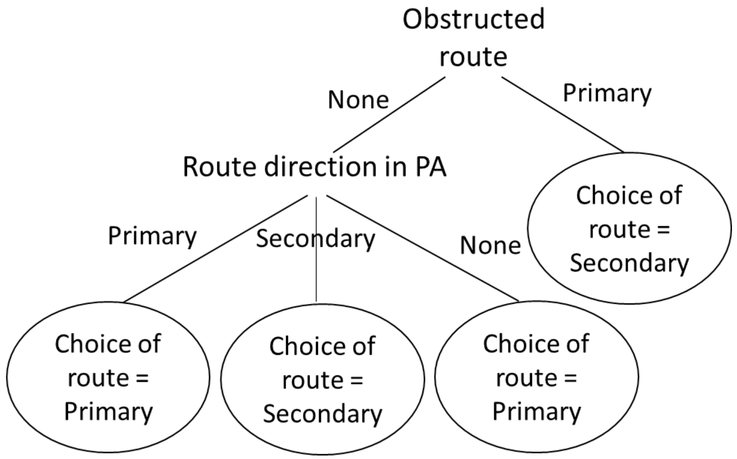
Table 4.4 shows the different decision trees for all participants in the study.

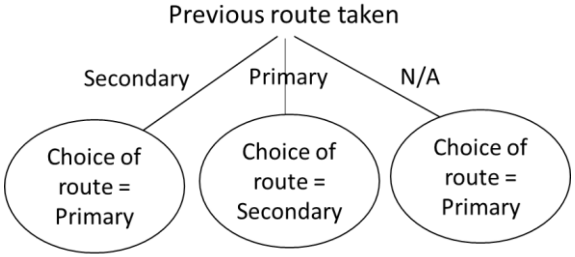
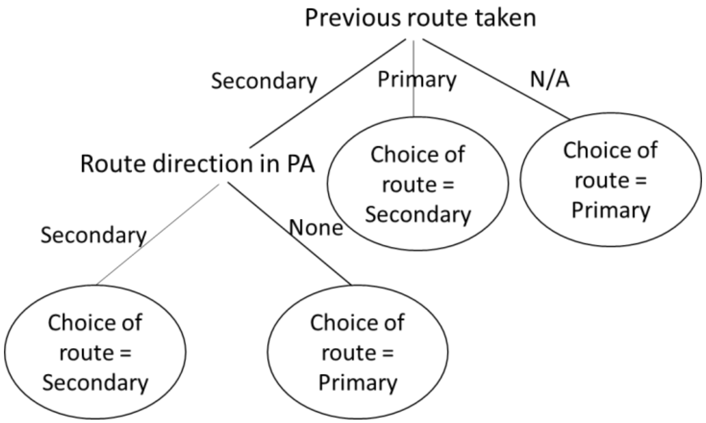
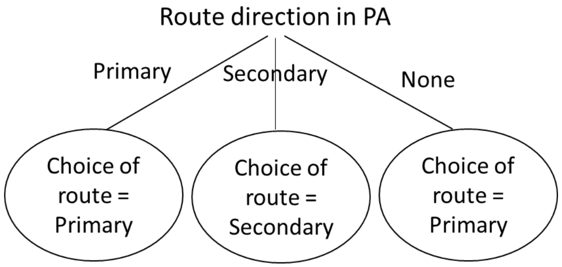
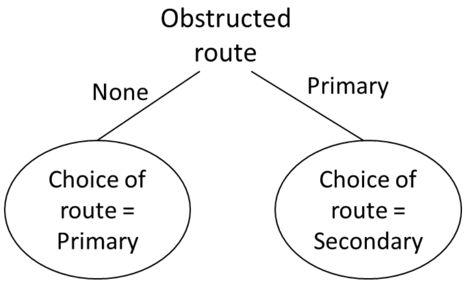
As shown, for 13 participants out of 17, the general understanding or strategy can be identified using a decision tree. For the remaining participants, the decision rules from the tree were the same as the ones in the data matrix and no generalization could be made.

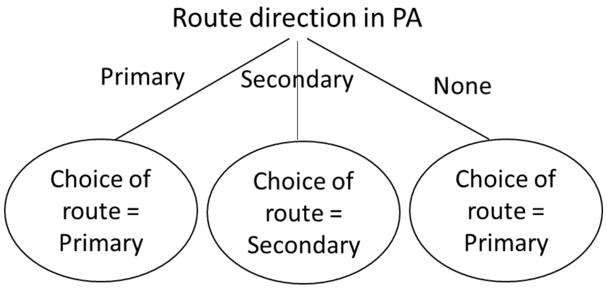
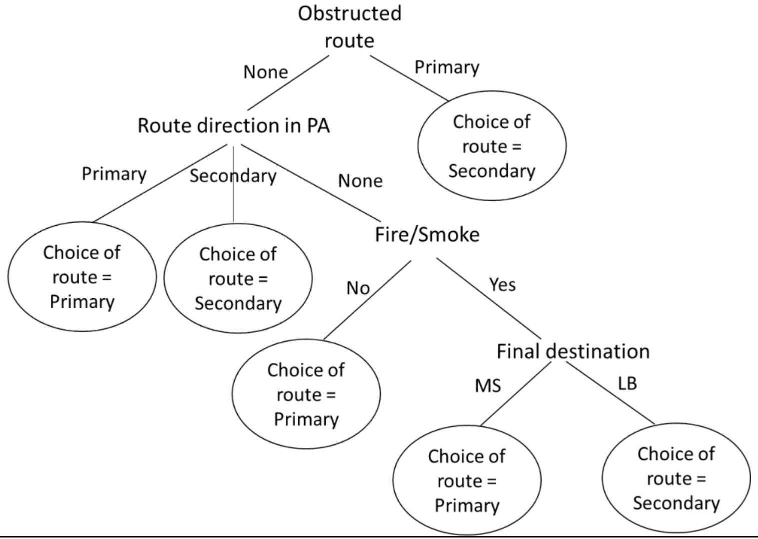
Given the training, it was expected that participants would be able to interpret the PA and choose an egress route based on the direction provided in the PA. Only 37% of the participants in Session 2 chose an egress route based on the route direction in PA. In Session 3 this declined to 25%. As shown in Table 4.4, participants were using other attributes, such as final destination, previous route taken, and obstructed route, to choose an egress route. Using such attributes for route selection can lead to failure (i.e. interaction with hazard). It is possible to improve the training program so that such failure is eliminated and all participants have a route choice strategy that leads to success.

Table 4.4: Decision tree for 17 participants

Participant No	***Decision Tree at the end of training in Session 2 (No hazard)***	***Decision Tree at the end of training in Session 3 (With Hazard)***
G1-03, G1-09, G1-10	 Final destination MS LB Choice of route = Primary Choice of route = Secondary	 Final destination MS LB Obstructed route Choice of route = Secondary None Primary Choice of route = Primary Choice of route = Secondary
G1-01, G1-04, G1-07, G1-08	 Route direction in PA Primary Secondary None Choice of route = Primary Choice of route = Secondary Choice of route = Primary	 Route direction in PA Primary Secondary None Choice of route = Primary Choice of route = Secondary Choice of route = Primary

Participant No	***Decision Tree at the end of training in Session 2 (No hazard)***	***Decision Tree at the end of training in Session 3 (With Hazard)***
G1-14	At any condition: Choice of route = Primary route	 Obstructed route None Primary Choice of route = Primary Choice of route = Secondary
G1-16	 Final destination MS LB Choice of route = Primary Choice of route = Secondary	 Obstructed route None Primary Route direction in PA Choice of route = Secondary Primary Secondary None Choice of route = Primary Choice of route = Secondary Choice of route = Primary
G1-13	At any condition: Choice of route = Secondary route	At any condition: Choice of route = Secondary route

Participant No	***Decision Tree at the end of training in Session 2 (No hazard)***	***Decision Tree at the end of training in Session 3 (With Hazard)***
G1-06	 Previous route taken Secondary Primary N/A Choice of route = Primary Choice of route = Secondary Choice of route = Primary	 Previous route taken Secondary Primary N/A Route direction in PA Choice of route = Secondary Choice of route = Primary Secondary None Choice of route = Secondary Choice of route = Primary
G1-05	 Route direction in PA Primary Secondary None Choice of route = Primary Choice of route = Secondary Choice of route = Primary	As long as the participant understands PA, s/he follows the same decision tree. If the participant fails to understand PA, then s/he makes a choice based on obstructed route. In case PA is not understandable, the decision tree is changed as below:  Obstructed route None Primary Choice of route = Primary Choice of route = Secondary

Participant No	***Decision Tree at the end of training in Session 2 (No hazard)***	***Decision Tree at the end of training in Session 3 (With Hazard)***
G1-02	 Route direction in PA Primary Secondary None Choice of route = Primary Choice of route = Secondary Choice of route = Primary	 Obstructed route None Primary Choice of route = Secondary Route direction in PA Primary Secondary None Choice of route = Primary Choice of route = Secondary Fire/Smoke No Yes Choice of route = Primary Final destination MS LB Choice of route = Primary Choice of route = Secondary
G1-15	Had a hard time with doors, so chose a route with fewer doors. Excluded from data set as the attribute is too specific for AVERT and won't apply in real life.	
G1-17, G1-11, G1-12	As the choice of route was random, decision tree does not give any more generalization than the data matrix. The decision tree contains the same decision rules as the data matrix, no behavioral pattern or strategy is identified.	

4.4.2 Prediction accuracy

The decision trees generated at the end of Session 2 were used to predict participants' performance in the testing scenarios in Session 2. As the participants moved into Session 3, the data matrices were re-trained and new examples were added. The updated decision trees were used to predict performance of the participants in the testing scenarios in Session 3. The prediction accuracy of the trees was then calculated using equation 4.4. Table 4.5 summarizes the result.

$$(\%) Prediction\ accuracy = \frac{n}{N} \times 100 \tag{4.4}$$

Where $n =$ number of test scenarios for which (predicted outcome = observed outcome) and $N =$ total number of test scenarios

Table 4.5: Classification accuracy of the decision trees

Participant no	(%) prediction accuracy of the decision trees
G1-01	100
G1-02	100
G1-03	100
G1-04	67
G1-05	100
G1-06	67
G1-07	100
G1-08	100
G1-09	100
G1-10	100
G1-13	100
G1-14	100
G1-16	100
Average	95

As shown in Table 4.5, the prediction accuracy of the decision trees is 95% on average. The trees predicted the performance of 11 participants with 100% accuracy (3 out of 3 testing scenarios), and 2 participants with 67% accuracy (2 out of 3 testing scenarios).

4.5 Limitations

There are a few limitations with the study that need to be mentioned. First of all, VE can only provide a certain degree of realism and cannot be considered as an exact representation of the real world operating conditions. Secondly, the participants of the study were naïve. It is anticipated that repetition of the same experiment with real offshore workers would result in a different set of route selection strategies. Finally, the training and testing data set used in the paper is limited. The small training data set increases the possibility of overfitting. Future work will aim for a larger data set with balanced classes to improve the prediction accuracy. Another improvement in future works will be to use more advanced decision tree algorithms (e.g. C4.5 or C5) that support tree pruning to avoid overfitting. Two common approaches of tree pruning are: 1) Stop the growing phase at a certain point even if the halting conditions in the growing phase are not met, and 2) Let the tree grow to its fullest height in the growing phase but then remove leaves iteratively based on some criterion. More details on the pruning process can be found in Han et al. (2011).

4.6 Conclusion

Though extensive research has been done on human behavior in some industries, limited studies are available on behavioral representation of offshore workers (Sharma et al., 2008, Baron et al., 1980, Woods, 1987, Cacciabue et al., 1992, Sasou et al., 1995). This paper presents a study that investigates peoples' route selection behavior in offshore emergency situations after a targeted training program. The decision tree algorithm is used to identify peoples' behavioral patterns during route selection. Results show that the trees can predict people's choice of route in future emergency scenarios with an average of 95% accuracy.

Identification of the route selection strategies can be useful in many ways. First, the decision trees can be used to predict the response of general personnel for a given situation. This can be extremely helpful in designing adaptive training so that individuals can reach competency faster. Next, the range of decision trees can help to detect the most effective strategy for a given situation. The ineffective strategies can be analyzed to see how and why they lead to poor performance, and find out ways of improvement. The identified strategies can also be used to assess the training curriculum and/or pedagogical approach. It is expected that a sound training process would ensure convergence to strategies that lead to success. Any systemic exception might be an indication of weakness of the training approach itself. Identifying and addressing of such weakness can yield better training curriculum or pedagogy. Finally, the strategies can be used to train human-like virtual operators. Virtual operators with different levels of skills (naïve versus expert) can be created by training them with different sets of strategies (successful versus erroneous).

References

Badino, M. (2004). An application of information theory to the problem of the scientific experiment. *Synthese*, *140*(3), 355-389.

Baron, S., Muralidharan, R., Lancraft, R., & Zacharias, G. (1980). *PROCRU: A model for analyzing crew procedures in approach to landing.* NASA.

Cacciabue, P., Decortis, F., Drozdowicz, B., Masson, M., & Nordvik, J. (1992). COSIMO: a cognitive simulation model of human decision making and behaviour in accident management of complex plants. *IEEE Transactions on Systems, Man and Cybernetics, 22*(5), 1058-1074.

Duffy, V. G. (2008). *Handbook of digital human modeling: research for applied ergonomics and human factors engineering.* CRC press Taylor & Francis Group.

Elkind, J. I., Card, S. K., & Hochberg, J. (2014). *Human performance models for computer-aided engineering.* Academic Press.

Flin, R. (1997). Crew resource management for teams in the offshore oil industry. *Team Performance Management, 3*(2), 121-129.

Flin, R., Slaven, G., & Stewart, K. (1996). Emergency decision making in the offshore oil and gas industry. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 38*(2), 262-277.

Han, J., Kamber, M., & Pei, J. (2011). *Data mining: concepts and techniques.* Elsevier.

House, A. W., Smith, J., MacKinnon, S., & Veitch, B. (2014). Interactive simulation for training offshore workers. *Oceans'14 MTS/IEEE Conference* (pp. 1-6). St. John's, NL: IEEE.

Joea, J. C., & Boringa, R. L. (2014). Individual Differences in Human Reliability Analysis. *12th Bi-Annual International Meeting of the Probabilistic Safety Assessment and Management (PSAM) Conference.*

Massaguer, D., Balasubramanian, V., Mehrotra, S., & Venkatasubramanian, N. (2006). Synthetic humans in emergency response drills. *Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems* (pp. 1469-1470). Hakodate, Hokkaido, Japan: ACM.

Musharraf, M., Smith, J., Khan, F., Veitch, B., & MacKinnon, S. (2016). Assessing offshore emergency evacuation behavior in a virtual environment using a Bayesian Network approach. *Reliability Engineering & System Safety, 152*, 28-37.

Rokach, L., & Maimon, O. (2014). *Data mining with decision trees: theory and applications.* Singapore: World Scientific Publishing Co. Pte. Ltd.

Sasou, K., Takano, K., Yoshimura, S., Haroko, K., & Kitamura, M. (1995). Modelling and simulation of operator team behaviour in nuclear power plants. *Proceedings of the HCI international.* Tokyo.

Sharma, S., Singh, H., & Prakash, A. (2008). Multi-agent modeling and simulation of human behavior in aircraft evacuations. *IEEE Transactions on aerospace and electronic systems, 44*(4), 1477-1488.

Shaw, M. J., Gentry, J. A., & Piramuthu, S. (1990). Inductive learning methods for knowledge-based decision support: a comparative analysis. *Computer Science in Economics and Management, 3*(2), 147-165.

Smith, J. (2015). *The effect of virtual environment training on participant competence and learning in offshore emergency egress scenarios.* Master of Engineering Thesis, Memorial University of Newfoundland, St. John's.

Smith, J., Musharraf, M., Veitch, B., & Khan, F. (2017). Can simulation-based mastery learning increase compliance: investigating decision making in virtual offshore emergency egress. *Proceedings of the 3rd Workshop and Symposium on Safety and Integrity management of operations in harsh environments (CRISE-3).*

Veitch, B., Billard, R., & Patterson, A. (2008). Emergency Response Training Using Simulators. *Offshore Technology Conference.*

Woods, D. D. (1987). Cognitive environment simulation: an artificial intelligence system for human performance assessment. *NUREG/CR-4862, 1-3*.

5. MODELING AND SIMULATION OF OFFSHORE PERSONNEL DURING EMERGENCY SITUATIONS

Mashrura Musharraf*, Faisal Khan, Brian Veitch
Centre for Risk, Integrity and Safety Engineering (C-RISE)
Faculty of Engineering & Applied Science,
Memorial University of Newfoundland,
St John’s, Newfoundland and Labrador, Canada A1B 3X5
* Correspondence author: Tel: + 1 709 864 6764; Email: mm6414@mun.ca

Co-authorship statement

A version of this manuscript has been published and presented in the 3rd Workshop and Symposium on Safety and Integrity Management of Operations in Harsh Environments (CRISE-3). The manuscript has also been accepted in the Journal of Safety Science with revision. The lead author Mashrura Musharraf performed the literature review, developed the human behavior model, implemented the model in the Integrated Performance Modeling Environment (IPME), performed the simulations, and prepared the draft of the manuscript. Co-authors Faisal Khan and Brian Veitch reviewed and revised the model and the results, and guided the simulation. They also reviewed and provided feedback on the manuscript. Mashrura Musharraf revised the manuscript based on the co-authors’ feedback and during the peer review process.

Abstract

The offshore oil industry functions in a team work culture, in which operations depend not only on individuals' competency, but also on team skills. Team skills are even more necessary when it comes to handling emergency conditions as they challenge personnel on board with high risk, time pressure, and complexity. This raises the need for training that goes beyond conventional training programs and incorporates team skills exercises. The major difficulty to design such training is that it involves practicing emergency exercises with a potentially large number of participants. Such large-scale team exercises suffer from both organizational and educational drawbacks. One solution to this problem is to use artificial agents that can reproduce the behavior of the team members. This paper presents a behavior model that can simulate the response of general personnel during emergency situations. The variability in human behavior is modeled using different performance influencing factors (PIFs). Empirical evidence is used to identify the sources of variability that are encoded in the agents to allow a realistic range of human behaviors. Though variability can come from both physical and mental differences, the focus of this paper is on the later. Focus is given to across-subject variability rather than within-subject variability.

5.1 Introduction

The offshore oil industry functions in a team work culture and operations usually involve a group of people working together. This makes teamwork an essential component of effective emergency responses. Members of a team not only need to understand their own roles and responsibilities, but also need to have clear understanding of the roles and responsibilities of the other team members. Such understanding is critical for emergency situations as most of the members will have different roles and responsibilities than their everyday duties (Flin, 1997). However, traditional training programs are often generic and are not designed to provide trainees with the understanding of social and cognitive aspects of effective team work.

O'Connor & Flin (2003) discuss the possibility of adopting the crew resource management technique, pioneered in the aviation industry, in offshore oil industries to enhance team performance. Crew resource management (CRM) is defined as “using all the available resources – information, equipment, and people – to achieve safe and efficient flight operations” (Moffat & Crichton, 2015). A significant part of the CRM training requires the trainees to participate in team training exercises using simulator flights. Organizing such team exercises for offshore industries may suffer from both organizational and educational drawbacks (Van Diggelen et al., 2010). Gathering all the team members at the same time and at the same location itself is a challenge. Even when it is possible, the financial requirement is high. Also, the members often have different training needs based on their competency levels. One solution to this problem is to develop a team training platform in a simulator where the roles of some of members are played by humans, while the roles of others are played by artificial intelligent agents (Van Diggelen et al., 2010). Though extensive research has been done to create artificial intelligent agents in military (Jones et al., 1999; Sampson & Ripingill Jr, 2003; Wray & Laird, 2003), aviation (McNally, 2005; Sharma, 2009), and nuclear power plant (Cacciabue et al., 1992; Chang & Mosleh, 2007a;

Dang, 1996) training simulators, no such model is available to date for offshore emergency training simulators.

This paper presents a computational human behavior simulation model (HBM), which is the first step to create such intelligent agents for an offshore emergency training simulator. Realism of agents largely depends on their underlying HBMs. HBMs are computational models that probabilistically simulate human behavior in different conditions. The purpose of the HBM presented in this paper is to reproduce the behavior of people working on offshore petroleum platforms, general personnel in particular, during emergency situations.

Unlike other human behavior models, the proposed model considers a larger fraction of the possible behavior space, which includes both correct and incorrect behaviors (Wray & Laird, 2003; McNally, 2005). Different performance influencing factors (PIFs) are used to model the variability across the behavior space. As use of subject matter experts' (SMEs') opinion often leads to a less reliable model (Chang & Mosleh, 2007c), empirical evidence is used in the development of the HBM. To this end, a two-level, three factor experiment was conducted to observe the influence of different PIFs on emergency response. Earlier works of the authors have discussed in detail the underlying mathematical models that capture the impact of external (Musharraf et al., in press) and internal PIFs (Musharraf et al., 2016) on human performance. Details of the learning and decision making process of individuals have been discussed in (Musharraf et al., 2017b). The goal of this paper is to present an HBM that integrates the different mathematical models and memory structure discussed in previous papers to produce automated probabilistic simulation of offshore workers' response under the pressure of an emergency. Prior to implementing the HBM in the actual simulator, it is modeled in the Integrated Performance Modeling Environment (IPME) simulation framework to define the implementation work scope and identify the technical challenges. Example results generated by the HBM during an IPME scenario simulation are discussed in this paper. Implementation of the HBM in the training simulator and validation of the HBM are discussed separately in (Musharraf et al., 2017).

5.2 Overview of the HBM

Modeling human behavior is a challenging area of research that needs considerations of both modeling and simulation, and behavioral and cognitive psychology (Goerger et al., 2005). There are qualitative models that focus mostly on the behavioral and cognitive psychology, and describe in detail the evolution of the human cognition process upon receiving an external stimulus from the environment (Trucco & Leva, 2007). Then, there are quantitative models that are based on the structure of the qualitative ones, but focus on the computational functionalities of modeling and simulation (Chang & Mosleh, 2007a). The HBM presented in this paper is a computational behavior simulation model that is a simplification of complex environmental settings and complex cognitive processes of human operators.

Section 5.2.1 introduces the different components of the HBM model. Section 5.2.2 describes how knowledge gained from training and experience is stored and retrieved during cognitive functions. The reasoning module is also discussed in this section.

5.2.1 Dynamic response model

The dynamic response model consists of four component models – an environment model, an operator model, a performance shaping model, and a task network model (after Chang & Mosleh, 2007a). The dynamic response model presented in this paper looks at individuals in isolation. Collaboration of team members and the concept of shared situation awareness is out of scope of this paper.

Environment model: The environment model includes external factors that define the circumstances or environment in which the individual is situated. This allows modeling human response under different environment conditions. External factors in the environment model include team-related factors (e.g. communication availability and quality, team composition), organization factors (e.g. safety and quality culture, procedure availability, adequacy, and quality), environment factors (e.g. temperatures, visibility), and conditioning events (e.g. latent failures) (Chang & Mosleh, 2007b).

Operator model: The operator model defines the characteristics of an operator in terms of internal factors. In the context of this paper, operator refers to general personnel working on offshore petroleum platforms. Though internal factors include both physical and non-physical attributes of the operator, this paper focuses on non-physical attributes only. The operator model allows modeling operators who may have different responses given the same environmental condition. Examples of internal factors used in the operator model include attention, bias, compliance, and efficacy of information use.

Task networking model: Task network modeling focuses on understanding the tasks that need to be simulated. The task network model graphically represents the sequence of tasks performed by an operator. Operators' behavior generally consists of different interrelated cognitive functions (Trucco & Leva, 2007). This paper considers four cognitive functions performed by the general personnel: perception, interpretation, decision making, and execution. Any function is decomposed into a series of sub functions, which in turn are decomposed into tasks for the development of the task network. Failure can happen at any stage of performing a task. Also, there can be more than one correct behavior or way to fail. The task network helps to identify possible deviations from the ideal behavior path(s) that may lead to error.

Performance shaping model: This model includes a set of performance shaping functions (PSFs). The PSFs generate the rules of behavior that govern the performance of general personnel while performing cognitive tasks. The response of general personnel depends on the state of the operator (e.g. stress, task related and non-task related load) and the current state of knowledge (e.g. scenario based knowledge from training and experience). The PSFs take the state of the operator and current state of knowledge into account and generate the associated operator response for a given set of PIFs. The PSFs used in the HBM development process are defined using a Bayesian Network (BN) approach. BNs have proven to be a powerful modeling tool due to their capability to 1) consider dependency among PIFs and associated actions, 2) quantify the impact of different PIFs on successful or erroneous behavior, and 3) update success or failure likelihood each time new evidence becomes available (Fenton & Neil, 2012; Podofillini & Dang, 2013; Sundaramurthi & Smidts, 2013). BNs have been widely used to model the impact of different PIFs on human performance or human error (Baraldi et al., 2009; Dang & Stempfel, 2012). Details of the PSF development is discussed in Section 5.3.

Figure 5.1 shows the interaction between the external world and the component models. At any time 't' the state of PIFs in the environment model and operator model are defined based on the events happening in the external world. The state of the internal and external factors defines the operator's state of mind. The PIFs also influence how information is memorized from training and experience, and retrieved when necessary. The PSF model takes the operator's state of mind and current state of knowledge into account, and generates the behavior rules that govern the operator's response during cognitive tasks.

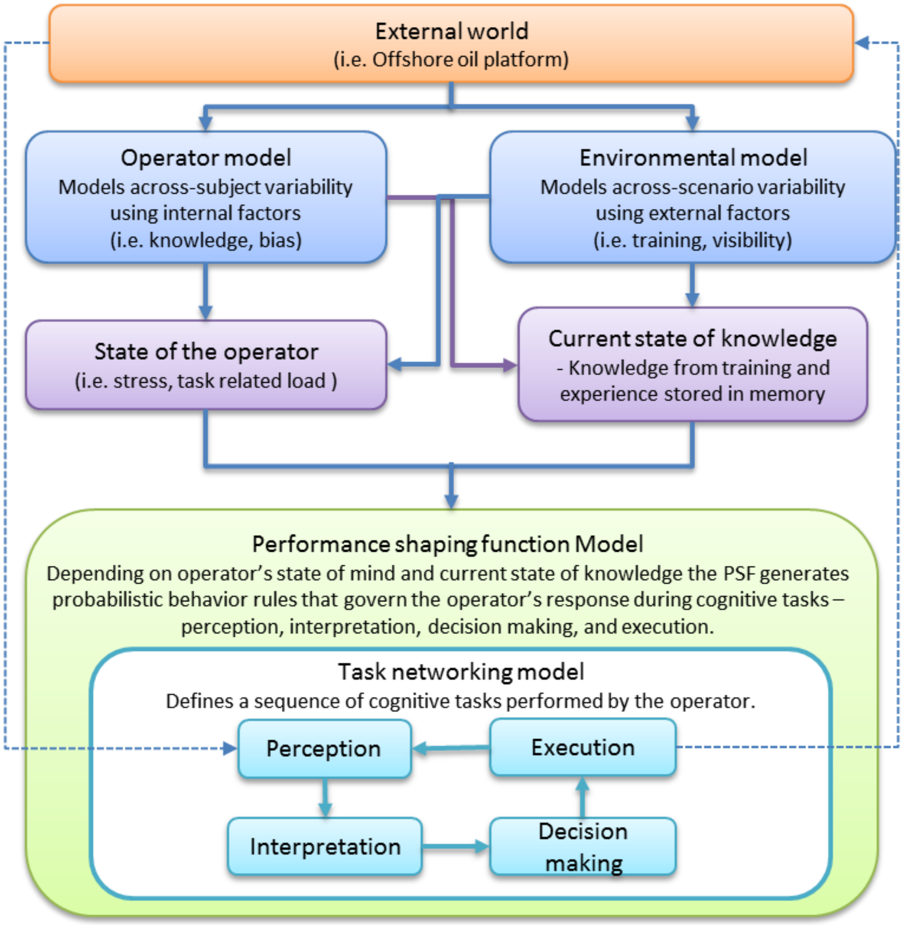

External world
(i.e. Offshore oil platform)
Operator model
Models across-subject variability using internal factors (i.e. knowledge, bias)
Environmental model
Models across-scenario variability using external factors (i.e. training, visibility)
State of the operator
(i.e. stress, task related load)
Current state of knowledge
- Knowledge from training and experience stored in memory
Performance shaping function Model
Depending on operator's state of mind and current state of knowledge the PSF generates probabilistic behavior rules that govern the operator's response during cognitive tasks – perception, interpretation, decision making, and execution.
Task networking model
Defines a sequence of cognitive tasks performed by the operator.
Perception
Execution
Interpretation
Decision making

Figure 5.1: Dynamic response model

5.2.2. Memory structure and cognitive functions

This section describes the memory structure and the cognitive functions as part of the HBM. The HBM used here simplifies the complex memory and cognitive processes of human operators.

The purpose of the HBM presented in this paper is to create intelligent agents for an offshore emergency training simulator. It is assumed that these agents' response to emergency situations depends in part on the knowledge they have stored in their memory.

A database representative of human memory is created in the HBM. The two main components of the memory structure are knowledge base and working memory, which are modeled based on the idea of long-term and short-term memory in the information processing model of Atkinson & Shiffrin (1968). According to the information processing model, memory consists of several 'stores' with different storage capacity. In the proposed HBM, the working memory has a finite capacity and stores the information relevant to the current cognitive process. The knowledge base has a theoretically infinite capacity and stores all the knowledge gained through training and experience.

A reasoning module, or inference engine, is added in the HBM to model the agents' reasoning process (after Li, 2013). Among different reasoning approaches, inductive reasoning is used (Li & Mosleh, in press). In inductive reasoning, generalizations are created from observed phenomena or principles. Decision tree induction is used in the HBM development (Musharraf et al., 2017b). Decision tree offers a visual representation of the reasoning process and has useful diagnostic capabilities. Compared to other methods, such as artificial neural networks, or support vector machines, decision trees can be constructed relatively quickly. Other benefits of decision trees are that they do not require any prior assumptions about the data, and can work with limited data compared to other techniques (Duffy, 2008). More on decision tree and its benefits can be found in Han et al. (2011). To make information retrieval fast and easier, besides reasoning, the inference engine in the HBM has the added functionality of storing the created generalizations once reasoning is completed.

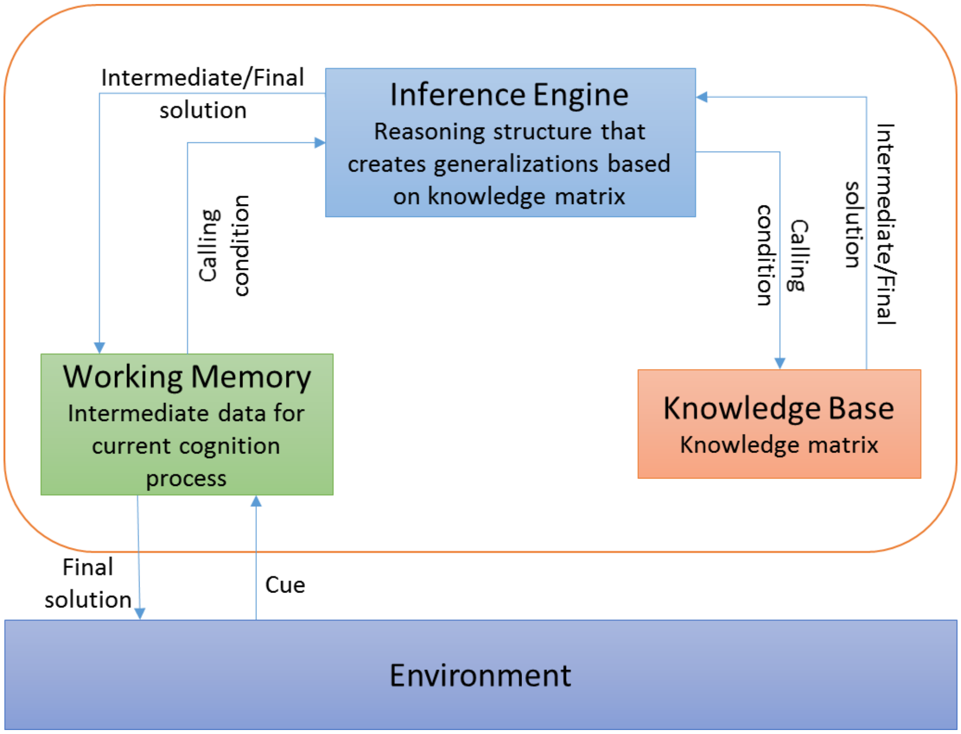

Intermediate/Final solution
Inference Engine
Reasoning structure that creates generalizations based on knowledge matrix
Calling condition
Calling condition
Intermediate/Final solution
Working Memory
Intermediate data for current cognition process
Knowledge Base
Knowledge matrix
Final solution
Cue
Environment

Figure 5.2: HBM main components of memory structure, reasoning module, and their interaction during the cognitive process

Figure 5.2 shows the interaction among the components of memory structure and reasoning module during the cognitive process. At the beginning of the cognitive process, cues are perceived from the environment. The perceived cues are interpreted to form a calling condition. A calling condition is a set of variables that takes values from a defined set (Thow-Yick, 1994). If a solution to the current calling condition is available in the working memory, it is immediately retrieved. Otherwise, the calling condition is transferred to the inference engine to find a solution. If a decision rule that matches the current calling condition is found, the solution is retrieved and sent to the working memory to act upon. If no matching decision rule is found, the calling condition is sent to the knowledge base.

Higher level analogy of the calling condition may be used at this stage to find a solution in the knowledge matrix. Once a solution is found, the next step is to execute a series of actions to implement it.

5.3 Case study: HBM for general personnel

This paper aims to develop an HBM for general personnel in the context of offshore emergencies. General personnel are individuals whose responsibility during an emergency is to follow the alarm(s) and public address (PA) announcement(s), and muster at their designated muster stations. As mentioned in Section 5.2.1, the focus of the HBM presented in this section is to reproduce behavior of individuals in isolation. Collaboration and shared situation awareness among team members are not considered at this stage.

The HBM presented in this paper uses empirical evidence. An experimental study was done to 1) populate the knowledge base, 2) define the inference process, and 3) investigate the influence of different PIFs on task performance. For clarity, only the part of the experimental study that is relevant to this paper will be discussed here. More details can be found in Smith (2015) and Musharraf et al. (2016).

The experiment was conducted using a virtual environment. A virtual environment is a computer aided simulation environment that allows trainees to gain artificial experience, including in dangerous scenarios. The virtual environment used in the case study is called the all-hands virtual emergency response trainer (AVERT). AVERT was designed to enhance offshore emergency response training. It is modeled after an offshore oil platform with high levels of detail and can create credible emergency scenarios by introducing hazards such as blackouts, fires, and explosions.

A total of 36 participants took part in the experimental study with a goal to learn how to muster during offshore emergency situations. Samples of convenience method was followed for participant recruitment (Ritter et al., 2012). Majority of the participants were university students. The participants were naïve concerning any detail of the experimental design, they were not employed in the offshore oil industry, and they were not familiar with the offshore platform. Among the 36 participants, 27 were males and 9 were females. The age range of the participants was 19-39 years, with a mean of 26.5 years and standard deviation of 4.4 years.

For the case study, the participants had to go through a range of offshore muster scenarios, from drills that required the participants to go to their primary muster station, to more complex emergency scenarios that required the participants to avoid hazards blocking their egress routes and muster at their lifeboat stations (House et al., 2014). The scenarios did not directly induce any operator state. Rather, the scenarios were designed such that the effect of different PIFs on individuals' performance during offshore emergency situations can be investigated. The underlying assumption is that the state of the PIFs implicitly induces a certain operator state. For example, consider a highly complex scenario where a participant's preferred route is blocked by a hazard, there is a blackout due to the hazard, the alarm changes mid-scenario, and the amount of incoming information through the PA is extensive. It is expected that this situation will induce a stressed state and high mental load (both task related and task non-related). More details on the different PIFs are discussed in subsections 5.3.2 and 5.3.3.

As stated in section 5.2.1, the four major cognitive functions considered in this paper are perception, interpretation, decision making, and execution. During the scenarios, the participants had to perceive the audio-visual cues provided through alarms and PA announcements. Next, the participants had to analyze the cues and interpret what the alarms and PA announcements mean (i.e. which route is obstructed, what is the recommended muster location). Once the participants were aware of the situation, they needed to evaluate the potential routes and choose a route to egress. Finally, the participants had to move along the egress route following all safety procedures (i.e. not running, and closing all fire/watertight doors). They needed to reach the muster location in a timely manner and muster there. More details on the task performed by general personnel is discussed in section 5.3.4.

Several performance metrics of the participants were recorded during each scenario. The metrics include: time to muster, time spent running, interaction with fire doors and watertight doors, interaction with hazards, and reporting at different muster locations. The data collected during the experiment are available in Musharraf et al. (2017a).

The following sections illustrate how the HBM presented in section 5.2 can be applied to model the behavior of general personnel in offshore emergency situations using the experimental data. Section 5.3.1 discusses how knowledge gained from training is stored in the memory. Section 5.3.2 to 5.3.5 discuss the different components of the response model.

5.3.1 Knowledge acquisition and storage

During the study, all participants were provided with some level of training (different training types are discussed in section 5.3.2). The knowledge participants gained from the training tutorials and scenarios was stored in the knowledge base in the form of a two-dimensional matrix. The inference engine used the decision tree induction to identify the general principles or problem-solving strategies based on the individual cases in the knowledge matrix. The knowledge matrix and the decision rules together form the current state of knowledge (Musharraf et al., 2017b). Knowledge acquisition and information retrieval are influenced by the state of the PIFs in the environmental model and operator model.

5.3.2 Environmental model

The environmental model focuses on external factors that define the situation. The external factors used in the experiment were training, visibility, and complexity.

A total of 36 participants took part in the study. Each participant was randomly assigned to one of two groups for training: 1) G1: high level training and 2) G2: low level training.

Both groups received a basic offshore emergency preparedness training. G1 received additional training tutorials and simulated training scenarios regarding situational awareness, alarms, PA announcements, and hazards. Performance of participants in both groups was subsequently tested using simulated testing scenarios.

In the simulated scenarios, visibility was varied at two different levels: *clearly visible* and *blackout*. In clearly visible conditions, there was enough ambient light to perform the assigned task. In the blackout conditions, the visibility was reduced by reducing the available ambient light. However, the participants could use a virtual flashlight in the blackout conditions to have functional visibility.

Complexity was also varied at two different levels: low and high. In low complexity conditions, there was no obstacle in the preferred evacuation route, and the responsibility assigned to the participant was minimal. High complexity situations were created by blocking the escape routes with hazards (i.e. smoke, fire, and explosion), and assigning more responsibilities to the participants. Complexity of the situation was also reflected in the alarm (static versus dynamic) and PA (direct versus indirect). To summarize, complexity was defined in terms of alarm, PA, presence of hazard, obstruction of routes, and amount of responsibility assigned to the participant. For the rest of the paper, these terms are used directly instead of complexity.

5.3.3 Operator model

The operator model focuses on the internal PIFs. It was observed during the study that given the same environmental conditions, participants' response to an emergency may vary depending on the internal PIFs. The internal PIFs assessed in the experiment were knowledge, bias, information use, compliance, and prioritization.

People learn and retain information differently, so given the same training, may have differences in knowledge. This difference is reflected in their problem-solving strategies. It is observed in the experimental data that even participants in the same training group had different knowledge-matrices and hence different strategies for solving problems (Musharraf et al., 2017b).

Bias can be defined as "the tendency of a human to make conclusions based on selected pieces of information while excluding information that does not agree with the conclusion" (Groth & Mosleh, 2012). While some participants were biased and inclined to previous experience, some were not. The *information use* measures an individual's effectiveness in using information presented to him/her. Individuals may favor some information over others due to bias. Some participants showed better information use efficacy than others. *Prioritization* is how an individual orders assigned tasks, or the goals that are to be achieved. Some participants prioritized personal safety over notifying others about the hazard. *Compliance* refers to an individual's commitment to follow directions and policies established by the organization or the industry. Some participants were safety compliant and followed the regulations. Others failed to follow the safety regulations while mustering.

5.3.4 Task networking model

Figure 5.3 shows the task network of general personnel during offshore emergency situations. To exploit the benefits of standard modeling, the task network is presented in the form of an activity diagram in Unified Modeling Language (UML). Activity diagrams are graphical representations of workflows of stepwise activities and actions and serve the same purpose as task network for the case study (Dumas & Ter Hofstede, 2001).

Figure 5.3 captures the main tasks done by general personnel during offshore emergencies. Besides the standard sequence of tasks, erroneous deviations at decision points are shown in the figure using dashed lines. In addition to the decision points, errors can also happen at the following task nodes:

a) Identify alarm and interpret PA – Some participants failed to identify the alarm and consequently went to the wrong final destination. These participants ended up registering at the wrong muster location. Some participants also failed to interpret the PA and misunderstood the presence and/or location of the hazard and obstruction of routes.

b) Re-assess situation – The same errors mentioned in (a) can happen at this task node.

c) Move along selected egress route – Some participants ran along the egress route instead of walking. Participants were trained to close fire/watertight doors while moving along the egress route, but some participants failed to close the doors while egressing.

d) Evaluation of egress paths – In each scenario, participants had to choose an egress route from a set of potential routes. Based on the location of the hazard, some routes may not be safe and must be avoided. Some participants failed to correctly evaluate the egress routes and chose a route that was not tenable. Failure at this node largely depends on failure at the interpretation and/or situation assessment nodes.

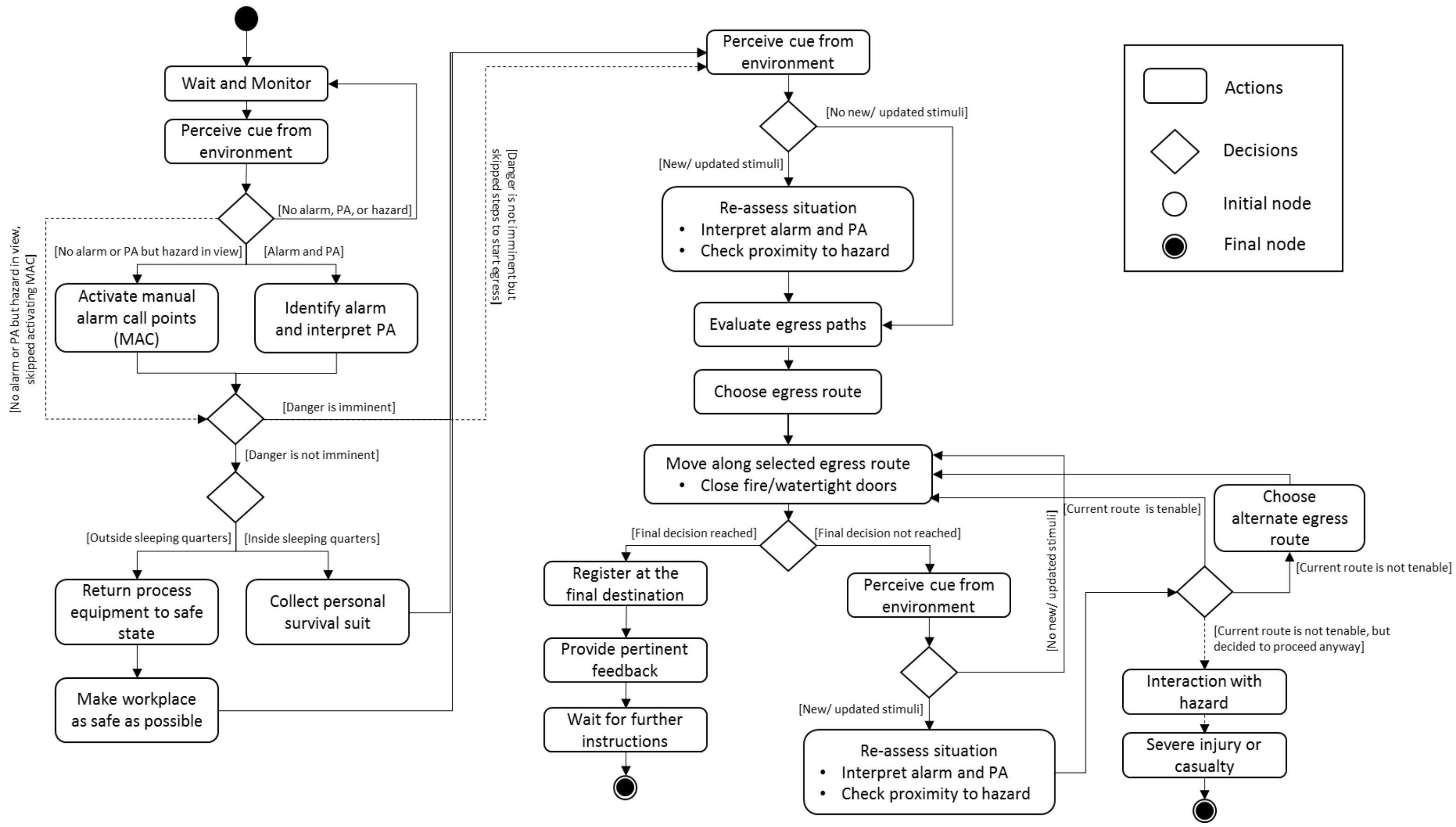

Wait and Monitor
Perceive cue from environment
[No alarm, PA, or hazard]
[No alarm or PA but hazard in view]
[Alarm and PA]
[No alarm or PA but hazard in view, skipped activating MAC]
Activate manual alarm call points (MAC)
Identify alarm and interpret PA
[Danger is imminent]
[Danger is not imminent but skipped steps to start egress]
[Danger is not imminent]
[Outside sleeping quarters]
[Inside sleeping quarters]
Return process equipment to safe state
Collect personal survival suit
Make workplace as safe as possible
Perceive cue from environment
[No new/ updated stimuli]
[New/ updated stimuli]
Re-assess situation
• Interpret alarm and PA
• Check proximity to hazard
Evaluate egress paths
Choose egress route
Move along selected egress route
• Close fire/watertight doors
[Final decision reached]
[Final decision not reached]
Register at the final destination
Provide pertinent feedback
Wait for further instructions
Perceive cue from environment
[No new/ updated stimuli]
[New/ updated stimuli]
Re-assess situation
• Interpret alarm and PA
• Check proximity to hazard
[Current route is tenable]
Choose alternate egress route
[Current route is not tenable]
[Current route is not tenable, but decided to proceed anyway]
Interaction with hazard
Severe injury or casualty
Actions
Decisions
Initial node
Final node

Figure 5.3: Activity diagram of general personnel during offshore emergency situations

5.3.5 Performance shaping model

As shown in the task network in Figure 5.3, at some decision nodes there are possible deviations from the standard behavior path. PSFs define the probabilities of such erroneous deviations. PSFs also define the failure probability at the task nodes mentioned in Section 5.3.4.

For example, during the training, the participants were trained to identify different alarms. Based on the training, the following alarm identification rule will be stored in the memory: two tone sound is a General Platform Alarm (GPA); constant tone sound is a Prepare to Abandon Platform Alarm (PAPA). During an emergency scenario, based on the state of the PIFs, a participant may retrieve and use the alarm identification rule correctly, or can make a mistake and misinterpret the alarm. The PSFs compute the probability of making such an error. They take the PIFs as inputs and uses a BN approach to compute the human error probability.

BNs are acyclic directed graphical models that represent conditional dependencies among a set of random variables Pearl (1988). While performing a task or exercise, errors can occur at different steps of the process. Each error is regarded as the outcome of the joint influence of different PIFs (as depicted in Figure 5.4). In the BN approach, error is the critical node, which depends on several PIFs that can influence the occurrence of the error. For example, misinterpretation of alarm may happen because the knowledge of the personnel is insufficient (PIF_1), or the efficacy of information use of the participant is low

(PIF_2), or both. Figure 5.4 shows the relationship between human error and PIFs. More detail on the BN development and probability calculation can be found in Musharraf et al. (2016) and Musharraf et al. (in press).

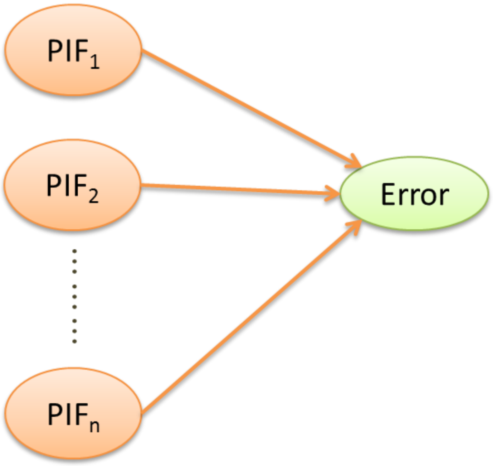


Figure 5.4: Human error while performing a task is the outcome of joint influence of PIF_1 to PIF_n (Musharraf et al., in press)

5.4 Simulation of the developed HBM

The purpose of the proposed HBM is to create human-like agents in the AVERT simulator enable team training. Before integrating the HBM into the agents in AVERT, the HBM was modeled in the IPME simulation framework to clearly define the implementation work scope and identify the technical challenges. During the experimental study, variability across participants was observed in terms of behavior (correct versus erroneous). Based on the aggregated score participants received in the scenarios, they can be classified into 3 broad categories: naïve (0-30%), in-between (31-79%), and competent (80-100%). To capture the same variability in agents, 3 types of operators were created using the HBM. The internal PIFs were used to encode the across-subject variability. Table 5.1 summarizes the internal PIF settings for the different types of operators.

Table 5.1: Internal PIF settings for the different types of operators

Operator type	*Internal PIF settings*
Competent operator	Knowledge = High, Bias = Low, Compliance = High, Efficacy of information use = High, Prioritization = Right
Naïve operator	Knowledge = Low, Bias = High, Compliance = Low, Efficacy of information use = Low, Prioritization = Wrong
In-between operator	In-between operators are representative of the behavior range between the two extremes of competent and naïve, and can be created by using different combinations of internal PIF values. Since *efficacy of information use* was found to be one of the most influential factors (Musharraf et al., 2016), a sample in-between operator was created with *efficacy of information use* as low and all other PIFs in optimal setting. Knowledge = High, Bias = Low, Compliance = High, Efficacy of information use = Low, Prioritization = Right

Section 5.4.1 discusses the probabilistic response generation for different types of operator using an example. Some sample equations, functions, and user interface screens are provided in Appendix A.

5.4.1 Probabilistic response generation

During any emergency scenario, the operators would need to perform the tasks summarized in Figure 5.3 as required. Depending on the state of the PIFs, the operators would either perform the task correctly or make an error. Section 5.3.4 discussed the potential errors. Probability of such error happening is calculated using a BN informed by the empirical data. To demonstrate the probabilistic response generation, a task node “Move along egress route while closing all fire/watertight doors” is used here as an example. In emergency situations, competent operators are expected to exhibit safe behavior and close all fire/watertight doors, whereas the in-between and naïve operators might deviate from the safe behavior and leave fire/watertight doors open. Figure 5.5 shows the BN that captures the causal dependency among the internal PIFs (knowledge, compliance) and human error (leaving fire/watertight doors open).

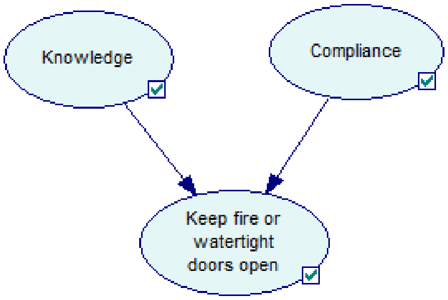


Figure 5.5: Causal dependency among the internal PIFs (knowledge, compliance) and human error (leaving fire/watertight doors open). A comprehensive BN with all PIFs and all potential errors can be found in Musharraf et al. (2016) and Musharraf et al. (in press).

Empirical data collected during the study were used to assess participants' internal PIFs. The assessment helped to inform the conditional probabilities in Figure 5.5. For example, participants with both high knowledge and high compliance had lower error probability (25%) than participants who had one (50%) or none (95%). More details on this can be found in Musharraf et al. (2016). The same idea was used to simulate the behaivor of different operator types. For the competent operators, the internal PIFs are in ideal settings, which means they will have a lower proability of error. For the naïve and in-between operators, the internal PIFs are in non-optimal or sub-optimal states and this increases their probaiblity of error.

The different types of operators were tested in several scenarios. Simulation results for a sample emergency scenario are summarized in the following subsections. In the scenario, there is a fire on the helideck that signals an alarm with a flashing green light and a two-tone sound. The participants must go to their muster station. The PA directs participants to use the primary route. None of the egress routes are obstructed. Values of external PIFs are set per scenario prior to simulation.

It has to be noted that the HBM is stochastic and even for the same combination of PIFs (both external and internal), the behavior of operators may vary across repeated simulation runs. The most frequently observed simulated behaviors of a competent operator, a naïve operator, and an in-between operator over 30 simulation runs of the same scenario are described in sections 5.4.2 to 5.4.4.

5.4.2 Competent operator

It was observed during the simulation that the operator successfully interpreted the audio-visual cues. The alarm type was correctly identified as the general platform alarm (GPA), and final destination was set to muster station. The operator also understood the PA clearly and followed the route direction given in the PA, which was the primary route. The operator walked while moving along the egress route and closed all fire and watertight doors.

5.4.3 Naïve operator

The naïve operator failed to interpret the alarm and the PA. The alarm was mistaken as a prepare to abandon platform alarm (PAPA) and the final destination was set as lifeboat station, which was incorrect. While choosing an egress route, the operator did not pay attention to the PA, rather went with the route s/he was most familiar with, which was the secondary route. The operator ran along the egress route and did not close the fire or watertight doors.

5.4.4 In-between operator

The in-between operator interpreted the alarm and PA correctly and set the final destination as the muster station. However, as the operator's efficacy of information use was low, s/he did not choose a route based on the PA. His/her strategy was to choose an egress route based on the final destination (take the primary route if the final destination is the muster station and try the secondary route in case the final destination is the lifeboat station). Though the strategy led the operator to the correct choice of route in this particular scenario, this strategy would fail in the case a route is obstructed and understanding the PA is vital.

The operator walked while moving along the egress route and closed all fire and watertight doors.

Figure 5.6 summarizes the observed erroneous behavior of different operator types over 30 simulation runs. As shown in Figure 5.6, the naïve operators exhibit the most erroneous behavior. The competent operators and in-between operators occasionally commit errors, but at a much lower rate compared to the naïve operators. One of the commonly observed errors in the ideal operators was leaving fire/watertight doors open under the pressure of emergency. This was also one of the most frequent mistakes by participants with high performance scores (80-100%).

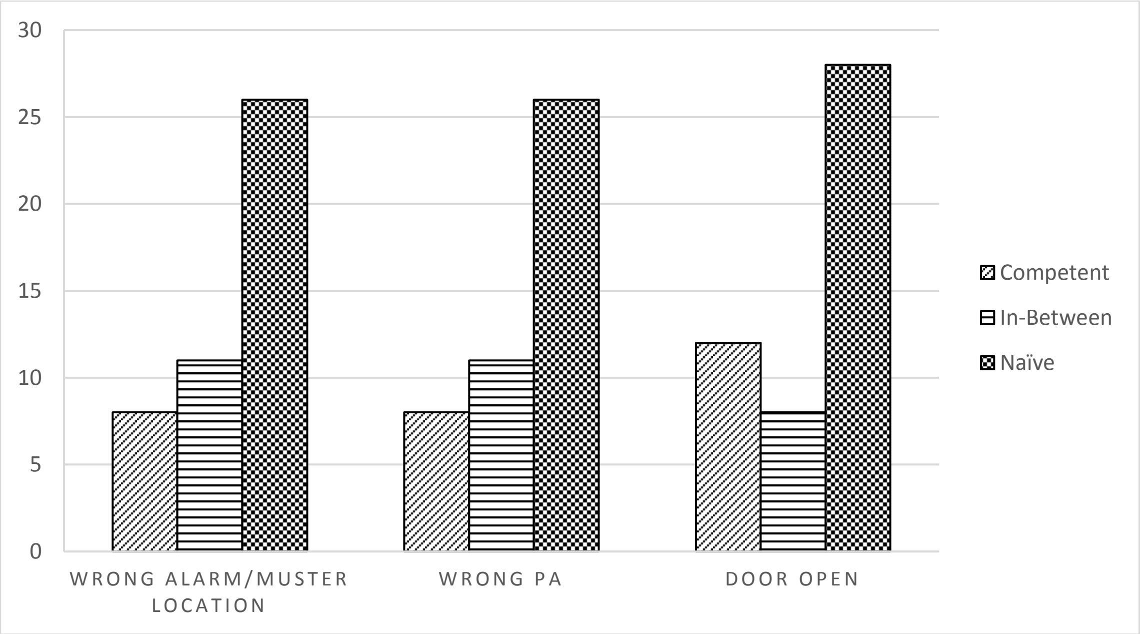


Figure 5.6: Observed erroneous behavior of different operator types over 30 simulation runs

As demonstrated by the results, the HBM is capable of modeling the across-subject variability, which will enable the creation of a heterogeneous training environment where each entity behaves somewhat differently. Training programs in AVERT can benefit from this. It will prepare the trainee for the variability that is inherent in human teammates. Trainees can gain experience of working with or commanding teams with varying skill levels. Variability can also help to keep the trainee more engaged during the simulation training by offering novel challenging situations.

5.5 Conclusion and future work

This paper presents a computational model of human behavior during offshore emergency situations. To understand the emergency response of general personnel, an experimental study was conducted using a virtual environment. The variability observed across scenarios was captured in the computational model using external PIFs. To encode the variability observed across different scenarios, external PIFs were used. To encode the variability observed across different subjects given the same scenario, internal PIFs were used. Simulation results show that the model can capture human behavior variability given different states of PIFs.

The next step is to integrate the HBM in AVERT to create intelligent agents. Results from the IPME simulation framework will guide the integration and implementation process. Validation of the HBM is out of the scope of this paper and is presented separately in Musharraf et al. (2017). The work on validation illustrates how the simulated behaviors correlate to the actual behaviors observed during the experimental study.

Acknowledgements

The authors acknowledge with gratitude the support of the Research & Development Corporation of Newfoundland and Labrador and the NSERC-Husky Energy Industrial Research Chair in Safety at Sea.

References

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In *Psychology of learning and motivation* (Vol. 2, pp. 89-195). Elsevier.

Baraldi, P., Conti, M., Librizzi, M., Zio, E., Podofillini, L., & Dang, V. (2009). A Bayesian network model for dependence assessment in human reliability analysis. *Proceedings of the Annual European Safety and Reliability Conference (ESREL)*, (pp. 223-230). Prague.

Cacciabue, P. C., Decortis, F., Drozdowicz, B., Masson, M., & Nordvik, J. P. (1992). COSIMO: a cognitive simulation model of human decision making and behavior in accident management of complex plants. *Systems, Man and Cybernetics, IEEE Transactions*(22(5)), 1058-1074.

Chang, Y. H., & Mosleh, A. (2007a). Cognitive modeling and dynamic probabilistic simulation of operating crew response to complex system accidents: Part 1: Overview of the IDAC Model. *Reliability Engineering & System Safety, 92*(8), 997-1013.

Chang, Y. H., & Mosleh, A. (2007b). Cognitive modeling and dynamic probabilistic simulation of operating crew response to complex system accidents. Part 2: IDAC performance influencing factors model. *Reliability Engineering & System Safety*, *92*(8), 1014-1040.

Chang, Y. H., & Mosleh, A. (2007c). Cognitive modeling and dynamic probabilistic simulation of operating crew response to complex system accidents: Part 5: Dynamic probabilistic simulation of the IDAC model. *Reliability Engineering & System Safety*, *92*(8), 1076-1101.

Dang, V. N. (1996). *Modeling operator cognition for accident sequence analysis: development of an operator-plant simulation.* Doctoral dissertation, Massachusetts Institute of Technology.

Dang, V., & Stempfel, Y. (2012). Evaluating the Bayesian belief network as a human reliability model - the effect of unreliable data. *Proceedings of the international conference on probabilistic safety assessment and management and the European safety and reliability conference PSAM 11 & ESREL 2012.* Helsinki, Finland.

Duffy, V. G. (2008). *Handbook of digital human modeling: research for applied ergonomics and human factors engineering.* CRC press Taylor & Francis Group.

Dumas, M., & Ter Hofstede, A. H. (2001). UML activity diagrams as a workflow specification language. *International Conference on the Unified Modeling Language* (pp. 76-90). Berlin Heidelberg: Springer .

Fenton, N., & Neil, M. (2012). *Risk assessment and decision analysis with Bayesian networks.* CRC Press.

Flin, R. (1997). Crew resource management for teams in the offshore oil industry. *Team Performance Management, 3*(2), 121-129.

Goerger, S. R., McGinnis, M. L., & Darken, R. P. (2005). A validation methodology for human behavior representation models. *The Journal of Defense Modeling and Simulation, 2*(1), 39-51.

Groth, K. M., & Mosleh, A. (2012). A data-informed PIF hierarchy for model-based Human Reliability Analysis. *Reliability Engineering & System Safety, 108*, 154-174.

Han, J., Kamber, M., & Pei, J. (2011). *Data mining: concepts and techniques.* Elsevier.

House, A. W., Smith, J., MacKinnon, S., & Veitch, B. (2014). Interactive simulation for training offshore workers. *Oceans'14 MTS/IEEE Conference* (pp. 1-6). St. John's, NL: IEEE.

Jones, R. M., Laird, J. E., Nielsen, P. E., Coulter, K. J., Kenny, P., & Koss, F. V. (1999). Automated intelligent pilots for combat flight simulation. AI magazine. *AI magazine*, *20*(1), 27.

Li, Y. (2013). *Modeling and simulation of operator knowledge-based behavior.* University of Maryland.

Li, Y., & Mosleh, A. (in press). Modeling and simulation of crew to crew response variability due to problem-solving styles. *Reliability Engineering & System Safety*.

McNally, B. H. (2005). An approach to human behavior modeling in an air force simulation. *Proceedings of the 37th conference on Winter simulation* (pp. 1118-1122). Winter Simulation Conference.

Moffat, S., & Crichton, M. (2015). Investigating non-technical skills through team behavioral markers in oil and gas simulation-based exercises. *Procedia Manufacturing, 3*, 1241-1247.

Musharraf, M., Khan, F., & Veitch, B. (2017). Validating human behavior representation model of general personnel during offshore emergency situations. *Submitted for review to the Journal of Fire Technology, special issue on Fire Evacuation Modeling*.

Musharraf, M., Smith, J., Khan, F., Veitch, B., & MacKinnon, S. (2016). Assessing offshore emergency evacuation behavior in a virtual environment using a Bayesian Network approach. *Reliability Engineering & System Safety, 152*, 28-37.

Musharraf, M., Smith, J., Khan, F., Veitch, B., & MacKinnon, S. (2017a). Human performance data collected in a virtual environment. *Data in Brief*, 213-215.

Musharraf, M., Smith, J., Khan, F., Veitch, B., & MacKinnon, S. (2017b). Identifying route selection strategies in offshore emergency situations using Decision Trees: A step towards adaptive training. *Manuscript submitted for review*.

Musharraf, M., Smith, J., Khan, F., Veitch, B., & MacKinnon, S. (in press). Incorporating individual differences in human reliability analysis: an extension to the virtual experimental technique. *Safety Science*.

O'Connor, P., & Flin, R. (2003). Crew resource management training for offshore oil production teams. *Safety Science, 41*(7), 591-609.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference.* San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Podofillini, L., & Dang, V. N. (2013). A Bayesian approach to treat expert-elicited probabilities in human reliability analysis model construction. *Reliability Engineering & System Safety, 117*, 52-64.

Ritter, F. E., Kim, J. W., Morgan, J. H., & Carlson, R. A. (2012). *Running behavioral studies with human participants: A practical guide.* Sage Publications.

Sampson, S. R., & Ripingill Jr, A. E. (2003). *System and method for training in military operations in urban terrain.* Washington, DC: U.S. Patent and Trademark Office.

Sharma, S. (2009). Avatarsim: A multi-agent system for emergency evacuation simulation. *Journal of Computational Methods in Sciences and Engineering, 9*(1, 2S1), 13-22.

Smith, J. (2015). *The effect of virtual environment training on partcipant competence and learning in offshore emergency egress scenarios.* Master of Engineering Thesis, Memorial University of Newfoundland, St. John's.

Sundaramurthi, R., & Smidts, C. (2013). Human reliability modeling for the Next Generation System Code. *Annals of Nuclear Energy, 52*, 137-156.

Thow-Yick, L. (1994). The basic entity model: A fundamental theoretical model of information and information processing. *Information Processing & Management, 30*(5), 647-661.

Trucco, P., & Leva, M. C. (2007). A probabilistic cognitive simulator for HRA studies (PROCOS). *Reliability Engineering & System Safety, 92*(8), 1117-1130.

Van Diggelen, J., Muller, T., & Van den Bosch, K. (2010). Using artificial team members for team training in virtual environments. In *Intelligent Virtual Agents* (pp. 28-34). Springer Berlin Heidelberg.

Wray, R. E., & Laird, J. E. (2003). Variability in human behavior modeling for military simulations. *Proceedings of Behavior Representation in Modeling and Simulation Conference (BRIMS).*

6. VALIDATING HUMAN BEHAVIOR REPRESENTATION MODEL OF GENERAL PERSONNEL DURING OFFSHORE EMERGENCY SITUATIONS

Mashrura Musharraf*, Faisal Khan, Brian Veitch

Centre for Risk, Integrity and Safety Engineering (C-RISE)

Faculty of Engineering & Applied Science,

Memorial University of Newfoundland,

St John’s, Newfoundland and Labrador, Canada A1B 3X5

* Correspondence author: Tel: + 1 709 864 6764; Email: mm6414@mun.ca

Co-authorship statement

A version of this manuscript has been submitted for review in the Journal of Fire Technology (Special issue on Fire Evacuation Modeling). The lead author Mashrura Musharraf performed the literature review, designed and conducted the validation experiment, performed the data analysis, and prepared the draft of the manuscript. Co-authors Faisal Khan and Brian Veitch supervised the experiment, reviewed and revised the results, and provided feedback on the manuscript. Mashrura Musharraf revised the manuscript based on the co-authors’ feedback.

Abstract

With the advancement of simulation-based training, intelligent agents that can display human-like behavior have become common. From military combat simulations to nuclear power plant simulation, agents have been widely used to facilitate team training (as team mates, opponents, or both). Credibility of these agents is vital to ensure a sound training process. Credibility of the agents largely depends on the credibility of the underlying human behavior representation model (HBM). This is why validation of the HBM is necessary to ensure realistic agent behavior. However, the non-deterministic nature of the HBM and the subjectivity in experts' judgment during the validation process make HBM validation more challenging compared to physics based models. This paper presents the validation process of an HBM of general personnel created for use in an offshore emergency training simulator. Three types of agents (naïve, in-between, and ideal) are created in the simulator using the HBM. The paper discusses the use of empirical evidence as referents, along with subject matter experts (SME). A two-level three factor experiment was conducted using 36 participants. Several performance metrics were collected during the experiment, including route selected for evacuation, time to muster, time spent running, interaction with fire doors and watertight doors, interaction with hazards, and reporting to the muster station. Data collected during the experimental study have been used in this paper to demonstrate how the use of empirical evidence can facilitate HBM validation. High-level tasks performed during HBM validation are discussed in detail. Special emphasis is given on acceptability criteria testing to ensure that the HBM performs adequately under different operating conditions. Results show that the proposed HBM

meets the acceptability criteria requirement for all types of agents. In general, the ideal agents exhibited safe behavior during offshore emergency egress, whereas the naïve and in-between agents showed erroneous behavior at times. For example, during the simulation runs of a critical emergency scenario where the primary egress route was obstructed by a hazard, the ideal agents either waited and listened to the public address (PA) announcement and followed an alternative egress route (60% cases), or they initially chose their preferred route but re-routed immediately after encountering the hazard (40% cases). In all cases, the in-between agents started with their preferred route and re-routed after encountering the hazard, and the naïve agents proceeded with their preferred route even when the route was compromised.

6.1 Introduction

Software agents, or computer generated forces (CGFs), are extensively used in a wide range of team training simulations. This includes military applications for training and rehearsal for combat situations (Karr et al., 1997). The use of virtual crew is also common in aviation and nuclear power plant simulation training (Chang & Mosleh, 2007). Realism of agents in any platform largely depends on the sophistication of the underlying human behavior representation model (HBM) (Smith, 1998). HBMs are computational models that probabilistically simulate human behavior in different conditions. To ensure agents have an acceptable level of credibility, the underlying HBM must be validated. However, due to the non-deterministic nature of the HBM, it is difficult to ensure that the HBM adequately represents the behavior it was designed to exhibit, and captures the behavior variability expected in the non-linear environment into which it will be integrated.

The difficulty has led face validation to become the most common form of validation for HBM (Anon., 2001b). Face validation is a method that is widely used to validate interactive real-time virtual simulations where user interaction bears significant importance for the simulation to be accredited (Sokolowski & Banks, 2010). As defined in the Defense Modeling and Simulation Office's (DMSO) Recommended Practices Guide (RPG), “in face validation technique, a subject matter expert (SME) drives through the scenario space by issuing commands or changing the simulating situation, observes the resulting behavior, and determines, often qualitatively, whether that behavior meets a user’s requirements for realism”. Face validation is listed as the least reliable and least completed HBM validation process (Anon., 2001b). SMEs’ judgments are drawn mostly from their own experience and can be biased. It is also hard to ensure their levels of consistency and accuracy when evaluating human performance versus simulated human behavior.

This paper presents the HBM validation of general personnel created to be used in an offshore emergency preparedness training simulator. The simulator is called the all-hands virtual emergency response trainer (AVERT). It is modeled after an offshore oil platform with high levels of graphical detail of the environment and can create credible emergency scenarios by introducing hazards such as blackouts, fires, and explosions. The current configuration of AVERT is intended to train general personnel in safe work practices (Smith et al., 2017). As of now, only individual training is enabled in AVERT. The HBM of the general personnel is the first step towards creating software agents to enable team training (Musharraf et al., 2017). To make sure that the HBM will contribute towards a sound team training process, the model must be validated before use. This paper describes the high-level tasks performed to validate the model. Besides SMEs' judgments, empirical evidence is used in this paper during the validation process.

Figure 6.1 summarizes the previous work leading to validation. An experimental study was conducted in AVERT to gather empirical evidence for – 1) an HBM development and 2) the HBM validation. Details of the HBM development have been discussed in Musharraf et al. (2017). The focus of the paper is the latter. Before going into the details of the validation process, Section 6.2 gives an overview of the experimental study and the HBM. Section 6.3 discusses the validation process in detail. Special attention is paid in this paper to the result validation. Section 6.4 lists the challenges faced during the validation. Section 6.5 summarizes and concludes the paper.

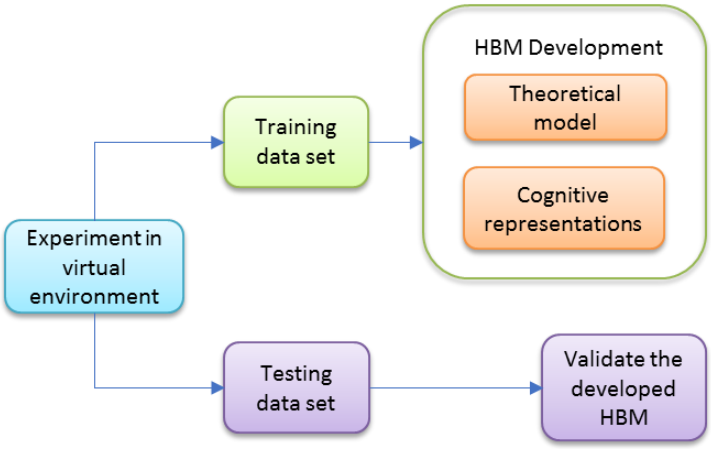

HBM Development
Theoretical model
Cognitive representations
Training data set
Experiment in virtual environment
Testing data set
Validate the developed HBM

Figure 6.1: Data collected during an experimental study conducted in AVERT were divided into two sets. The training data set was used to develop the HBM, and the testing data set was used to validate it.

6.2 Overview of the experimental study and HBM

6.2.1 Experimental study

This section provides an overview of the experimental study and the HBM of general personnel in the context of offshore emergencies. General personnel are individuals whose responsibility during an emergency is to muster at their designated muster stations (Smith et al., 2017).

The experiment was conducted using the AVERT simulator. A total of 36 participants took part in the study, each with the goal to learn how to successfully muster during offshore emergency situations. Each participant was randomly assigned to one of the two groups 1) G1: high level training, or 2) G2: low level training. The sample size was determined by an iterative process using equation 1.

$$n = \frac{(t_{\alpha/2})^2 \sigma^2}{B^2} \tag{1}$$

Here, n is the sample size, $t_{\alpha/2}$ is the t-score for a 95% confidence interval, σ is the estimated standard deviation informed by a prior study (Bradbury-Squires, 2013), and B is the acceptable margin of error. Originally, a sample of 40 participants was targeted for a confidence interval of 95% where a margin of error of (+/-) 10% was considered acceptable. As the study investigated the effect of training, a minimum of 15 participants in each training group was required. From the original recruitment, 4 participants withdrew for various reasons. This resulted in a sample size of 36 participants, increasing the margin of error to (+/-) 11%. 17 participants were assigned to G1, and 19 participants were assigned to G2. A more detailed discussion on the sample size can be found in Smith (2015).

Among the 36 participants, 27 were males and 9 were females. The age range was from 19 to 39 years. Participants were recruited using samples of convenience (Ritter et al., 2012). Naïve participants, mostly university students, were recruited for the study. The participants were not aware of any detail of the experimental design, they were not employed in the offshore oil industry, and therefore they were not familiar with the offshore platform.

Each participant attended 3 sessions on 3 separate days. Both groups received basic offshore emergency preparedness training in session 1. G1 then received 4 practice scenarios in the same session. In session 2 and session 3, G1 received additional training tutorials and practice scenarios (4 scenarios per session) regarding situation awareness, alarms, public address (PA) announcements, and hazards. G2 did not receive any further training or practice opportunities in session 2 and 3. Performance of participants in both groups was tested in each session using 4 test scenarios. In total, participants in G1 performed in 24 scenarios and participants in G2 performed in 12 scenarios.

The simulated scenarios ranged from drills that required the participants to go to their primary muster station, to more complex emergency scenarios that required the participants to avoid hazards blocking their egress routes and muster at their lifeboat stations (House et al., 2014). Appendix B includes the schematic diagrams of different egress routes. The scenarios were designed such that the effect of different performance influencing factors (PIFs) on individuals’ performance during offshore emergency situations could be investigated. PIFs are factors that can specifically decrement or improve human performance during a task (e.g. complexity, visibility) (Blackman et al., 2008). PIFs are also referred to as behavior moderators.

Besides training, the other two external PIFs investigated in the study are visibility and complexity. Both visibility and complexity were varied into two levels across the scenarios (more on this is discussed in Section 6.2.2.1). Several performance metrics were collected during each scenario for each participant. The metrics included route selected for evacuation, time to muster, time spent running, interaction with fire doors and watertight doors, interaction with hazards, and reporting to the muster station. Data collected in different scenarios were divided into two sets. Data collected in some scenarios (for all participants) were used to develop the HBM. Data collected in the remaining scenarios (for all participants) were used to validate the HBM.

Section 6.2.2 discusses the dynamic response generation for a given situation. The probabilistic aspects of the HBM are described in this section. Section 6.2.3 describes how knowledge gained from training and experience form the current state of knowledge. It also explains the underlying information processing approach.

6.2.2 Dynamic response model

The dynamic response model has been documented in detail in the authors' previous work (Musharraf et al., 2017). This section presents a brief overview of the model to facilitate the understanding of the validation process, which is the focus of this paper.

The dynamic response model consists of four component models: an environment model, an operator model, a performance shaping model, and a task network model. The environmental model defines the situation, or environment, using external PIFs. The operator model defines the characteristics of the operator using internal PIFs. The task network model graphically represents the sequence of tasks performed by the operator. The performance shaping model generates the rules of behavior of the operators depending on the state of different PIFs and the current state of knowledge.

Figure 6.2 shows the interaction between the external world and the component models. At any time 't' the state of PIFs in the environmental model and operator model are defined based on the events happening in the external world. The state of the internal and external factors defines the operator's state of mind. The PIFs also define how information is gained from training and experience. The performance shaping function (PSF) model takes the operator's state of mind and current state of knowledge into account, and generates the behavior rules that govern the operator's response during tasks.

The following subsections introduce the four component models. The process of knowledge acquisition and retrieval is modeled based on the idea of knowledge based system architecture (Negnevitsky, 2005). This is discussed separately in Section 6.2.3.

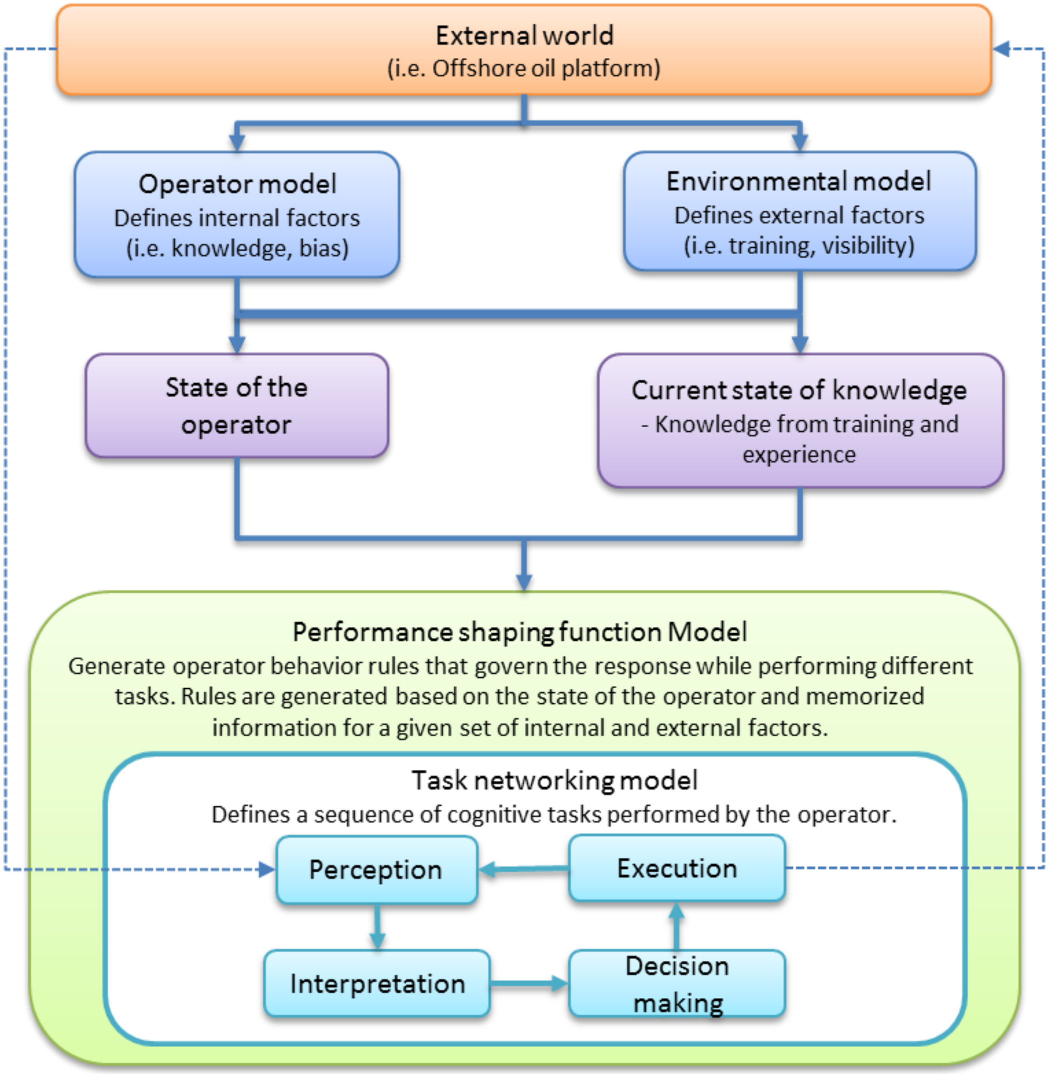

External world
(i.e. Offshore oil platform)
Operator model
Defines internal factors
(i.e. knowledge, bias)
Environmental model
Defines external factors
(i.e. training, visibility)
State of the operator
Current state of knowledge
- Knowledge from training and experience
Performance shaping function Model
Generate operator behavior rules that govern the response while performing different tasks. Rules are generated based on the state of the operator and memorized information for a given set of internal and external factors.
Task networking model
Defines a sequence of cognitive tasks performed by the operator.
Perception
Execution
Interpretation
Decision making

Figure 6.2: Interaction between the external world and the four component models: operator model, environmental model, performance shaping function model, and task networking model (after Musharraf et al., 2017).

6.2.2.1 Environmental model

The environmental model includes external PIFs that define the situation or environment the individual is in. The external PIFs that were used in the experiment are training, visibility, and complexity.

As noted previously, each participant was assigned to either G1 or G2 for training. During the study, G1 received more advanced training compared to G2.

In the simulated scenarios, visibility was varied at two different levels: clearly visible and blackout. In clearly visible conditions, there was enough ambient light to perform the assigned task. In the blackout conditions, the visibility was reduced by reducing the available ambient light. However, the participants could use a virtual flashlight in the blackout conditions to have functional visibility.

Complexity was also varied at two different levels: low and high. In low complexity conditions, there was no obstacle in the preferred evacuation route, and the responsibility assigned to the participant was minimal. High complexity situations were created by blocking the escape routes with hazards (i.e. smoke, fire, and explosion), and assigning more responsibilities to the participants. Complexity of the situation was also reflected in the alarm (static versus dynamic) and PA (direct versus indirect). To summarize, alarm, PA, presence of hazard, obstruction of routes, and amount of responsibility assigned to the participant, together defined complexity.

6.2.2.2 Operator model

Operator model focuses on the internal PIFs (Groth & Mosleh, 2012). It was observed during the study that given the same environmental conditions, participants' response to an emergency may vary depending on the internal PIFs. The internal PIFs investigated in the experiment were knowledge, bias, information use, compliance, and prioritization.

It was observed in the experimental data that given the same training, people focus on different scenario attributes before making a decision. This resulted in different knowledge-matrices and hence different problem-solving strategies across participants (Musharraf et al., 2017a).

During the study, while some participants were biased (inclined to previous experience), some were not. Some participants effectively used the information presented to them, while some failed to do so. Some participants prioritized personal safety over notifying others about the hazard. Some participants were safety compliant and followed the regulations; others failed to follow the safety regulations under the pressure of emergency.

6.2.2.3 Task networking model

Task network modeling focuses on understanding the tasks that need to be simulated. The task network model graphically represents the sequence of tasks performed by an operator.

Figure 6.3 shows the task network of general personnel during offshore emergency egress. Besides the standard sequence of tasks, erroneous deviations at decision points are shown in the figure using dashed lines. Additional to the decision points, errors can also occur during the following: identifying the alarm and interpreting the PA, re-assessing the situation (i.e. interpreting updated alarm and PA, checking proximity to hazard), moving along the selected egress route, and evaluating egress paths.

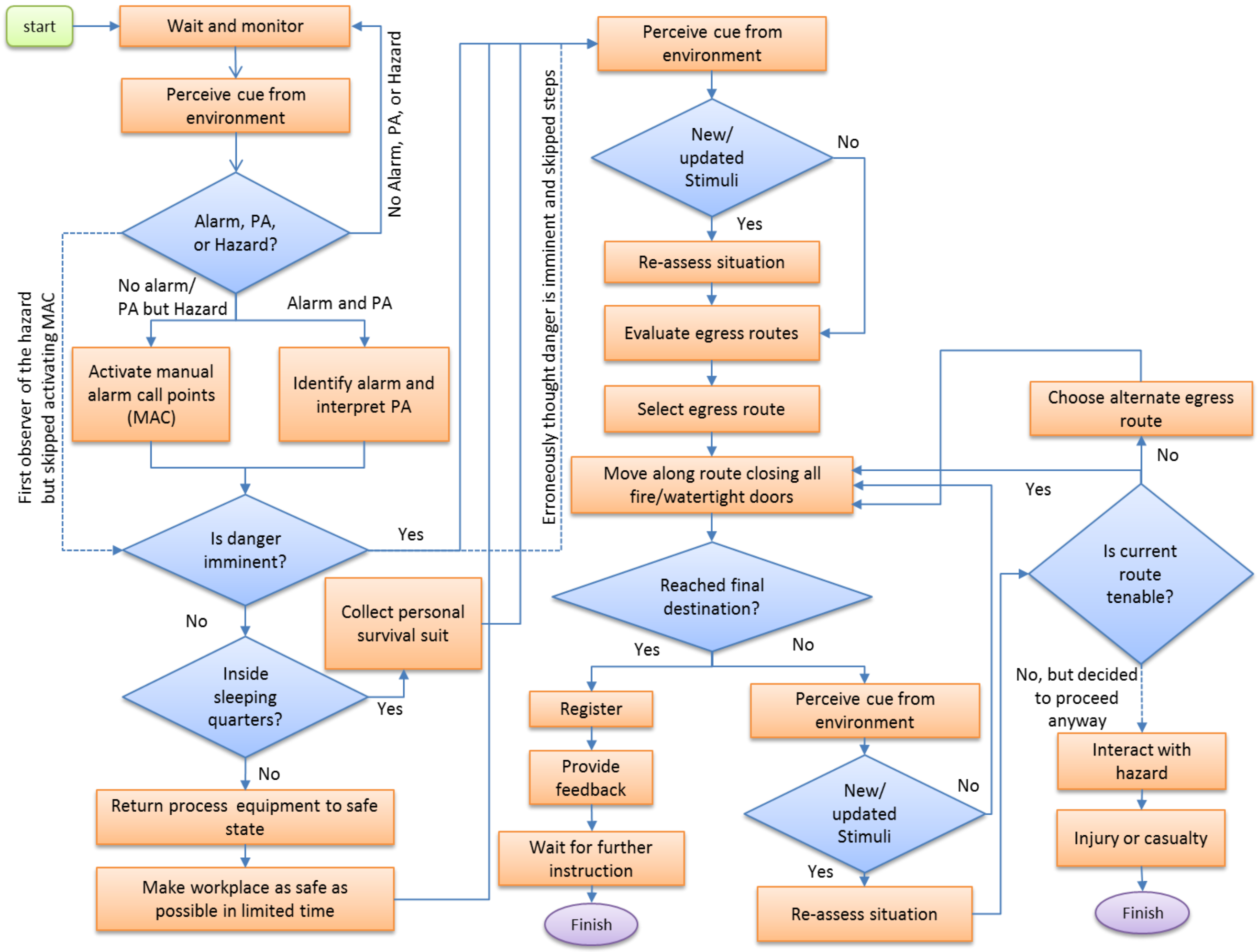

start
Wait and monitor
Perceive cue from environment
Alarm, PA, or Hazard?
No Alarm, PA, or Hazard
No alarm/ PA but Hazard
Alarm and PA
Activate manual alarm call points (MAC)
Identify alarm and interpret PA
First observer of the hazard but skipped activating MAC
Is danger imminent?
Yes
No
Inside sleeping quarters?
Yes
Collect personal survival suit
No
Return process equipment to safe state
Make workplace as safe as possible in limited time
Erroneously thought danger is imminent and skipped steps
Perceive cue from environment
New/ updated Stimuli
No
Yes
Re-assess situation
Evaluate egress routes
Select egress route
Move along route closing all fire/watertight doors
Reached final destination?
Yes
No
Register
Provide feedback
Wait for further instruction
Finish
Perceive cue from environment
New/ updated Stimuli
No
Yes
Re-assess situation
Choose alternate egress route
No
Yes
Is current route tenable?
No, but decided to proceed anyway
Interact with hazard
Injury or casualty
Finish

Figure 6.3: Sequence of tasks performed by general personnel during offshore emergency egress. Solid lines represent standard sequence, whereas dashed line represent possible erroneous deviation.

6.2.2.4 Performance shaping model

This model includes a set of performance shaping functions (PSFs). The PSFs generate the rules of behavior that govern the performance of general personnel while performing different tasks (Chang & Mosleh, 2007). The response of general personnel depends on the state of the operator and the current state of knowledge. The PSFs take the state of the operator and current state of knowledge into account and generate the associated operator response for a given set of PIFs.

For example, during the training, the participants were trained to identify different alarms. Based on the training, the following alarm identification rule will be stored in the memory: a two-tone sound is a General platform alarm (GPA); a constant tone sound is a Prepare to abandon platform alarm (PAPA). During an emergency scenario, based on the state of the PIFs, a participant may retrieve and use the alarm identification rule correctly, or can make a mistake and misinterpret the alarm. The PSFs compute the probability of making such an error. They take the PIFs as inputs and use a Bayesian network (BN) approach to compute the human error probability.

Figure 6.4 shows the relationship between human error and PIFs. Here, error is modeled as the outcome of joint influence of PIFs. More detail on the BN development and probability calculation can be found in Musharraf et al. (2016) and Musharraf et al. (2017b).

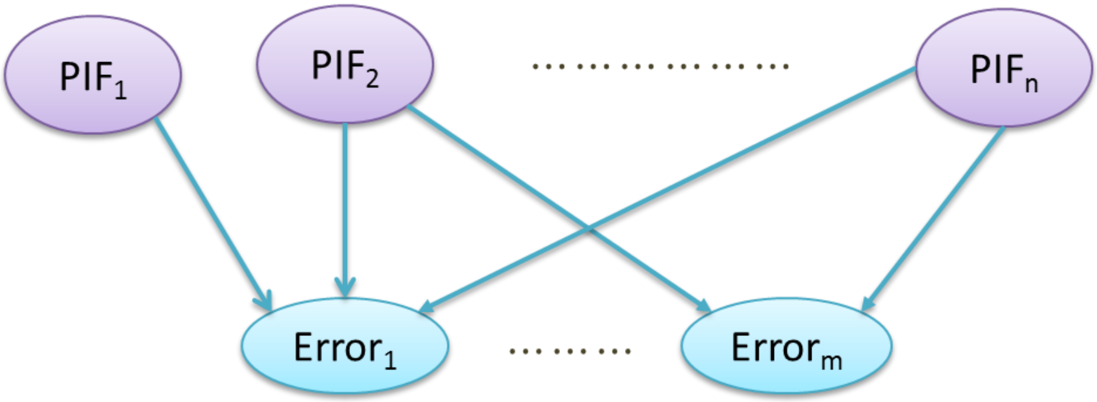


Figure 6.4: Human error$(Error_1\ to\ Error_m)$ **while performing a task is the outcome of joint influence of performance influencing factors** $(PIF_1\ to\ PIF_n)$.

6.2.3 Memory structure and cognitive functions

In the HBM, the knowledge-matrices and problem-solving strategies are stored in a database representative of human memory. This data base is referred to as the memory structure. Organization of information in the memory structure is modeled based on the idea of knowledge based system architecture (Kendal & Creen, 2007). The underlying assumption is that people's response to emergency situations depends in part on the knowledge they have stored in their memory. The three main components of the memory structure are: knowledge base, working memory, and inference engine. The working memory has a finite capacity and stores the information relevant to the problem that is currently being solved. The knowledge base has a theoretically infinite capacity and stores all the knowledge gained through training and experience. The inference engine is an intermediate memory space that contains generalized decision rules based on the content in the knowledge base.

The process involved in creating generalizations from observed phenomena or principles is referred to as induction. Among the available induction approaches, decision tree is used in the development of the HBM. In decision tree induction, data are divided into certain groups based on the information gain of the attributes, with the goal that the examples in any particular group will belong to the same class (Han et al., 2011). Decision tree offers a visual representation of the reasoning process and has valuable diagnostic capabilities. Compared to other methods, such as artificial neural networks or support vector machines, decision trees can be constructed relatively quickly. Other benefits of decision tree, which are particularly important for this paper, are that it does not require any prior assumptions about the data and can work with limited data compared to other techniques (Duffy, 2008). More on decision tree induction can be found in Musharraf et al. (2017a).

During information processing to solve the problem at hand, cues are perceived from the environment. The perceived cues are interpreted to form a calling condition. A calling condition is a set of variables that takes values from a defined set (Thow-Yick, 1994). If a solution to the current calling condition is available in the working memory, it is immediately retrieved. Otherwise the calling condition is transferred to the inference engine to find a solution. If a decision rule that matches the current calling condition is found, the solution is retrieved and sent to working memory to act upon. If no matching decision rule is found, the calling condition is sent to the knowledge base. More abstract relationships between the calling condition and solutions stored in the knowledge matrix may be analyzed at this stage. Once a solution is found, the next step is to execute a series of actions to implement it.

6.3 Validation process for the HBM

As listed in DMSO's "Validation of Human behavior representation", any HBM validation process needs to perform the following high-level tasks:

1. Collect as complete a set of requirements and acceptability criteria as possible.
2. Identify referent(s) to assess the credibility of the HBM model.
3. Validate the HBM’s conceptual model using referent and requirements.
4. Validate the HBM’s knowledge base using referent and requirements.
5. Analyze the HBM’s conceptual model and knowledge base to identify complex areas of the model that need attention in future validation activities.
6. Validate the integrated HBM using referent and requirements. This is called result validation and involves acceptability criteria testing by exercising testing scenarios to ensure that the integrated HBM performs adequately under different operating conditions. Before the result validation, the HBM must be integrated in the virtual environment. In this paper, the model is integrated in AVERT to create software agents performing as general personnel. The complex areas identified in the previous step are used at the result validation step to design credible test scenarios.

The following subsections will describe each step of the HBM validation in detail.

6.3.1 Collect as complete a set of requirements and acceptability criteria as possible

The first step of validating the HBM was to make a list of requirements that will set the foundation of validation. Since listing a complete set of requirements was challenging in an early stage, attention was paid to fulfill the minimum requirements first. The minimum requirements include a detailed task analysis of the person the HBM is intended to represent and the definition of level of skills that the simulation must include. A list of PIFs were defined as well. Fidelity of the simulated behavior largely depends on how well the effect of these PIFs are modeled.

Detailed task analysis of general personnel represented by the HBM: A detailed task analysis of general personnel during offshore emergency situations was done based on (DiMattia, Khan, & Amyotte, 2005). The corresponding activity diagram and possible deviations from ideal behavior are presented in Section 6.2.2.3. For an agent to be acceptable, it must be able to perform all the tasks outlined in Figure 6.3. Like humans, agents can also make mistakes while performing a task.

The AVERT configuration during the experimental study did not allow performing the following tasks – returning process equipment to safe state, making workplace as safe as possible, and collecting personal survival suit. Though some of these functionalities are in the current configuration and hence integrated in the agents, they will not be discussed further in this paper since a benchmark empirical evidence is not available for comparison.

Level of skills of general personnel: The HBM was developed to represent three different levels of skill: ideal personnel, naïve personnel, and in-between personnel. The levels of skill were defined in terms of internal PIFs. The internal PIFs and corresponding states are listed in Table 6.1. Details of the internal PIFs were discussed in Section 6.2.2.2. Different levels of skill were achieved by varying the state of the internal PIFs.

Table 6.1: Internal PIFs and corresponding possible states

Internal PIFs	***Possible states***
Bias	Yes, No
Compliance	High, Low
Efficacy of information use	High, Low
Knowledge	High, Low
Prioritization	Right, Wrong
Preference of route	Primary route, Secondary route

Ideal agents were created by setting the internal PIFs in the following way: knowledge as high, bias as low, compliance as high, efficacy of information use as high, and prioritization as right. To create naïve agents, internal PIFs were set in the exact opposite way: knowledge as low, bias as high, compliance as low, efficacy of information use as low, and prioritization as wrong. In-between agents are representative of the behavior range between the two extremes of ideal and naïve. These agents can be created by using different combinations of the internal PIF values. The example in-between agents used in this paper were created by setting knowledge as low, bias as low, compliance as high, prioritization as right, and efficacy of information use as low.

PIFs or behavior moderators: Two kinds of PIFs – internal and external - are used in this paper. Internal PIFs were presented in Table 6.1. Besides internal PIFs, a list of external PIFs are used to delimit the range of situations in which the HBM is expected to perform. Table 6.2 provides a list of external factors and the corresponding possible states. Details of the external factors were discussed in Section 6.2.2.1.

Table 6.2: External PIFs and corresponding possible states

External PIFs	***Possible states***
Alarm (Audio & visual cues)	GPA, PAPA
Route direction in PA	None, Primary route, Secondary route
Obstruction of routes	None, Primary route, Secondary route
Presence of hazard	Yes, No
Visibility	High, Low

6.3.2 Identify referent(s) to assess the credibility of the HBM

Referent refers to a codified body of knowledge about a thing being simulated. During validation, referent provides the information with which the simulation outcomes are compared. Among the six different model correspondences listed in the “Key concepts of VV&A”, domain correspondence is used in this paper (RPG: Reference Document - Key Concepts of VV&A, 2001). Two domain referents were used at different steps of the HBM validation. At the earlier stage, inputs from SMEs were used to validate the conceptual model. As the validation progressed, empirical evidence collected during the experimental study discussed in Section 6.2.1 was used for knowledge base validation and integrated HBM validation.

6.3.3 Validate the HBM’s conceptual model using referent and requirements

The conceptual model of the HBM of general personnel consists of the followings:

1. Tasks the HBM must perform: As discussed in Section 6.3.1, a detailed task analysis for the general personnel was performed. The analysis resulted in a list of tasks that the HBM must perform during offshore emergency situations. The tasks are presented in Figure 6.3.
2. Objects and properties of those objects that the HBM can sense: All the external PIFs listed in Table 6.2 fall in this category.
3. Objects and properties of those objects that the HBM can explicitly change through its actions: During a simulated scenario in AVERT, the agent may interact with and change the properties of the following objects: Doors, Muster board, T-Card, Personal protective equipment (PPE), and Manual alarm call point (MAC).
4. Effects of the internal factors that can moderate the model’s response: The internal PIFs that can moderate the model’s response are listed in Table 6.1. These PIFs were used to capture the across subject variability during the same scenario. The PSFs discussed in Section 6.2.2.4 were used to model the effects of the internal PIFs on response. The PSFs take the internal PIFs as inputs and use a BN to generate a dynamic response.
5. Knowledge that the HBM must possess to manifest the proper responses to the proper situations: The rules of behavior integrated in the HBM were generated from the same training tutorials and training scenarios used to train the participants in the experimental study. The training tutorial was used to generate rules of thumbs that

> help with platform familiarity, interpretation of the audio-visual cues from the alarm and PA, and safety procedures like closing fire/watertight doors and not running. The training scenarios were used to generate behavior rules that help the agent in choosing the egress route (including possible re-route depending on the situation) during an emergency scenario. It must be noted that given the same training, participants were observed to learn and infer things differently. Consequently, the content of the knowledge base varied from person to person. The same concept was followed while developing the knowledge base of the agents. The contents of the knowledge base in naïve, ideal, and in-between agents were different, allowing behavior variability.

Once the conceptual model was developed, it was checked with the SMEs to make sure of the following:

Sets of situations and responses are sufficient to accommodate the scenarios required to achieve the purpose: In the experimental study, sets of situations were chosen based on SME guidance and industry standards. The same set of situations were used for training and testing the agents. The external PIFs listed in Table 6.2 were used to create the range of credible situations. This list was reviewed by the SMEs. Though not comprehensive, the list was considered sufficient to model a reasonable set of situations.

The possible range of responses – both ideal and erroneous – were also reviewed by the SMEs. The responses captured by the task network in Figure 6.3 were considered sufficient and no further revision was suggested. However, the SMEs suggested the following additions to the potential error list:

- Not securing workspace before mustering
- Forgetting to register at the final destination – not doing the T-Card
- Not providing relevant feedback at the final destination
- Premature evacuation – getting in the lifeboat and driving away

As the current configuration of AVERT could not allow for the suggested additions, these additions are considered as future work and are out of scope of this paper.

Influences of internal PIFs or behavior moderators are adequately represented: The list of internal PIFs in Table 6.1 was reviewed by the SMEs. The list was considered sufficient to model the across subject variability. A BN approach was used to model the influence of the PIFs on performance (Musharraf et al., 2016a). The influence of internal PIFs on task performance cannot be studied in a controlled experiment. Goerger (2004) identifes this as the most complex phase of creating and modifying BN for HBM development. In this paper, the influence of internal PIFs is defined by the analyst based on the observations during the experimental study. Musharraf et al. (2016a) shows a successful demonstration of the BN in assessing the internal PIFs.

6.3.4 Validate the HBM's knowledge base using referent and requirements

As stated in Section 6.3.3, the rules of behavior were generated from the same training content used to train the participants in the experimental study. Decision rules with causal if/then association were generated from the observation of participants' performance during the training scenarios. Participants' performance in test scenarios was then predicted using the generated decision rules. The predicted performance was compared to the observed performance of the participants in the same scenarios, using data that had been set aside for validation. The prediction accuracy of the trees can be calculated using equation 1.

$$(\%)Prediction\ accuracy = \frac{n}{N} \times 100 \tag{1}$$

where $n =$ number of test scenarios for which (predicted outcome = observed outcome) and $N =$ total number of test scenarios.

An average of 95% prediction accuracy was achieved for the decision rules in the knowledge base. For 85% of participants, the prediction was accurate for all test scenarios (i.e. 100% accuracy). More detail on this can be found in Musharraf et al. (2017a).

6.3.5 Analyze the HBM's conceptual model and knowledge base to identify complex areas of the model that need attention in future validation activities

The conceptual model helps to define number, ranges, and intersections of the inputs (internal and external PIFs) and outputs of the HBM. It also defines the number and intersections of different situations to which the HBM must be able to respond. All this information contributes to identifying the complex areas of the HBM. The knowledge base validation also contributes by recognizing where the complexities of the knowledge lie.

A few complex areas in the behavior space were identified during conceptual model and knowledge base validation. These include coping with alarm changes and PA changes during an emergency situation, understanding route obstruction from the PA and using this knowledge while evaluating different egress routes, re-routing to avoid potential interaction with hazard, and understanding the responsibility to raise the alarm when the first observer of any hazard. Identification of the complex areas guided the test scenario design in the next step.

6.3.6 Result validation

Result validation consists of two steps. The first step is to perform a fundamental testing to see if the agent can demonstrate adequate individual skills. In the context of this paper, this means the agent must be able to perform the tasks of a general personnel during different offshore emergency situations. Since the focus of this step is to make sure that competent performance of a general personnel can be achieved by the agent, among the three types of the agents (i.e. naïve, ideal, and in-between) only performance of ideal agents will be analyzed at this step. The next step is to test that the HBM is able to capture the expected range of behaviors under different operating conditions. Besides ideal agents, performance of naïve and in-between agents will also be analyzed at this step. At both steps, test results will be compared against the acceptability criteria to determine the validity of the HBM.

Result validation involves 4 critical steps. Steps of result validation are described in detail in the following subsections.

6.3.6.1 Developing the HBM test plan

The fundamental steps of developing the test plan were to define the objectives and the type of testing that will be done. The objective of the test was to see if the HBM performs adequately under different operating conditions. Only aspects of agents related to the HBM were investigated. General aspects, such as look and feel of the agents, were out of the scope of the result validation. Among different testing approaches, acceptability criteria testing was selected as suggested by Anon. (2001a).

Acceptance criteria define the desirable behavior outcomes of the HBM. Defining acceptability criteria for an HBM is challenging due to its stochastic nature. Even for the same operating conditions and the same set of PIFs, there can be more than one acceptable behavior outcome. The behaviors observed during the experimental study were used as a benchmark to define an acceptable range. In the original study, participants' performance was tested in a set of testing scenarios. Each participant achieved an aggregated performance score at the end of the test scenario (Smith , 2015). The score of a participant can be used to classify them into one of the following categories: naïve (0-30%), in-between (31-79%), and ideal (80-100%). Behaviors observed in each category were used to define the acceptability criteria for naïve, in-between, and ideal agents.

The observed behavior of participants across emergency scenarios are listed in Table 6.3. As shown in Table 6.3, there are expected ideal behavior outcomes for each task. However, erroneous deviations were also observed during scenarios. All these define the acceptable behavior range for the HBM.

Behavior of ideal agents is expected to incline towards the ideal behavior outcomes observed during the experimental study. Naïve agents are expected to incline more towards the erroneous deviations. In-between agents are expected to have a combination of ideal and erroneous behavior outcomes. Figure 6.5 summarizes the HBM test plan.

It should be noted that all tasks shown in Figure 6.3 are not listed in Table 6.3 as tasks such as perception do not have observable outcomes.

Table 6.3: List of observed behavior outcomes during offshore emergency scenarios

Task	***Demonstrated by***	***Acceptable behavior range***	
		Ideal behavior outcome	***Possible deviation***
Identify Alarm	Mustering at the correct final destination	Muster at the muster station in case of GPA and the starboard side lifeboat station in case of PAPA	Misidentify GPA as PAPA or vice versa and muster at the incorrect location.
Evaluate egress path and choose egress route	Moving along a chosen egress route	Choose an egress route according to the direction of the PA. No interaction with hazard and/or rerouting expected if the correct route is chosen.	Choose an incorrect route and - reroute as soon as hazard is observed keep following the untenable route
Register at the final destination	Mustering at the final destination	Same as the task – identify alarm	Same as the task – identify alarm. Additionally, muster at the port side lifeboat station instead of starboard side lifeboat station.
Following safety procedure while moving along the egress route	Walking instead of running and keeping all fire/watertight doors closed	Walking and keeping all fire/watertight doors closed	Running and keeping fire/watertight doors open
Raising an alarm in case of first observer of a hazard	Activating manual alarm call points (MAC)	Activating MAC if there is a hazard on sight and no alarm or PA is currently activated.	Skip activating MAC and muster prematurely.

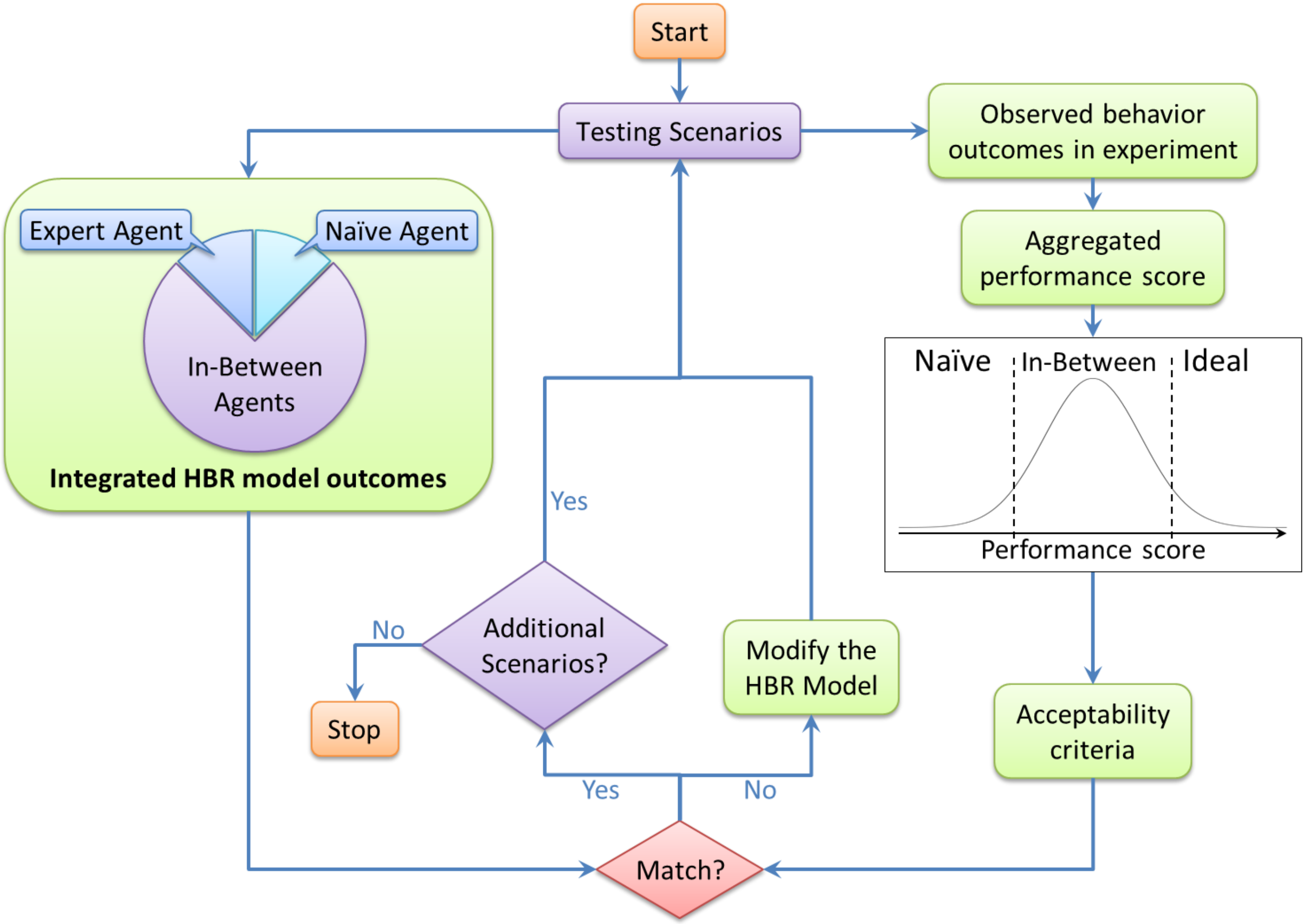


Figure 6.5: Fundamental steps in the HBM test plan. During the test, behavior of all three types of agents is compared to the acceptability criteria defined using empirical evidence

6.3.6.2 Designing the HBM testing scenarios

Design of test scenarios to validate the HBM was guided by the complex areas identified during conceptual model and knowledge base validation. First, a basic static scenario was designed. This scenario covered the fundamental task that the agent must be able to perform. Next, complexities were added by following means: dynamic changes of alarm and PA, obstruction of primary route with hazard, obstruction of secondary route with hazard, and introduction of a hazard on sight without any alarm or PA.

Table 6.4 summarizes the test scenarios for the HBM validation exercise.

Table 6.4: Test scenarios

Scenario Name	***Context***
Scn1	Agent starts in the cabin. A GPA sounds followed by a PA announcement notifying of a man overboard (MOB) drill. The agent must go to muster station using either primary or secondary egress route.
Scn2	Agent starts in the cabin. Fire erupts in the galley signaling a GPA. The agent must go to the muster station, but re-route to the lifeboat station due to the fire and smoke spreading to the adjacent muster station. The primary egress route and the muster station are compromised by the hazards.
Scn3	Agent starts in the cabin. A fire and explosion on the helideck signal a GPA. High winds cause the smoke to engulf a portion of the platform exterior. The agent must go to the muster station, but re-route to the lifeboat station due to the increase in emergency severity and the alarm change to PAPA. The hazard blocks the secondary egress route.
Scn4	Agent starts in the C-Deck Hallway and watches smoke coming out the cabin. The alarm is not triggered and no PA is available. Agent must raise the alarm and go to muster station using either primary or secondary egress route.

6.3.6.3 Conducting the tests and assessing test results against the acceptability criteria

The three types of agents were tested in the 4 scenarios listed in Table 6.4. As observed in the experimental study, agents were allowed to have a preferred egress route. With 3 types of agents, 2 possible preferred routes, and 4 different scenarios, a total of 24 combinations had to be tested. Figure 6.6 shows the possible combinations.

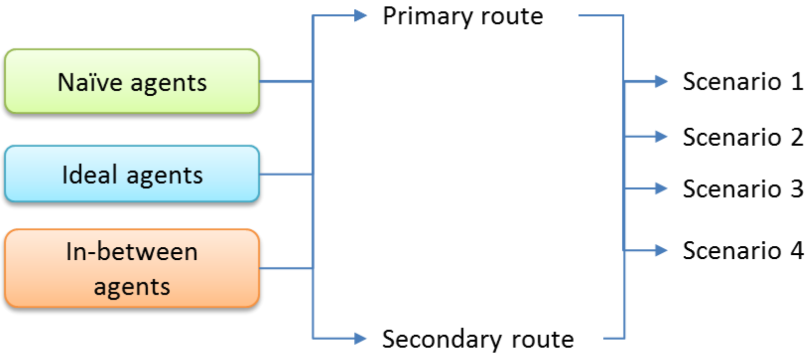


Figure 6.6: Possible combinations for testing and data collection

10 simulation runs were conducted for each combination, giving a total of 240 simulation runs. As the HBM is stochastic, even for the same combination, behavior may vary across repeated simulation conditions. The most common behaviors observed for each combination are discussed below.

Naïve agents: As stated in Section 6.3.1, naïve agents were created by setting knowledge as low, bias as high, compliance as low, efficacy of information use as low, and prioritization as wrong. Table 6.5 lists the most common behaviors of naïve agents observed in the testing scenarios.

Table 6.5: Common behaviors of naïve agents in the testing scenarios

Scenario	***Observed behavior***	
	Preference: Primary route	***Preference: Secondary route***
Scn1	Agent starts in the cabin. As GPA sounds, the agent takes the primary route to egress and goes to the mess hall. While egressing, the agent keeps the fire/watertight doors open and runs instead of walks.	Agent starts in the cabin. As GPA sounds, the agent takes the secondary route to egress and goes to the mess hall through starboard side. While egressing, the agent keeps the fire/watertight doors open and runs instead of walks.
Scn2	Agent starts in the cabin. As GPA sounds, the agent takes the primary egress route and heads toward the mess hall. As the agent arrives at the mess hall, it finds the muster station compromised. The agent interacts with the hazard. As the alarm changes to PAPA, the agent decides to reroute to the lifeboat station. The agent mistakenly goes to the port side lifeboat instead of the starboard side lifeboat. While egressing, the agent keeps the fire/watertight doors open and runs instead of walks.	Agent starts in the cabin. As GPA sounds, the agent takes the secondary egress route and heads toward the mess hall. On the way to the mess hall, the alarm changes to PAPA. The agent reroutes to the lifeboat station. The agent mistakenly goes to the port side lifeboat instead of the starboard side lifeboat. While egressing, the agent keeps the fire/watertight doors open and runs instead of walks.
Scn3	Agent starts in the cabin. As GPA sounds, the agent takes the primary egress route and heads toward the mess hall. As the alarm changes to PAPA, the agent decides to reroute to the lifeboat station. The agent mistakenly goes to the port side lifeboat instead of the starboard side lifeboat. While egressing, the agent keeps the fire/watertight doors open and runs instead of walks.	Agent starts in the cabin. As GPA sounds, the agent takes the secondary egress route and heads toward the mess hall. On the way, the agent sees the smoke, but goes through it anyway. As the alarm changes, the agent reroutes to the lifeboat station. The agent mistakenly goes to the port side lifeboat instead of the starboard side lifeboat. While egressing, the agent keeps the fire/watertight doors open and runs instead of walks.

Scenario	***Observed behavior***	
	Preference: Primary route	***Preference: Secondary route***
Scn4	Agent starts in the C-Deck Hallway (farthest from the cabin). The agent considers the danger imminent and starts to egress without activating the MAC. Agent takes the primary egress route to the mess hall and musters there. While egressing, the agent keeps the fire/watertight doors open and runs instead of walks.	Agent starts in the C-Deck Hallway (farthest from the cabin). The agent considers the danger imminent and starts to egress without activating the MAC. Agent takes the primary egress route to the mess hall and musters there. While egressing, the agent keeps the fire/watertight doors open and runs instead of walks.

As shown in Table 6.5, the naïve agents followed their preferred route irrespective of the obstruction of the route with hazard. This led to interaction with the hazard. The other common mistakes were: confusing the portside lifeboat station as the starboard side lifeboat station, running, keeping fire/watertight doors open, and not activating the MAC as the first observer of the hazard. By comparing the performance of the naïve agents against the acceptability criteria listed in Table 6.3, it is found that in most cases the performance matches the erroneous behavior outcomes, as expected.

Ideal agents: Ideal agents were created by setting the internal PIFs in the following way: knowledge as high, bias as low, compliance as high, efficacy of information use as high, and prioritization as right. Table 6.6 lists the most common behaviors of ideal agents observed in the testing scenarios.

Table 6.6: Common behaviors of ideal agents in the testing scenarios

Scenario	*Observed behavior*	
	Preference: Primary route	***Preference: Secondary route***
Scn1	Agent starts in the cabin. As GPA sounds, the agent takes the primary route to egress and goes to the mess hall. While egressing, the agent closes all fire/watertight doors and walks.	Agent starts in the cabin. As GPA sounds, the agent takes the secondary route to egress and goes to the mess hall through the starboard side. While egressing, the agent closes all fire/watertight doors and walks.
Scn2	Agent starts in the cabin. As GPA sounds, it waits and listens to the PA. Based on the PA, it takes the secondary route and heads toward the mess hall. As the alarm changes to PAPA, it goes to the lifeboat station instead, and musters there. While egressing, the agent closes all fire/watertight doors and walks.	Agent starts in the cabin. As GPA sounds, it waits and listens to the PA. Based on the PA, it takes the secondary route and heads toward the mess hall. As the alarm changes to PAPA, it goes to the lifeboat station instead, and musters there. While egressing, the agent closes all fire/watertight doors and walks.
Scn3	Agent starts in the cabin. As GPA sounds, it waits and listens to the PA. Based on the PA, it takes the primary route and heads toward mess hall. As alarm changes to PAPA, it goes to the lifeboat station through the mess hall and musters there. While egressing, the agent closes all fire/watertight doors and walks.	Agent starts in the cabin. As GPA sounds, it waits and listens to the PA. Based on the PA, it takes the primary route and heads toward the mess hall. As alarm changes to PAPA, it goes to the lifeboat station and musters there. While egressing, the agent closes all fire/watertight doors and walks.
Scn4	Agent starts in the C-Deck Hallway (farthest from the cabin). As a first observer of the hazard, the agent activates the MAC. It takes the primary egress route to the mess hall and musters there. While egressing, the agent closes all fire/watertight doors and walks.	**a.** Agent starts in the C-Deck Hallway (farthest from the cabin). As a first observer of the hazard, the agent activates the MAC. It takes the primary egress route to the mess hall and musters there. While egressing, the agent closes all fire/watertight doors and walks.

		b. Agent starts in the C-Deck Hallway (farthest from the cabin). As a first observer of the hazard, the agent activates the MAC. It approaches the secondary stairwell from a different way than usual to avoid the hazard. The agent takes the secondary route to the mess hall and musters there. While egressing, the agent closes all fire/watertight doors and walks.

The ideal agents were able to perform all the tasks of general personnel during emergency situations. Though the probability of the ideal agents making a mistake is non-zero, their most common behaviors match the ideal behaviors as defined in Table 6.3. Irrespective of their preferred route, ideal agents were able to pick a route to avoid potential interaction with hazards. They registered at the correct muster location and followed the safety procedures while egressing. They were also able to activate the MAC as the first observer of the hazard. The occasional errors made by the ideal agents were running and not closing the fire/watertight doors. These were also the most common erroneous behaviors of the participants with high scores in the experimental study.

It is also worth noting that unlike naïve agents, the ideal agents did not always prefer one route over another. The preference often depended on the final destination or presence of hazards. For example, an ideal agent may prefer the primary route when there is no hazard, but take the secondary route in the presence of a hazard.

In-between agents: As discussed in Section 6.3.1, the example in-between agents used in this paper were created by setting knowledge as low, bias as low, compliance as high, prioritization as right, and efficacy of information use as low. Table 6.7 lists the most common behaviors of in-between agents observed in the testing scenarios.

Table 6.7: Common behaviors of in-between agents in the testing scenarios

Scenario	***Observed behavior***	
	Preference: Primary route	***Preference: Secondary route***
Scn1	Agent starts in the cabin. As GPA sounds, the agent takes the primary route to egress and goes to the mess hall. While egressing, the agent keeps all fire/watertight doors open but walks.	Agent starts in the cabin. As GPA sounds, the agent takes the secondary route to egress and goes to the mess hall through starboard side. While egressing, the agent keeps all fire/watertight doors open and runs instead of walks.
Scn2	Agent starts in the cabin. As GPA sounds, the agent takes the primary route to mess hall. As it reaches the mess hall, it realizes that the mess hall is compromised. The agent immediately reroutes. It goes up a deck and takes the secondary route to egress. In the meantime, alarm changes to PAPA and the agent musters at the lifeboat station. While egressing, the agent keeps all fire/watertight doors open but walks.	Agent starts in the cabin. As GPA sounds, the agent takes the secondary route and heads towards mess hall. As the alarm changes to PAPA on the way, the agent musters at the lifeboat station instead. While egressing, the agent closes all fire/watertight doors and walks.
Scn3	Agent starts in the cabin. As GPA sounds, it waits and listens to the PA. Based on the PA, it takes the primary route and heads towards mess hall. As alarm changes to PAPA, it goes to the lifeboat station through mess hall and musters there. While egressing, the agent keeps all fire/watertight doors open and runs instead of walks.	Agent starts in the cabin. As GPA sounds, the agent takes the secondary route to egress. It reroutes immediately after finding out that the route is compromised with smoke. It takes the primary route and goes to starboard side lifeboat through mess hall. While egressing, the agent closes all fire/watertight doors and walks.

Scenario	***Preference: Primary route***	***Preference: Secondary route***
Scn4	Agent starts in the C-Deck Hallway (farthest from the cabin). The agent considers the danger imminent and start to egress without activating the MAC. Agent takes primary egress route to mess hall and musters there. While egressing, the agent closes all fire/watertight doors and walks.	Agent starts in the C-Deck Hallway (farthest from the cabin). The agent considers the danger imminent and starts to egress without activating the MAC. It approaches the secondary stairwell from a different way than the usual to avoid the hazard. The agent takes the secondary route to mess hall and musters there. While egressing, the agent closes all fire/watertight doors but runs instead of walks.

As shown in Table 6.7, behaviors of in-between agents lie somewhere between the two extremes of ideal and naïve agents. The in-between agents may fail to interpret the PA and take the ideal route from the beginning. As soon as they realize that the current route is untenable, they reroute immediately. The agents may not follow the safety procedure at all times, and may forget to activate the MAC. As expected, in all scenarios, the in-between agents had a combination of ideal and erroneous behavior outcomes.

A summary of agent behavior in Scn2 is presented in Figures 6.7 and 6.8. Figure 6.7 shows the number of erroneous behaviors observed for each type of agent during the 10 simulation runs for this scenario. As show in the figure, though ideal agents occasionally commit errors, in general they exhibit safer behavior compared to the in-between and naïve agents. In all simulation runs, the ideal and in-between agents manage to avoid interaction with hazards by either choosing the safest route from the beginning, or by rerouting immediately after encountering the hazard. Figure 6.8 summarizes the route choice of all types of agents in Scn2. Behavior of agents in other scenarios can be summarized in the same way.

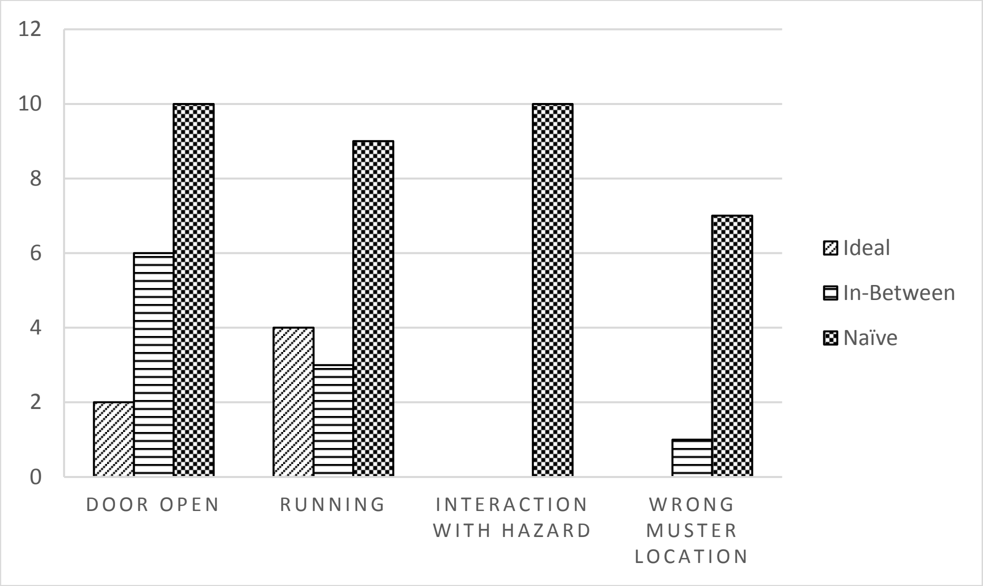


Figure 6.7: Observed erroneous behavior of 3 types of agents in Scn2 (preference = primary route) for 10 simulation runs each.

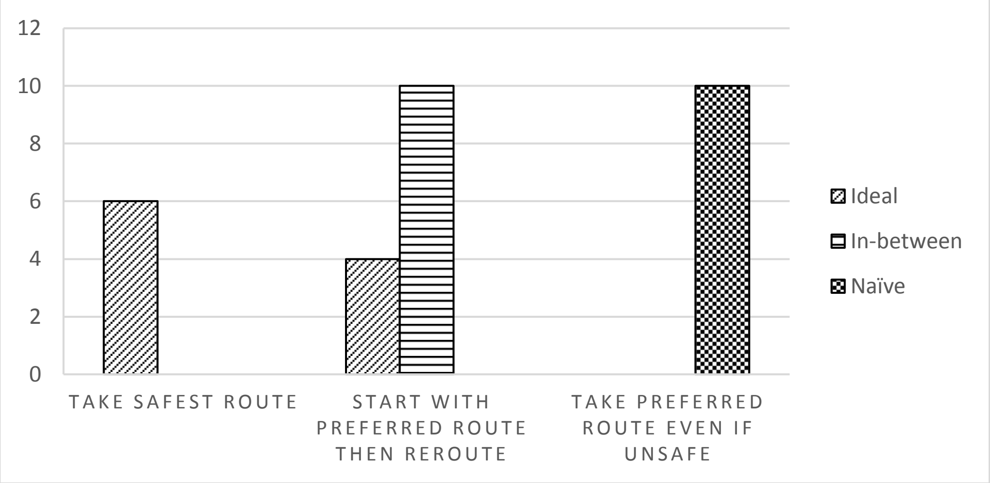


Figure 6.8: Route choice of 3 types of agents in Scn2 (preference = primary route) for 10 simulation runs each.

6.4 Limitations and future work

One of the biggest challenges faced during result validation was to determine the number of simulation runs. With a deterministic simulation, just one run is enough. The answer is not so simple for stochastic simulations. Though some guidelines are available for determining the number of runs for quantitative stochastic simulations, such guidelines are not available for qualitative simulations like behavior simulation (Byrne, 2013). In this paper, a total of 240 runs were conducted simply for feasibility. However, that only allowed 10 runs for each combination. In future, more runs per combination will be conducted to increase the confidence in the results (Ritter et al., 2011).

The testing conducted during result validation did not involve SMEs. It is often recommended in literature to involve SMEs in the testing process by either direct participation or by review of the testing report. Authors plan to get the test reports reviewed by SMEs in future. Authors also plan to extend the behavior spectrum by including worksite scenarios and having more than one agent simultaneously.

It must be noted that virtual environments can provide a certain degree of realism and should not be expected to be an exact counterpart to real life emergency situations. The goal of the validation presented in this paper is to make sure that the agents behave as realistically as the participants in the study. More sophisticated behavior of the agents (i.e. as in real world settings) is out of the scope of this paper.

6.5 Conclusion

With the advancement of simulators as a training tool, use of software agents to enable team training has become quite common. Credibility of these agents is critical to ensure a sound training process. Validation of the underlying HBM of these agents is the first step to ensure such credibility. This paper presents the validation process of an HBM of general personnel, created for an offshore emergency training simulator. Unlike traditional HBM validation processes that use experts' opinion, empirical evidence was used in this paper. Use of empirical evidence makes the validation more objective and reliable.

All high-level tasks of validation are discussed in detail with special emphasis given on the acceptability criteria testing. Three types of agents – naïve, ideal, and in-between – were tested during validation. Results show that the integrated HBM meets the acceptability criteria requirement for all types of agents. This indicates that the agents have potential to be used as team members for crew training in offshore emergency situations. A combination of different types of agents will allow creating a heterogeneous training environment, which would be a closer representation of the actual working environment.

Future work includes improving the result validation by conducting more simulation runs and getting the test reports reviewed by SMEs. The authors also plan to extend the behavior spectrum by including worksite scenarios and team simulation.

Acknowledgements

The authors acknowledge with gratitude the support of NSERC-Husky Energy Industrial Research Chair in Safety at Sea and the Research & Development Corporation of Newfoundland and Labrador.

References

Anon. (2001a). *Verification, Validation, and Accreditation (VV&A) Recommended Practices Guide (RPG): Reference Document - Human Behavior Representation (HBR) Literature Review.* Department of Defense Modeling and Simulation Office (DMSO). Retrieved from https://vva.msco.mil/Ref_Docs/HBR/beh-ref-pr.pdf

Anon. (2001b). *Verification, Validation, and Accreditation (VV&A) Recommended Practices Guide (RPG): Special Topic - Validation of Human Behavior Representations.* Department of Defense Modeling and Simulation Office (DMSO). Retrieved from https://vva.msco.mil/Special_Topics/HBR-validation/hbr-validation-pr.pdf

Anon. (2011). *Modeling and Simulation (M&S) Verification, Validation, and Accreditation (VV&A) Recommended Practices Guide (RPG): Validation Referent.* Department of Defense Modeling and Simulation Office (DMSO). Retrieved from https://vva.msco.mil/Special_Topics/Validation_Referent/val_ref-pr.pdf

Blackman, H. S., Gertman, D. I., & Boring, R. L. (2008). Human error quantification using performance shaping factors in the SPAR-H method. *Proceedings of the human factors and ergonomics society annual meeting* (pp. 1733-1737). Sage CA: Los Angeles: CA: SAGE Publications.

Bradbury-Squires, D. (2013). *Simulation training in a virtual environment of an offshore oil installation.* St. John’s, NL: Memorial University of Newfoundland.

Byrne, M. D. (2013). How many times should a stochastic model be run? An approach based on confidence intervals. *Proceedings of the 12th International conference on cognitive modeling.* Ottawa.

Chang, Y. H., & Mosleh, A. (2007). Cognitive modeling and dynamic probabilistic simulation of operating crew response to complex system accidents: Part 1: Overview of the IDAC Model. *Reliability Engineering & System Safety, 92*(8), 997-1013.

DiMattia, D. G., Khan, F. I., & Amyotte, P. R. (2005). Determination of human error probabilities for offshore platform musters. *Journal of loss prevention in the process industries, 18*(4), 488-501.

Duffy, V. G. (2008). *Handbook of digital human modeling: research for applied ergonomics and human factors engineering.* CRC press.

Goerger, S. R. (2004). *Validating human behavioral models for combat simulations using techniques for the evaluation of human performance.* MONTEREY, CA: NAVAL POSTGRADUATE SCHOOL.

Groth, K. M., & Mosleh, A. (2012). A data-informed PIF hierarchy for model-based Human Reliability Analysis. *Reliability Engineering & System Safety, 108,* 154-174.

Han, J., Kamber, M., & Pei, J. (2011). *Data mining: concepts and techniques.* Elsevier.

House, A. W., Smith, J., MacKinnon, S., & Veitch, B. (2014). Interactive simulation for training offshore workers. *Oceans'14 MTS/IEEE Conference* (pp. 1-6). St. John's, NL: IEEE.

Karr, C. R., Reece, D., & Franceschini, R. (1997). Synthetic soldiers in military training simulators. *IEEE spectrum, 34*(3), 39-45.

Kendal, S. L., & Creen, M. (2007). *An introduction to knowledge engineering.* London: Springer.

Musharraf, M., Khan, F., & Veitch, B. (2017). *Modeling and simulation of personnel response during offshore emergency situations.* 3rd Workshop and Symposium on Safety and Integrity management of operations in harsh environments.

Musharraf, M., Smith, J., Khan, F., Veitch, B., & MacKinnon, S. (2016). Assessing offshore emergency evacuation behavior in a virtual environment using a Bayesian Network approach. *Reliability Engineering & System Safety, 152,* 28-37.

Musharraf, M., Smith, J., Khan, F., Veitch, B., & MacKinnon, S. (2017a). Identifying route selection strategies in offshore emergency situations using Decision Trees: A step towards adaptive training. *Manuscript submitted for review*.

Musharraf, M., Smith, J., Khan, F., Veitch, B., & MacKinnon, S. (2017b). Incorporating individual differences in human reliability analysis: An extension to the virtual experimental technique. *Safety Science*.

Negnevitsky, M. (2005). *Artificial intelligence: a guide to intelligent systems.* Pearson Education.

Ritter, F. E., Kim, J. W., Morgan, J. H., & Carlson, R. A. (2012). *Running behavioral studies with human participants: A practical guide.* Sage Publications.

Ritter, F. E., Schoelles, M. J., Quigley, K. S., & Klein, L. C. (2011). Determining the number of simulation runs: Treating simulations as theories by not sampling their behavior. In *Human-in-the-loop simulations* (pp. 97-116). Springer.

Smith, J. (2015). *The effect of virtual environment training on partcipant competence and learning in offshore emergency egress scenarios.* St. John's, NL: Faculty of Engineering and Applied Science, Memorial University on Newfoundland.

Smith, J., Musharraf, M., Veitch, B., & Khan, F. (2017). *Can simulation-based mastery learning increase compliance: investigating decsion making in virtual offshore emergency egress.* 3rd Workshop and Symposium on Safety and Integrity management of operations in harsh environments.

Smith, R. D. (1998). Essential techniques for military modeling and simulation. *Proceedings of the 30th conference on winter simulation* (pp. 805-812). IEEE Computer Society Press.

Sokolowski, J. A., & Banks, C. M. (2010). *Modeling and simulation fundamentals: theoretical underpinnings and practical domains.* John Wiley & Sons.

Thow-Yick, L. (1994). The basic entity model: A fundamental theoretical model of information and information processing. *Information Processing & Management, 30*(5), 647-661.

7. HUMAN PERFORMANCE DATA COLLECTED IN A VIRTUAL ENVIRONMENT

Mashrura Musharraf, Jennifer Smith, Faisal Khan**, Brian Veitch, Scott MacKinnon*
Centre for Risk, Integrity and Safety Engineering (C-RISE),
Faculty of Engineering & Applied Science,
Memorial University of Newfoundland,
St John’s, Newfoundland and Labrador, Canada A1B 3X5
*Department of Mechanics and Maritime Sciences,
Chalmers University, Gothenburg, Sweden
** Correspondence author: Tel: + 1 709 864 8939; Email: fikhan@mun.ca

Co-authorship statement

A version of this manuscript has been published in the Data in brief journal. Authors Mashrura Musharraf and Jennifer Smith designed the experiment, conducted the experiment, and performed necessary data collection. Co-authors Faisal Khan, Brian Veitch, and Scott MacKinnon supervised the experimental study. Mashrura Musharraf prepared the draft of the manuscript. All co-authors reviewed and provided feedback on the manuscript. Mashrura Musharraf revised the manuscript based on the co-authors’ feedback and during the peer review process.

Abstract

This data article describes the experimental data used in the research article "Incorporating individual differences in human reliability analysis: an extension to the virtual experimental technique" (Musharraf et al., in press). The article provides human performance data for 36 individuals collected using a virtual environment. Each participant was assigned to one of two groups for training: 1) G1: high level training and 2) G2: low level training. Participants' performance was tested in 4 different virtual scenarios with different levels of visibility and complexity. Several performance metrics of the participants were recorded during each scenario. The metrics include: time to muster, time spent running, interaction with fire doors and watertight doors, interaction with hazards, and reporting at different muster locations

Specifications Table

Subject area	Engineering, Human factors
More specific subject area	Safety & Risk, Human Reliability Analysis
Type of data	Text files
How data was acquired	Data were collected by conducting an experiment in a virtual environment. The virtual environment used is called the all-hands virtual emergency response trainer (AVERT) and was developed at the Memorial University. AVERT was designed to enhance offshore emergency response training. The virtual environment is modeled after an offshore oil installation platform with high levels of detail. It is capable of creating credible emergency scenarios by introducing hazards such as blackouts, fires and explosions. Human performance data of 36 individuals tested in simulated emergency scenarios in AVERT were collected.
Data format	Filtered and processed

Experimental factors	The participants were naïve concerning any detail of the experimental design, they were not employed in the offshore oil and gas industry, and were not familiar with the AVERT simulator prior to the experiment. Their ages ranged from 19-39 years. Information regarding participants' gaming and marine experience was collected prior to the experiment. This information guided the assignment of participants into different training groups. Participants were provided with basic offshore emergency preparedness training tutorials before performing in any simulated emergency scenarios.
Experimental features	Two performance influencing factors (PIFs) – visibility and complexity – were each tested at two different levels to create $2^2 = 4$ virtual testing scenarios. Participants' performance was tested in the scenarios and the following performance metrics were collected: time to muster, time spent running, interaction with fire doors and watertight doors, interaction with hazards, and reporting at different muster locations (i.e. mess hall/muster station, lifeboat starboard side, lifeboat port side).
Data source location	Memorial University of Newfoundland, St. John's, NL, Canada
Data accessibility	The data are with this article.

Value of the data

- The data serve as a benchmark for human performance in emergency situations.
- The data allow objective assessment of human reliability rather than subjective assessments that rely on expert judgement.
- The data enable investigating the effects of different PIFs on human performance.
- The data provide the information that each human is different and the effect of PIFs on performance can vary from individual to individual.
- Analysis of the data can provide direction towards adaptive training.

7.1 Data

Human performance data for 36 individuals in 4 testing scenarios are associated with this article. The testing scenarios were created in AVERT. Two PIFs – visibility and complexity – were varied in the scenarios. Details of the 4 testing scenarios can be found in Table 3 in (Musharraf et al., in press).

Performance metrics recorded during the scenarios include: time to muster, time spent running, interaction with fire doors and watertight doors, interaction with hazards, and reporting at different muster locations.

The 4 supplementary text files summarize the performance metrics of 36 individuals in the 4 simulated emergency scenarios.

7.2 Experimental Design, Materials and Methods

The data presented in this article were originally collected during an experimental study presented in (Smith J. , 2015) and (Musharraf et al., 2016). Though a broad range of human performance data were collected during the study, this article only presents the data relevant to the article “Incorporating individual differences in human reliability analysis: an extension to the virtual experimental technique” (Musharraf et al., in press).

A total of 36 participants took part in the study with a goal to learn how to perform a successful offshore emergency evacuation. The participants were naïve concerning any detail of the experimental design, they were not employed in the offshore oil and gas industry, and therefore they were not familiar with the offshore platform. Each participant was assigned to one of two groups for training: 1) G1: high level training and 2) G2: low level training. Participants in both groups received basic offshore emergency preparedness training. Participants in G1 received additional training tutorials and practice scenarios on alarms and hazards.

Once a participant was assigned to a group, his/her training level remained static (either low or high) for the rest of the study. The PIFs visibility and complexity, on the other hand, were set to different levels to investigate how these PIFs influence each participant. The schematic diagram of the experimental design can be found in (Musharraf et al., in press).

Acknowledgements

The authors acknowledge with gratitude the support of the Research & Development Corporation of Newfoundland and Labrador and the NSERC-Husky Energy Industrial Research Chair in Safety at Sea.

Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.dib.2017.09.029.

References

Musharraf, M., Smith, J., Khan, F., Veitch, B., & MacKinnon, S. (2016). Assessing offshore emergency evacuation behavior in a virtual environment using a Bayesian Network approach. *Reliability Engineering & System Safety*, *152*, 28-37.

Musharraf, M., Smith, J., Khan, F., Veitch, B., & MacKinnon, S. (in press). Incorporating individual differences in human reliability analysis: an extension to the virtual experimental technique. *Safety Science*.

Smith, J. (2015). *The effect of virtual environment training on participant competence and learning in offshore emergency egress scenarios.* Master of Engineering Thesis, Memorial University of Newfoundland, St. John's.

8. CONCLUSIONS & RECOMMENDATIONS

8.1 Conclusions

Post-accident analysis of disasters like Piper-Alpha shows that offshore emergencies are uncertain, dynamic, and stressful. Assistance cannot be reached immediately and successful handling of emergencies often depends on the competency of the personnel on board. Both individual and team competency are essential. The conventional training programs mostly focus on increasing individual competency, as organizing large-scale team exercises is often a challenge. The organizational and educational drawbacks of such team exercises make them unfeasible. This research proposes the use of artificial intelligent agents to enable team training in a VE. Computational models of human behavior are developed in this research that can be used to create such intelligent agents. Though similar works have been done for the military, aviation, and nuclear power plant industries, no such work is available to date for the offshore industries. This research will be first of its kind in the context of offshore emergencies.

Both SMEs' opinion and empirical evidence are used during the development and validation phases of the HBM. The first part of the research focuses on understanding human behavior by conducting an experiment in a VE. A range of emergency scenarios was created in the VE by varying the level of different external PIFs. Influence of these external PIFs on human performance was investigated during the study. Internal PIFs of participants of the experiment were also assessed using the data collected during the experiment. Participants' knowledge acquisition and inference processes were studied in the research. In the second part, understanding from the experimental study was used to develop the HBM. A BN approach was used to model the effect of external and internal PIFs. Special attention was paid to ensure that the model takes individual differences of participants into account while modeling the effect of different PIFs on human performance. The BN models allowed the consideration of dependency among the PIFs and associated performance. They also allowed updating prior probabilities with incoming new evidence. To model the reasoning processes of participants during emergency situations, decision trees were used. The decision trees offer a visual representation of the reasoning process, and valuable diagnostic capabilities. Once the HBM was developed, it was used to create agents with varying levels of skill. Three types of agents – naïve, ideal, and in-between – were created using the HBM. The third part of the research focused on validating the HBM. All high-level tasks of HBM validation were performed with special emphasis given on the result validation. Result validation shows that the integrated HBM meets the acceptability criteria requirement for all types of agents.

Outcomes of this research may help to advance emergency preparedness training and to improve safety. The mathematical models, BNs, and decision trees, developed in this research may be used to predict people's reliability in emergency situations. Diagnostic aspects of BN and decision trees may be used to identify the strengths and weakness of individuals. Such diagnosis may help design adaptive training to overcome weaknesses and reach competency faster. The diagnosis may also help the personnel selection process for different roles. Models like decision trees may be used to assess the efficacy of the training curriculum and/or pedagogical approaches. It is expected that a sound training approach would result in converging problem-solving strategies that lead to success. Any systemic exception might indicate weaknesses in the training approach itself. In addition to improving individual training, HBMs developed in this research may be used to facilitate team training in a VE. Agents created using the HBM can be used to create a heterogenous training environment where trainees can gather experience of working with, or commanding, teams with varying skill levels. This may help to keep trainees more focused and engaged during the training by providing novel challenges.

8.2 Technical challenges and recommendations

A few technical challenges faced during the research and according recommendations are presented below:

- Virtual environments can only provide a certain degree of realism. The work done in this research is centered around the virtual environment AVERT. Though AVERT represents an offshore oil installation platform with high levels of detail, it can not be taken as an exact counterpart to the real world operational environment.

  It is not feasible to compare the outcomes achieved in AVERT to outcomes in real emergency situations. However, special attention can be paid while designing emergency scenarios in VE to make sure they closely represent real life emergencies. Facts and findings in the literature and investigation reports of previous accidents can be used to create credible emergency scenarios.

- Human behavior and cognition processes are extremely complex, and both the number and the magnitude of factors that can influence human performance are very high. Using a comprehensive list of PIFs while keeping the number of virtual scenarios feasible is challenging. As the number of PIFs and associated magnitudes increase, the number of virtual scenarios needed to quantify computational models like BN also increase exponentially.

  Instead of investigating the effect of all possible PIFs at the same time, it is recommended to identify the ones that are vital for a given context. This can help to keep the number of scenarios manageable. A fractional factorial design can also be used instead of a full factorial one when possible. Use of concepts like Noisy OR may help minimize the data requirement for quantifying a BN.

- Blocking nuisance factors that have some effect on the response, but are not of particular interest to the experimenter is extremely difficult while conducting experiments with humans. As stated above, both the number and the magnitude of factors that can influence human performance is very high. It is nearly impossible to block all nuisance factors using the available blocking techniques while conducting experiments with humans (Montgomery, 2017). For example, in the experiment done in this research – three controlled factors training, visibility, and complexity – were of interest. However, it was observed during the experiment that some participants' performance in the virtual scenarios might be influenced by their gaming experience.

Though the two groups were balanced in terms of their self-reported video gaming experience, it was not possible to block the effect of such experience at an individual level.

While it may be impossible to block all possible nuisance factors during an experiment, it is recommended that sufficient research be done prior to experiment to identify these factors. This may allow the investigator to design the experiment in a way so that effects of nuisance factors are minimized. Even when elimination or reduction of the effect of nuisance factors is not possible, being aware of the factors can help interpret the results more accurately.

- Having a meaningful sample size can be challenging while conducting experiments with humans. Because this research looked into individual differences, rather than group statistics, having a meaningful sample size was even harder. Gathering sufficient data for training the models (i.e. BN, decision trees) while keeping the number of exposures to virtual scenarios manageable for an individual was a challenge. One possible improvement could be to allow the participants more time on the scenarios rather than lecture based tutorials. More data points in the training data set can help increase the prediction accuracy of the computational models.

- Incorporation of individual differences presents new challenges in the conventional verification and validation paradigm. Since conditional probabilities in the BN can be different for each individual, quantification of parameterization confidence suggested in a conventional validity framework is nearly impossible (Pitchforth & Mengersen, 2013). Though a validation exercise was not conducted for the BN models developed in this research, the models were later integrated into an HBM. The HBM is validated with the underlying assumption that the uncertainty involved in the BN is negligible.

- Balancing variation and validation of an HBM is one of the biggest challenges. Due to the stochastic nature of the HBM, it is not sufficient to expose the HBM to each scenario just one time. The HBM must be tested in the same scenario multiple times. No standard guideline is available that defines the number of such exposures. Ritter's (2011) suggestion regarding confidence parameters can be useful when deciding the number of required exposures.

It is worth mentioning that the experiments conducted in this research was designed to achieve multiple objectives at the same time and as a result there were a few constraints that needed to be maintained. For example, Smith (2015) used the collected data to study the learning effect of participants across scenarios. This prevented the scenarios to be randomized, which would be ideal for this research. Due to the challenges associated with experiments with humans (i.e. recruitment, time commitment of participants), it is often worth combining multiple research objectives into one experiment. In such cases, it is recommended that potential constraints resulting from the merger are analyzed prior to the experiment.

8.3 Future work

Some guidance on future works that can help advance the offshore emergency training are discussed below.

- An interesting future work would be investigating the effect of the different types of agents on human behavior and training efficacy.

- The external PIFs studied in this research were training, visibility, and complexity. Though the PIFs satisfied the intended variability encoding in the agents, a more relevant and informative set of PIFs would be useful. The AVERT configuration has made significant progress since the time of the study and the additional features have relaxed the constraints on the choice of PIFs. Future works should take advantage of the additional features to design experiments with more realistic PIFs. Moyle et al. (2017) has already conducted an experiment in AVERT with such PIFs (complexity, stress, and uncertainty), which are more realistic representatives of offshore emergency situations.

- Though PIFs like complexity and stress are more realistic, it is hard to define different degrees of these PIFs. Objective measurement of whether a virtual scenario is highly stressful or not can be challenging. Moreover, what seems to be a stressful situation for one, might not be stressful for others. Future works should look into measurable indicators, such as physiological data, to define the different degrees of the PIFs.

- This research focused on the casual dependency among the PIFs and human performance. The dependency among the PIFs themselves were not investigated in this research. Consideration of such dependency may help improve the accuracy of the computational models and should be considered in future research.

- This research demonstrates how BN can be used to improve human reliability assessment in offshore emergencies. BNs are just one of the many potential Bayesian methods that can improve HRA. Groth et al. (2014) discusses the advantages of applying other Bayesian methods and associated computational techniques to facilitate HRA with special emphasis given on Bayesian inference. Groth's work demonstrates the use of data collected in a nuclear power plant simulator to update the initial human error probability assigned by the experts. Similar work can be done using the data collected during this study to improve HRA in offshore oil & gas industries.

- Participants of the experimental study done in this research were naïve. They were not employed in the offshore oil and gas industry, and therefore they were not familiar with the offshore platform. It is anticipated that data collected from real offshore workers may provide different results. Since the goal is to use the models to improve safety in oil and gas industries, future works should consider participants with relevant experience.

- The focus of the research was on the behavioral outcomes of the agents. General aspects such as look and feel were out of the scope of this research. These aspects are very important for the agents to be realistic enough to facilitate team training in VEs. Future works should pay attention to the details of how behaviors of agents are portrayed in the VE. Another essential improvement of the agents is their ability to communicate. The range of team training scenarios will heavily depend on the sophistication of the communication module of the agents.

- This research focused on the behavior of general personnel working on offshore oil and gas installations. Since the motivation behind the research is to enable team training in VEs, emergency response behavior of other team members needs to be studied and modeled as well. This might require looking into the development and use of a shared mental model framework, especially while modeling co-operative decision makings (i.e. decisions in emergency co-ordination centre).

References

Groth, K. M., Smith, C. L., & Swiler, L. P. (2014). A Bayesian method for using simulator data to enhance human error probabilities assigned by existing HRA methods. *Reliability Engineering & System Safety, 128*, 32-40.

Montgomery, D. C. (2017). *Design and analysis of experiments.* John Wiley & Sons.

Moyle, A., Musharraf, M., Smith, J., Veitch, B., & Khan, F. (2017). Human Reliability Analysis using virtual emergency (VE) scenarios via a Bayesian Network model. *Proceedings of the 3rd Workshop and Symposium on Safety and Integrity Management of Operations in Harsh Environments (C-RISE3).*

Pitchforth, J., & Mengersen, K. (2013). A proposed validation framework for expert elicited Bayesian Networks. *Expert Systems with Applications, 40*(1), 162-167.

Ritter, F. E., Schoelles, M. J., Quigley, K. S., & Klein, L. C. (2011). Determining the number of simulation runs: Treating simulations as theories by not sampling their behavior. In *Human-in-the-loop simulations* (pp. 97-116). Springer.

Smith, J. (2015). *The effect of virtual environment training on participant competence and learning in offshore emergency egress scenarios.* Master of Engineering Thesis, Memorial University of Newfoundland, St. John's.

Appendix A

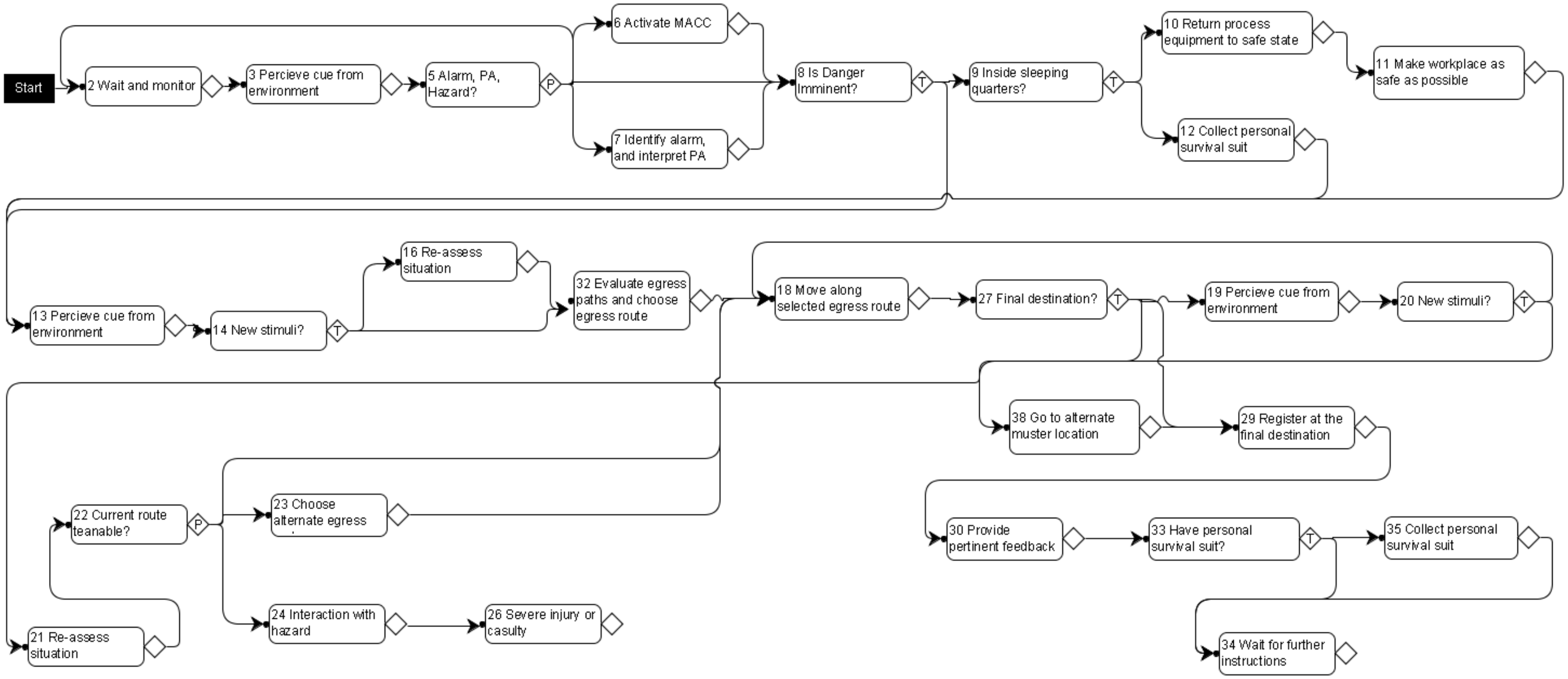
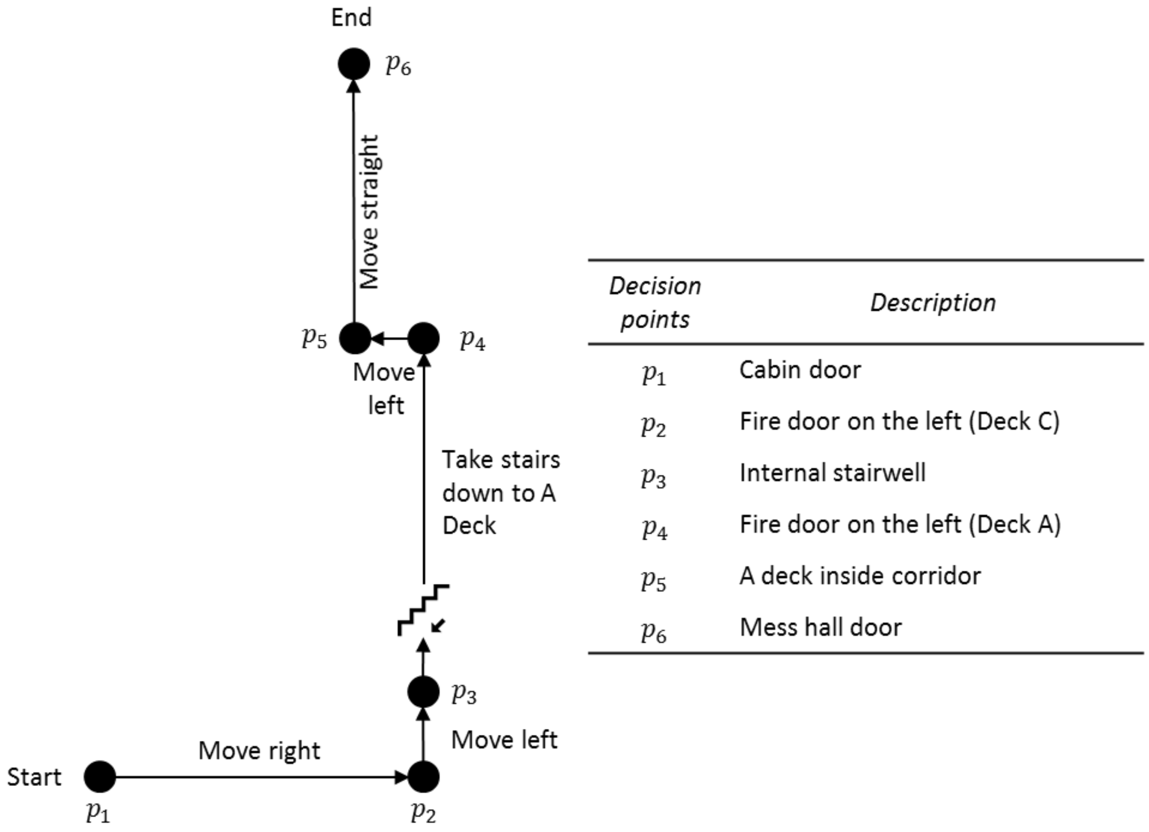

Start
2 Wait and monitor
3 Percieve cue from environment
5 Alarm, PA, Hazard?
P
6 Activate MACC
7 Identify alarm, and interpret PA
8 Is Danger Imminent?
T
9 Inside sleeping quarters?
T
10 Return process equipment to safe state
11 Make workplace as safe as possible
12 Collect personal survival suit
13 Percieve cue from environment
14 New stimuli?
T
16 Re-assess situation
32 Evaluate egress paths and choose egress route
18 Move along selected egress route
27 Final destination?
T
19 Percieve cue from environment
20 New stimuli?
T
38 Go to alternate muster location
29 Register at the final destination
22 Current route teanable?
P
23 Choose alternate egress
21 Re-assess situation
24 Interaction with hazard
26 Severe injury or casulty
30 Provide pertinent feedback
33 Have personal survival suit?
T
35 Collect personal survival suit
34 Wait for further instructions

Figure A.1: IPME network screenshot (based on the task network in Figure 5.3)

Sample function AlarmInterpretation() and associated equations

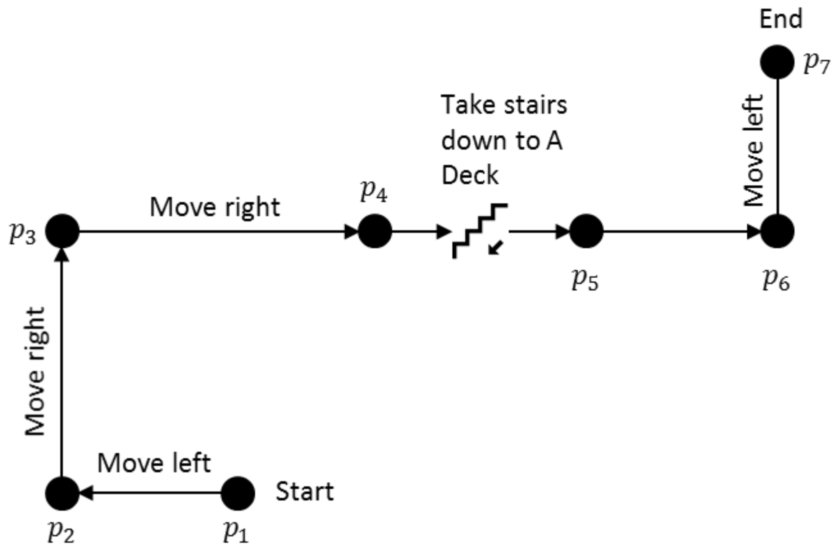
```
/*
Represents agent's interpretation of the alarm and PA
@modifies PercievedAlarmType, FinalDestination, PercievedObstructedRoute, PercievedPA
*/
void AlarmInterpretation()
{
    double CurrentProb = randInt (0,100);
    double ProbOfFailure = 100*PSF_InterpretAlarm();

    //if agent is interpreting correctly
    if (CurrentProb >= ProbOfFailure)
    {
        if (VisualForAlarm == "Flashing Green" && AudioForAlarm == "Two tone")
        {
               PercievedAlarmType = "GPA";
               FinalDestination = "Muster station";
        }
        else if (VisualForAlarm == "Steady Green" && AudioForAlarm == "Constant tone")
        {
               PercievedAlarmType = "PAPA";
               FinalDestination = "Lifeboat station";
        }
         PercievedObstructedRoute = ObstructedRoute;
         PercievedPA = PA;
    }

    //if agent is making mistake because of the influence of the PSFs
    else
    {
       if (VisualForAlarm == "Flashing Green" && AudioForAlarm == "Two tone")
       {
              PercievedAlarmType = "PAPA";
              FinalDestination = "Lifeboat station";
        }
        else if (VisualForAlarm == "Steady Green" && AudioForAlarm == "Constant tone")
        {
               PercievedAlarmType = "GPA";
               FinalDestination = "Muster station";
        }
       PercievedObstructedRoute = "None";
       PercievedPA = "None";
    }
    if (NewStimuli == "Yes") NewStimuli = "No";

}


/*
Calculates probability of failure for the task Alarm & PA interpretation based on the state of the PSFs.
@returns probability of failure for interpret alarm
*/
double PSF_InterpretAlarm()
{
    double ProbOfIntFailure = 0.0;
    if ((AssignedOp.Knowledge.Value == HiMedLow.High) && (AssignedOp.EfficacyOfInformationUse.Value ==
    HiMedLow.High))
            ProbOfIntFailure = 0.25;
    else if ((AssignedOp.Knowledge.Value == HiMedLow.Low) && (AssignedOp.EfficacyOfInformationUse.Value
    == HiMedLow.High))
            ProbOfIntFailure = 0.5;
    else if ((AssignedOp.Knowledge.Value == HiMedLow.High) && (AssignedOp.EfficacyOfInformationUse.Value
    == HiMedLow.Low))
            ProbOfIntFailure = 0.5;
    else if ((AssignedOp.Knowledge.Value == HiMedLow.Low) && (AssignedOp.EfficacyOfInformationUse.Value
    == HiMedLow.Low))
            ProbOfIntFailure = 0.95;
    return ProbOfIntFailure;
}
```

Appendix B



Decision points	*Description*
p_1	Cabin door
p_2	Fire door on the left (Deck C)
p_3	Internal stairwell
p_4	Fire door on the left (Deck A)
p_5	A deck inside corridor
p_6	Mess hall door

Figure B.1: Schematic diagram of primary egress route from cabin to mess hall.



Decision points	*Description*
p_1	Cabin door
p_2	Corridor on right
p_3	Fire door
p_4	External stairwell
p_5	End of stairs (Deck A)
p_6	A blue colored pipe
p_7	Mess hall fire door

Figure B.2: Schematic diagram of secondary egress route from cabin to mess hall.