HUMAN RELIABILITY ANALYSIS USING VIRTUAL EMERGENCY SCENARIO VIA A BAYESIAN NETWORK MODEL

by

© Daphne Allison Blundon

A Thesis submitted to the School of Graduate Studies

in partial fulfillment of the requirements for the degree of

Masters of Engineering

Faculty of Engineering and Applied Science

Memorial University of Newfoundland

May 2019

St. John's Newfoundland and Labrador

Abstract

Human reliability assessments (HRA) are typically completed by eliciting expert opinion. Data used are subjective and are prone to uncertainty and errors. This thesis outlines an HRA method using a Bayesian network (BN) model to evaluate human performance in emergency scenarios using a virtual environment (VE). VE can be used to simulate emergency situations to evaluate human performance in an environment that is controlled and safe and gives access to data that is based on an experimental method, rather than expert opinion. This method involves selecting appropriate performance shaping factors (PSFs) that are varied into different states to create credible scenarios in the VE to observe human performance. The virtual experimental technique provides a way to collect data to quantify a BN. The BN approach is suited to the assessment of human reliability due to its ability to 1) characterize dependency among different performance shaping factors (PSFs) and human errors, 2) incorporate new evidence as it becomes available, and 3) quantify the impact of different PSFs on different individuals. This paper presents an extension of the work done by Musharraf et al. (2014) by introducing PSFs that were purposively selected based on the ability to implement them in the VE, their relevance to real-life situations, and whether they could be controlled to minimize the effects of variables other than the chosen PSF. The PSFs used in this paper are complexity, stress, and uncertainty.

Acknowledgment

First, I'd like to acknowledge and thank to my graduate advisor, Dr. Brian Veitch. Your patience, understanding and support has truly been appreciated. Conducting this research has been insightful and being a part of your research team has been a valuable and rewarding experience.

I'd also like to acknowledge the support of the Memorial University Safety at Sea research team. Their knowledge of virtual environment development, pedagogical strategies, and offshore safety processes has been a major factor to the success of this research. It was a pleasure to work with you all.

A special thank you is due to Mashrura Musharraf. Your exciting and interesting lecture on Human Reliability Methods is what led me to pursue this research. Thank you for all the support and guidance.

Thanks are also due to my co-investigators, Kyle Doody and Sinh Duc Bui. Working as a team to accomplish a common goal has been a rich and rewarding experience. I wish you all the best in your future endeavours!

Lastly, I'd like to thank my family. To my parents, your unwavering support and encouragement has always given me confidence to strive for success. You are appreciated more than you will ever know. To my beautiful children, thank you for reminding me to always reach for the stars and never stop learning! May you always pursue your dreams and goals with passion and determination.

Table of Contents

List of Figures

List of Tables

List of Acronyms

AVERT – All-Hands Virtual Environment Response Trainer

BN – Bayesian Network

CPT – Conditional Probability Table

FPSO - Floating, Production, Storage and Offloading

HRA – Human Reliability Analysis

ICEHR – Interdisciplinary Committee on Ethics in Human Research

IDAC – Information, Decision, and Action in crew Context

OERC - Ocean Engineering Research Centre

OIM - Offshore Installation Manager

PA – Public Address

PSF – Performance Shaping Factor

SIS – Simulator-Induced Sickness

SBML - Simulator-Based Mastery of Learning

SMS - Safety Management System

SSQ – Simulator Sickness Questionnaire

VE – Virtual Environment

VIMS – Visually Induced Motion Sickness

Chapter 1 Introduction

Offshore oil and gas accidents have resulted in loss of life, loss of capital assets, and damage to the environment. Accident investigations provide valuable information about the root causes of accidents. The root causes of a significant number of accidents have been attributed to human error, or employees failing to follow safety policies and procedures (Dimattia et al., 2005). However, the information from accident investigations is a reactive way of gathering information about human reliability.

Human reliability assessment (HRA) techniques are used to predict people's performance, including their response to emergencies. Human performance data required to perform HRA in emergency scenarios are not readily available. Due to lack of available data on human performance in emergencies, HRA analysis is usually done using expert opinion (Groth et al., 2014). Expert opinion, while valid in certain circumstances, can lead to uncertainty in results (Musharraf et al., 2013). This uncertainty can be reduced by use of a virtual environment (VE) to gather data on people's performance during emergency situations. Another limitation of most HRA methods is that they do not consider dependency amongst performance shaping factors and associated actions (Musharraf et al., 2013).

Offshore incidents such as the Piper Alpha in 1988 and the Deepwater Horizon in 2012 are examples where human factors affected performance during an emergency (Flin & Slaven, 1995; Flin et al., 1996; Norazahar et al., 2014). The incidents began due to technical failure, however human performance during the emergency situation was inadequate due to lack of emergency preparedness. The inquiry into the Piper Alpha accident (Cullen, 1990) concluded that the Offshore Installation Manager (OIM) may not have been competent in dealing with emergency situations. The OIM was assumed to be appointed by job-specific requirements, rather than based on an assessment of their ability to handle and manage an emergency situation (Flin & Slaven, 1995). Deepwater Horizon was a major accident that required personnel to evacuate the platform. An investigation into the accident concluded that there was a lack of emergency preparedness, which resulted in personnel errors in the evacuation procedures. Personnel had difficulty in moving through the emergency routes due to emergency conditions, and were disorganized at the muster point (Norazahar et al., 2014). Current training practices do not allow for training in conditions other than ideal. By conducting a HRA using a simulated environment, information regarding personnel's performance during realistic emergency situations can be gathered, which might be useful for providing additional training and heightened preparedness.

The research presented in this thesis proposes a Human Reliability Assessment technique that utilizes a Bayesian network (BN) model to assess human reliability in emergency scenarios using a VE. The BN model has the ability to characterize dependency among different attributes (Groth et al., 2012b). Attributes can be any factor that the researcher is interested in. For this research, the attributes of interest are performance shaping factors (PSFs) and human errors. This approach provides a means to quantify the impact of different factors on an overall event. The virtual experiment technique provides a way to collect data to quantify a BN. By using a BN in combination with a VE, the uncertainty in results can be reduced, and a person's response in emergency situations can be assessed.

Musharraf et al. (2018, 2014, 2013) presented a VE based technique to predict an individual's failure probability in emergency situations. They used data from a VE to quantify a BN to assess human reliability in offshore emergency conditions. Three PSFs were considered in the experiment: training, visibility, and complexity. These were selected based on a task analysis and the capability of the VE. The factors were varied at different levels, and data to quantify the BN were collected. Results from the study indicated that this method is a viable way to overcome uncertainty associated with expert opinion, to create a realistic way to demonstrate dependency amongst PSFs, and to estimate human error more accurately. However, choice of the PSFs in the study was constrained by the VE's capability and did not provide an ideal representation of offshore emergency scenarios.

The research presented in this thesis extends the same experimental technique, but aims to address the limitations of the previous study. Scenarios for the VE were developed that accurately represent emergency situations on an offshore oil and gas facility. The PSFs selected for this research are complexity, stress, and uncertainty.. The lack of knowledge about how these factors affect individuals' performance in emergencies undermines organizations' ability to manage safety. This is especially important in the offshore oil and gas industry where facilities tend to be remote and external emergency response is not immediate (Flin et al., 1996).

This thesis starts with a literature review into the background of BN, VE, and HRA techniques. The experimental design and procedures are described in sections 3 and 4. Data collection and analysis follow, and finally results and discussion are presented. The conclusion section summarizes the thesis and suggests future lines of enquiry.

Chapter 2 Literature Review

A literature review was completed on the various aspects of the proposed experiment before research began. This was done to determine the current state of knowledge of HRA methods, BNs, and virtual environments. The review investigated what other researchers have done. This information was considered in the development of this research to ensure that the current work would contribute to the field of HRA research and practice.

The subjects reviewed and researched include current HRA methods (Section 2.1), PSF selection (Section 2.2), BN overview (Section 2.3), BN approach to HRA (Section 2.4), and virtual environments (Section 2.5).

2.1 Conventional HRA methods

HRA methods have been used since the 1960s (Bell & Holroyd, 2009). Early HRA methods considered a human as a mechanical component that can fail while performing a task (Kirwan, 1994). These methods are referred to as the first generation methods. Human Error Assessment and Reduction Technique (HEART) and Success Likelihood Index Method (SLIM) are examples of first generation methods. These methods focused on physical tasks that caused or may cause an error. As interest in HRA methods increased, it was realized that the cognitive aspects of humans need to be considered when assessing human reliability (Bell & Holroyd, 2009). This led to the introduction of second generation HRA methods. Rather than focusing on the frequency of the error, the second generation methods focus on the causes of the errors. Simplified Plant Risk Human Reliability Assessment (SPAR-H) is an example of second generation methods. Both first and second generation methods aim to assess a human error probability (HEP). Quantification of HEP often involves investigating the effect of performance shaping factors (PSFs) on human performance. These are factors that can influence human performance (Groth & Mosleh, 2012a). The majority of second generation HRA methods use expert opinion to define the relationship between PSFs and human performance. The HEART, SLIM, SPAR-H, and IDAC methods are outlined below to give an understanding of conventional HRA methods and how expert opinion influences HEP calculation and overall Human Reliability Assessment.

HEART is a simple and quick assessment that was first outlined in 1985 (Kirwan, 1994). It elicits expert opinion, requiring a minimum of one assessor who has adequate education and experience in the field of interest (Bell & Holroyd, 2009). An HEP is calculated by selecting from a list of generic nominal HEP values and a specific list of PSFs (called Error Producing Conditions (EPCs) for the HEART method) that relate to the task to be assessed. The nominal HEP list and EPC list are specific to the HEART method, however selection from the list is based on expert opinion (Kirwan, 1994).

An expert will select a certain number of EPCs to be included in the assessment from the provided lists based on their knowledge and experience in the event being assessed (Kirwan, 1996). The values associated with the selected EPCs are known as the Maximum Effect (ME) and are specific to each EPC. The next step in the HEART assessment is to assign an Assessed Proportion of Effect (APOE) to each EPC selected by the expert and calculate an Assessed Effect (AE) as per equation 1. The APOE is the percentage of effect that each EPC is estimated to have on the overall HEP (the value is between 0 and 1). The selection of APOE is based on expert judgement and very little guidance has been given on how to assign it (Kirwan, 1997). The calculated AE of each EPC is then multiplied by the initial generic nominal HEP selected (equation 2).

$$AE_{EPC_i} = \left(\left(ME_{EPC_i} - 1\right) * APOE_{EPC_i}\right) + 1 \qquad (1)$$

$$HEP = Nominal\ HEP * \prod_i AE_{EPC_i} \qquad (2)$$

As described, the HEART method is very subjective and the final HEP can vary drastically depending on values selected from the predetermined lists and expert opinion (Kirwan, 1994). It is also specific to the situation being assessed and the end result may not be transferable to other situations (Kirwan, 1996).

SLIM uses a cohort of experts to complete an assessment. The cohort comprises of at least three experts who are knowledgeable and experienced in the subject matter being assessed. They breakdown the event to be analyzed into manageable tasks, and determine a set of PSFs through discussion and review. The cohort continues to use expert judgement to rate each PSF on a scale of 1 to 9 (sub-optimal to optimal) (R in equation 3), and allocate a weighting to the PSF by level of importance (W in equation 4). A Success Likelihood Index (SLI) is calculated and input into a simple logarithmic equation, as shown in equations 3 and 4. The equation (4) converts the SLI into a probability. The values of a and b in equation 4 are estimated values based on previous SLIM calculations in a similar situation (Kirwan, 1994).

$$SLI_j = \sum_i R_{ij} W_i \tag{3}$$

$$\log P = a\, SLI + b \tag{4}$$

The process of PSF selection and weighting can vary depending on the selected cohort (Kirwan, 1994). Dimattia et al. (2005) presented a SLIM technique for offshore platform musters where a cohort of 24 experts selected 6 PSFs for the assessment of the event. In comparison, the SLIM method that Musharraf et al. (2013) completed to validate her proposed BN Model used 3 PSFs and only 1 expert. As well, like HEART, SLIM is based on the situation assessed and cannot be updated as new information about a situation becomes available (Musharraf et al., 2013).

The SPAR- H method is widely used in the nuclear industry (Groth & Swiler, 2013). The most recent version was updated in 1999 and its format is based upon the HEART method (Bell & Holroyd, 2009). The SPAR-H Method has a set list of 8 PSFs with a set level describing each PSF. Three PSFs are normally selected for evaluation by an expert assessor (Bell & Holroyd, 2009). This method employs a pre-determined set of worksheets to calculate a HEP. This worksheet can vary slightly depending on the level of HRA analysis the assessor is completing (Bell and Holroyd, 2009). While the equations in the worksheet remain the same, the multipliers for PSFs and nominal HEP values are different depending on the type of assessment selected (Groth & Swiler, 2013).

It is suggested that the worksheet simplifies the HEP calculation process and therefore is very useful when a detailed assessment is not required (Bell & Holroyd, 2009).

The Information, Decision and Action in crew Context model (IDAC) is a second generation HRA Method. Second generation models are predictive and include experimental validation. These models are different from the first generation models, which are primarily based on subjective assessment (Chang & Mosleh, 2007). IDAC is a causal model of an operator's behavior at the cognitive level. The model probabilistically simulates operators' behavior under the influence of PSFs present during an event (Chang & Mosleh, 2007). It is primarily used for operational type assessments and lacks information with regard to maintenance activities or emergency response decisions making (Groth & Mosleh, 2012a).

Behaviour is assessed by an expert using generic rules that determine the dynamic responses of the operator (Chang & Mosleh, 2007). It is represented by three kinds of responses. They are the information response, decision response, and action response. The information response determines the information to be assessed, processed and prioritized, and defines the problem in the scenario that is to be addressed. The decision response determines a strategy and makes decisions or develops solutions based on knowledge. The action response creates an action sequence based upon the decisions made in the previous response step. Cognitive and psychological states of the operator being assessed factor into all these responses. (Musharraf et al., 2013). The IDAC model consists of internal and external PSFs organized in a map of information that incorporates each of these responses. Assessment is conducted in a bottom up approach from the scenario to be assessed to the final human performance (Groth & Mosleh, 2012a).

An HEP is calculated using several data sheets that provide the information gathered from site visits and operator interviews regarding the scenario and event to be assessed in combination with assigned probabilities of each response as a function of the PSF state. (Chang & Moelsh, 2007; Groth & Mosleh, 2012b).

As noted in all the HRA methods reviewed, expert judgement and opinion is a very important factor into the HEP calculation. Results can vary across experts with different opinions and background. It is also noted that the data is specific to an event or situation being assessed. There is a need to reduce or remove the dependency on expert opinion for HRAs. The method of utilizing a simulator to gather real-time data that is presented in this thesis addressed that need.

2.2 PSF Selection

PSFs are the factors that can affect human performance (Groth & Mosleh, 2012a). For many HRA methods, PSFs are applied in a human error probability (HEP) calculation to increase or decrease the value based on the affect it may have. Careful selection of PSFs is an important step in HRA development (Groth & Mosleh, 2012b). A review was completed of the list of PSFs for conventional HRA methods. As seen in section 2.1, conventional methods have a specified list of PSFs for HEP calculation (Kirwan, 1994; Bell & Holroyd, 2009). The PSFs and their corresponding values used in traditional HRA methods are usually specific to the purpose and industry for which the method is applied (Kim & Jung, 2003). When using a specific HRA method to calculate HEP, the PSF selection is based on expert opinion and the situation to be evaluated.

There is not a standard set of PSFs to be used, so when comparing results between assessors or between methods, there can be errors or deficiencies in the results (Groth & Mosleh, 2012a; Kim & Jung, 2003). This is seen as a challenge in selecting PSFs to be used for the analysis of an experiment. Several researchers have also recognized this challenge and have made attempts to create a generic PSF list for HRA practitioners to use in assessments.

In Groth & Mosleh (2012a), the author reviews multiple HRA methods and databases to create a comprehensive list of PSFs that can be used in HRA methods. It is stated that previous PSFs in use are not defined specifically enough to ensure consistent interpretation across various methods. Groth's paper (2012a) describes a hierarchy of PSFs and categorizes them into groups such as organizational, team, person, stress, and machine based factors. Each PSF in the categories are defined so they can be applied to any HRA methods.

The author suggests that the hierarchy can be used for different analysis methods, such as computer modeling, manual interpretation, and HEP calculations in HRA methods. This point is applicable to the experiment described in this thesis as there is no specific list for the method proposed.

Another author, Rangra et al. (2015), has also completed research on PSFs for use in HRA assessments. Rangra et al.'s focus is specific to emergency scenarios for the railway industry. The paper presents a relevant list of PSFs to be used for a railway specific HRA.

Kim & Jung (2003) has completed work similar to Groth & Mosleh (2012a) and Rangra et al. (2015). This author's focus is on proposing a taxonomy of PSFs for emergency tasks in nuclear power plant operations. The process the author completed to create a full set of PSFs was to review traditional HRA methods and several databases of HRA and incident investigation information. The author reviewed the PSF information, selected a list of PSFs, and defined sub items that can be applied to certain emergency situations.

The PSF list was categorized into four sections – human, system, task, environment – which were then defined further into subgroups. The subgroups are based on frequency of use in HRA methods. The terminology for each PSF is specific and practical.

Groth & Mosleh (2012a), Rangra et al. (2015), and Kim & Jung (2003) agree that the situation the PSFs are being applied to is required to be clearly defined in terms of the unit of analysis (e.g. person versus team), and if there are sub events to be considered. Careful definition is important so PSFs selected are independent and do not overlap (Groth & Mosleh, 2012a). They also all state that the definition of PSFs is important to selection for HRA. Training and experience are examples of PSFs that require further definition. Experience differs for each member of a crew, but training could be the same. They are often considered to be the same for accident investigation using HRA methods so having them further defined would make the assessment more valid (Rangra et al., 2015). By considering the process by which the hierarchy is created by Groth & Mosleh (2012a), Kim & Jung (2013), and Rangra et al. (2015), selection of PSFs can be more informative and potentially provide more accurate results in a HRA. PSFs that are defined accurately can be incorporated to the correct task context and ensure that the selected PSFs do not overlap (Kim & Jung , 2003).

Groth & Mosleh (2012a) suggests starting with an expanded PSF list and as more details about the situation being assessed is known, narrowing down the PSFs to be used for a HRA. This method is followed for PSF selection for the experiment presented in this thesis. The proposed HRA method involves data collection from a virtual environment (Moyle et al., 2017). A list of PSFs was selected based on the literature review discussed above. The list was narrowed down based on criteria such as the ability to replicate the PSF in a virtual environment, if a PSF could be present in the situation to be assessed (i.e. if a high and low level can be created for assessment). This review was necessary in order to select PSFs that were interesting and can provide adequate results for assessment (Moyle et al., 2017). The selection of PSFs guided the experimental design and data collection process. Further details on PSF selection is in section 3.4.

2.3 BN Overview

The experiment discussed in this thesis uses a BN technique to model dependency between PSFs and human error, and to calculate a HEP (Moyle et al., 2017). A BN is a statistical model that is used to represent causal dependency amongst different variables. It is composed of a set of nodes (variables) that are connected by arcs that represent causal dependency. The independent nodes (i.e. nodes with no predecessor) are called root nodes. Probabilities of root nodes are based on the initial belief of the possibility of each possible state of the corresponding variable. The dependent nodes are called child nodes and are associated with a conditional probability component. This component captures how different possible states of parent nodes can affect the probability of the child nodes (Musharraf et al., 2013). Figure 1 shows a simple BN that represents the causal dependency between PSFs and human error. Here PSFs are the root nodes, and errors are the child nodes. Each node can have several possible states (e.g. PSFs can be high or low, errors can be yes or no).

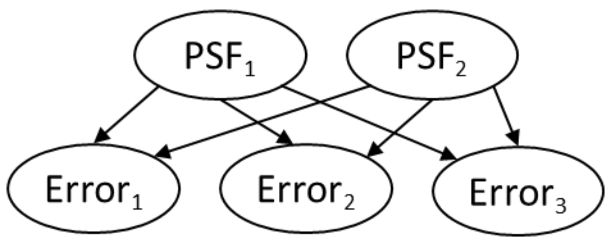

Figure 1: Causal dependency between PSF and Human Errors

Bayes' Theorem (equation 5) is used to calculate the marginal probability of each child node when information on the root node and conditional probabilities on variable states is known. The probability of event A (Errors), given that event B (PSF) has occurred, can be calculated using equation 5 (Haldar & Mahadevan, 2000).

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (5)$$

By applying the product rule, the marginal probability of Event A can be written in terms of a conditional probability distribution, as shown in equation 6 (Haldar & Mahadevan, 2000).

$$P(A) = \sum_i P(A|B_i)\,P(B_i) \qquad (6)$$

The conditional probability distribution is presented in the form of a Conditional Probability Table (CPT) within the BN model. Each dependent node has an associated CPT. Table 1 shows an example CPT for $Error_1$ from Figure 1. Given the states of the parent PSF nodes PSF_1 and PSF_2, the probability of $Error_1$ happening can be calculated (Musharraf et al., 2013). Equation 7 illustrates how the probability of $Error_1$ happening can be calculated, given the conditional probabilities (Moyle et al., 2017).

Table 1: Example CPT

PSF_1	*High*		*Low*	
PSF_2	*High*	*Low*	*High*	*Low*
$Error_1$ - Yes	1	0	1	0
$Error_1$ - No	0	1	0	1

$$P(Error_1 = Yes)$$
$$= P(Error_1|PSF_1 = High, PSF_2 = High) \times P(PSF_1 = High) \times P(PSF_2 = High)$$
$$+ P(Error_1|PSF_1 = High, PSF_2 = Low) \times P(PSF_1 = High) \times P(PSF_2 = Low)$$
$$+ P(Error_1|PSF_1 = Low, PSF_2 = High) \times P(PSF_1 = Low) \times P(PSF_2 = High)$$

$$+ P(Error_1|PSF_1 = Low, PSF_2 = Low) \times P(PSF_1 = Low) \times P(PSF_2 = Low) \qquad (7)$$

The intricacy of this equation highlights that quantifying BN can be complex and are normally completed using software, such as Genie Modeller (Anonymous, 2017).

A BN is useful in applying updated evidence of a node. This concept allows the investigation of the cause-effect relationship between the variables. Both forward analysis (event B to A) and backward analysis (event A to B) are possible. This is useful in determining root causes of events when additional evidence on a node's occurrence is provided (Groth & Swiler, 2013; Musharraf et al., 2014).

2.4 BN approach to HRA

Traditional HRA methods consider specific equations to calculate a HEP. However, application of a BN in HRA is a practice that is gaining popularity amongst practitioners (Martins & Maturana, 2013). The benefit of a BN is that it can demonstrate the influence of PSFs on a HEP calculation (Groth & Mosleh, 2012b). By modelling a HEP calculation using a BN, the strength that a PSF can have on human errors can be quantified (Groth & Mosleh, 2012b). A BN allows for manipulation of its nodes which can determine if a PSF negatively or positively affects human performance. The BN HRA model can be populated by data and information from several sources including databases, a panel of expert opinions, or as this research aims to demonstrate, by information gathered from task performance in a virtual environment.

Groth & Swiler (2013) presents a method of transitioning the SPAR-H method into a BN. The PSFs listed for the SPAR-H method become the input (root nodes) of a BN model. This shows HRA researchers that traditional methods can be combined with a more modern approach in order to assess the influence of various dependencies on a situation (Groth & Swiler, 2013).

In another paper, Groth & Mosleh (2012b) derives a BN from the IDAC model and HERA database. The HERA database was developed by the US Nuclear Regulatory Commission (NRC) and maintains a list of information about human factors from various nuclear power plants (Groth & Mosleh, 2012b). The author selected a list of PSFs from the above methods based on the incident being reviewed from the nuclear power industry. For the transition to a BN, PSFs were connected via arcs depending on an assessed value. The PSFs were then assigned to an error context group which defined how likely the PSF will result in a human error. The BN in this model shows causal dependencies between PSFs and from PSFs to an error context (Groth & Swiler, 2013).

The research provided in Martins & Maturana (2013) reviews the use of a BN to analyze human reliability. The method is applied to predict the risk of collision accidents of ships in the maritime industry. The use of BNs for a formal safety assessment in the maritime industry was approved by the International Maritime Organization (IMO) in 2006 (Martins & Maturana, 2013). The BN was developed based on a fault tree of a ship collision. A fault tree shows the connection of events to an overall top event. This mapping allows for a simple transition to a BN (Martins & Maturana, 2013). The authors state the BN would be able to determine the elements (PSFs) that greatly impact the event or situation, not just an end risk value that is calculated in a fault tree (Martins & Maturana, 2013). The fault tree is extended to include a BN of PSFs that provide inputs to the basic events of the fault tree. The basic events that are expanded are potential human errors that may occur (Martins & Maturana, 2013). Similar to the work by Groth & Mosleh (2012a) and Groth & Swiler (2013), the data for the BN calculated was populated by expert opinion. Martins & Maturana (2013) provides a method that is different than Groth & Mosleh (2012a) and Groth & Swiler (2013), however it is still focused on the use of subject matter experts to provide data to populate a BN to calculate a HEP.

Another author has completed research into utilizing BNs for safety assessments in the maritime industry. Hänninen et al. (2014) presents a BN model that demonstrates the link between elements of the maritime safety management system (SMS) (referred to as a submodel) and past safety indicators such as accident involvement, reported incidents, and port inspections (referred to as the main model). Like the models previously described, the data for the BN model is elicited from experts experienced in the industry and situation being assessed. For development of the submodel, the experts involved in this research selected points of interest from the SMS, which became nodes of the submodel BN. The submodel links to the main model via a node called "safety". The "safety" node is an abstract variable. This is to represent that the main model nodes should not be viewed as safety management indicators directly, but can influence the belief of the nodes in the submodel. The method is more focused on interactions between business areas of the SMS rather than human behavior, and therefore can predict consequences of changes in safety management procedures (Hänninen et al., 2014).

Musharraf et al. (2013, 2014, 2018) has also completed work on the development of a BN model for HRA methods. The author has investigated different data gathering methods and various applications of a BN in assessing human performance during offshore oil and gas emergency situations. Musharraf et al. (2013) begins by converting an initial expert opinion on performance shaping factors to an initial probability value using the Dempster-Shafer Theory Method. This method uses a mathematical formula to convert a qualitative belief to a quantitative value (Musharraf et al., 2013). This value is the prior probability that is input to a BN for the HRA model in order to calculate a HEP. The PSFs were selected from the IDAC model, similar to Groth & Mosleh (2012b). The BN model was developed by connecting the selected PSFs to task events from the situation being assessed. The connections illustrate the dependencies between PSFs and the task events.

In Musharraf et al. (2014), a BN model was developed using a different approach. PSFs were selected based on a review of the situation to be assessed and how they could be replicated in a virtual environment. The use of a virtual environment to gather data is innovative as it relies upon the capabilities of a simulator to capture performance data and uses this data as inputs into the BN model. This process replaces the dependency on expert opinions for HEP calculation and HRA modelling (Musharraf et al., 2014). However, this model calculated a general HEP for participants as a group rather than for individuals. This does not allow for an assessment to be completed accurately on an individual in order to determine how PSFs specifically affect a participant. Therefore, Musharraf et al. (2017) expanded again on the previous research to include more inclusive PSFs and a BN model for assessing individual reactions to PSFs and to calculate a HEP for each participant.

Along with the movement to develop BN models for use in HRA analysis, there has been a review of how methods are developed and recommendations made on how to improve them (Mkrtchyan et al., 2015). One of the recommendations is to enhance the explanation of how the BN model was developed. A finding from the review stated above is that BNs are typically presented to the reader without an explanation as to how or why it was structured in a particular way (Mkrtchyan et al., 2015). Clarification of how to build BNs, such as node selection and definition, structure development, and how to populate conditional probability tables (CPTs), needs to be included so researchers can appreciate and understand how to develop a BN for use in a HRA. For example, defining the node based on the application and the definition of the node state (yes or no, high or low) can further assist HRA practitioners in developing a BN for their own assessments.

The review also identified several positive features in the use of BNs for a HRA. Positive features included ability to graphically represent a complex situation showing dependency between PSFs and task errors, and the ability to combine data from various sources in a mathematical model (Mkrtchyan, et al., 2015). These aspects have been echoed by authors such as Groth & Swiler (2013) and Musharraf et al. (2013, 2014,

2017) in their research. Groth & Swiler (2013) also states that while the BN can be used to perform a causal reasoning between PSFs and task errors, it can also be used to determine the strength of influence of different PSFs on human error (Musharraf et al., 2014). The mathematical concept of this is outlined in section 2.3.

HRA data currently comes from a variety of sources, such as expert opinion, previously calculated data, and a database of recorded data from actual incidents. Variability in a BN calculation using expert opinions arise due to the different backgrounds and knowledge of experts (Hanninen et al., 2014). Mkrtchyan et al. (2015) states expert data are an important source of data, but also says that capturing data from a simulator can be useful for populating a BN and can eliminate the need for expert opinion. Again, Groth & Swiler (2013), Rangra et al. (2015), and Musharraf et al. (2013, 2014, 2018) state the importance of simulator data in the development of using a BN. Musharraf et al. (2013, 2014, 2018) use simulator data as the basis for research presented in all three papers.

2.5 VE Overview

Using a virtual environment for safety training is expected to improve emergency preparedness for regular and ad hoc workers on offshore platforms. It is also expected to reduce the induction period for workers new to an offshore platform by providing a medium to introduce them to the new place of work before they get there. Simulators also give insights into a participant’s reaction to stressful situations and can be used as a screening tool to identify those who do not do well during emergency situations (Flin & Slaven, 1995).

This section reviews the use of virtual environments to gather information on a participants' strengths and weaknesses during performance in different egress and evacuation conditions. The virtual environment used in this research is a simulator called the All-hands Virtual Environment Response Trainer (AVERT) (Veitch et al., 2008). AVERT (Figure 2) allows users to gain experience in naturalistic emergency scenarios in a safe and controlled manner. AVERT replicates an offshore oil and gas facility in which users can learn knowledge and skills regarding offshore emergency safety procedures. Training programs can be designed in AVERT to introduce users to platform layout, alarm types, potential hazards, and appropriate responses (Moyle et al., 2017).

Users are required to undergo a training program that introduces them to an offshore oil and gas facility layout, alarm types, and how to respond to the alarms in an appropriate manner. The experiment presented in this thesis expands on this training to expose participants to realistic emergency scenarios that are at a higher stress level than the training scenarios. This allows researchers to gather information on performance in an emergency scenario that would otherwise be unavailable until after an emergency occurred (Moyle et al., 2017).

Musharraf et al. (2014, 2018) used AVERT to collect data on participants' performance in emergency situations. They created scenarios using three different PSFs and recorded performance data for each participant. This data was input into a BN-HRA for analysis resulting in a HEP calculation. Groth et al. (2014) also incorporated simulator data in a HRA to enhance HEP calculations.

Figure 2: Example AVERT Scenes

Simulators are a beneficial way of capturing human response data in controlled conditions. (Mkrtchyan et al., 2015). There are arguments that simulator data does not represent expected conditions, however a simulator can enhance or diminish a situation by incorporating performance shaping factors. PSFs are manipulations of the environment and are usually factors of interest in an experiment (Moyle et al., 2017). By inputting the collected data into a BN, the probability of error (whether a participant completes a task successfully or not) given a PSF is present can be calculated. The value is more accurate than calculating a probability of error without knowing the influence of factors (Groth et al., 2014).

Chapter 3 Design of Experiment

3.1 Methodology

The methodology followed for the experiment is outlined in Figure 3. The context of interest was emergency response on an offshore petroleum installation. The author developed and implemented the experiment in conjunction with an experiment for the assessment of emergency procedure skill retention.

A task analysis of this activity was the first step in the research. The tasks of interest were the safety compliance procedures associated with emergency situations. After the task analysis, a list of factors that can influence task performance was generated. The next step was to model the dependency between the PSFs and task performance using BN. Scenarios were created in AVERT by the author to gather the data required to quantify a BN (Moyle & Veitch, 2017).

The author collected data by observing the performance of 37 volunteer human participants in the virtual scenarios (Moyle et al., 2017). The data were integrated into the BN and the reliability of participants was assessed. Sections 5 and 6 further describe this process in detail.

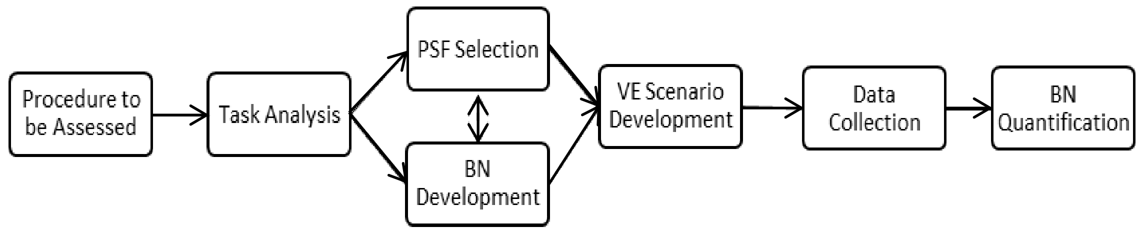

Figure 3: HRA Experiment Methodology

3.2 Experimental setup

The 37 human participants were recruited from a cohort that completed an initial Simulator-Based Mastery of Learning (SBML) experiment (Smith & Veitch, 2019). Participants were required to complete a training program that was developed in AVERT that would provide all the necessary knowledge and skills required for completing an offshore emergency scenario. The training program introduced the participants to the virtual model of a Floating, Production, Storage and Offloading vessel (FPSO), and provided an environment that taught offshore emergency rules and requirements in training blocks. Participants were tested and retrained until they were fully competent in the training material. Figure 4 shows how the SBML, Retention and HRA experiments relate to each other. The SBML & Retention portion of the flow chart shows the training and testing sequence participants completed. They were first introduced to the AVERT program and the model of the FPSO in a habituation Stage. Participants then completed a series of training (block numbers 1 through 8) and testing scenarios that gradually increased in difficulty. Once all testing scenarios were passed successfully, the participant was deemed to be competent in performing offshore safety procedures in AVERT (Smith & Veitch, 2019). 6-9 months after the initial SBML training, the participants completed another study to determine the effectiveness of the SBML training in retaining the information taught, including the extent to which their skills had declined. Participants were exposed to the virtual environment again and were re-tested using the initial testing scenarios from the SBML training. Data on the retention of skills and knowledge of offshore emergency rules were evaluated (Doody et al., 2017). Participants were tested, and retrained if necessary in training blocks, until all scenarios were completed successfully.

If participants completed the retention study successfully and thus demonstrated competence in performing offshore safety procedures, they were invited to participate in the experiment addressed in this thesis, which is shown as the HRA study portion of Figure 4. This experiment assessed a participant’s application of the skills and knowledge learned during the SBML and Retention study by exposing the participants to a series of 8 scenarios with varying degrees of performance shaping factors. This is shown in the figure as HRA 1 through HRA 8. It has to be noted that the participants did not complete the scenarios in the same order, but were exposed to a random sequence.

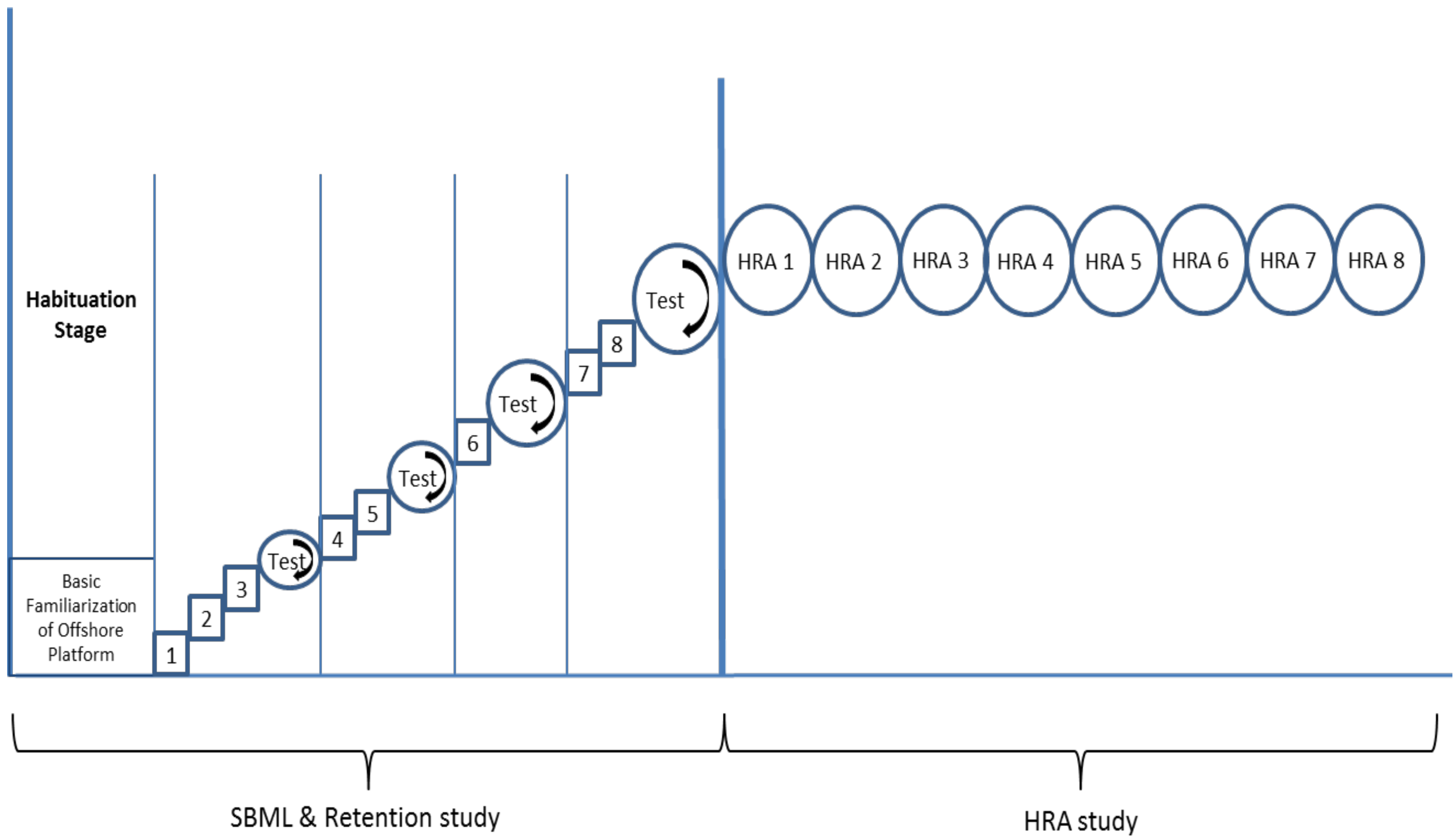

Figure 4: Experimental Setup Flow Chart

3.3 Procedure Selection & Task Analysis

The muster sequence required to be completed during an offshore emergency situation is the procedure to be assessed for the research. The area within the AVERT model to be used was the Accommodation block.

A task analysis from Dimattia et al. (2005) on muster actions required during an emergency situation was used as the basis for creating the BN model and development of scenarios. Dimattia et al. (2005) presents a list of actions required, beginning at the time of muster initiation, through to platform egress and evacuation. This is when individuals are more likely to be exposed to hazards like heat, smoke, pressure, and to high levels of stress.

The task hierarchy of egress and evacuation during a muster scenario is outlined in Figure 5. This list was developed by a cohort of experts that participated in the SLIM analysis for Dimattia et al. (2005). The required actions are independent of the type of emergency and are to be completed regardless of the type of incident. Incidents that require actions for platform abandonment were not considered for this research and are not discussed. The goal of this thesis is to investigate the influence of the selected PSFs on task performance during a muster sequence only (Moyle et al., 2017).

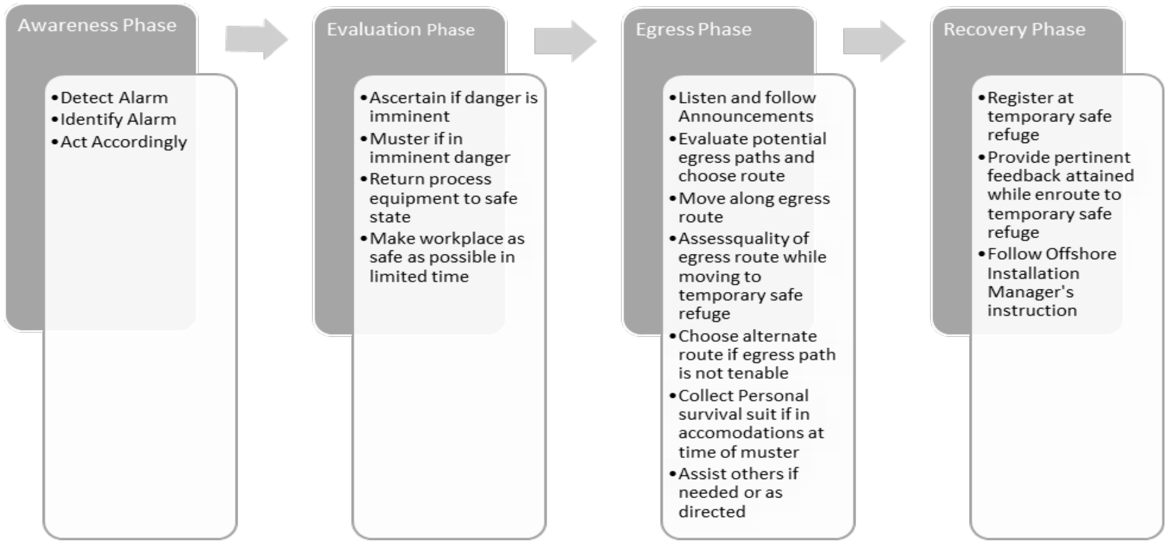

Figure 5: Task Hierarchy for safety procedures during an offshore emergency situation. Adapted from DiMattia, et al. (2005)

3.4 PSF Selection and Experimental Design

Varying the intensity of PSFs can affect participants' ability to complete emergency scenarios successfully. Observing participants' performance at different levels of the PSFs can help assess their strengths and weaknesses during an assigned task (Musharraf et al., 2014). According to Boring (2010), keeping the PSF list relatively short for analysis (three or four) reduces the chance of having PSFs that overlap in definition. This agrees with Kirwan (1994) who states that 3 or 4 PSFs give more accurate results in an HRA. The choice on how many to use is often based on expert opinion, or can be a systematic process where a set is reviewed, quantitative analysis is performed, and the PSF set refined (Groth & Mosleh, 2012b).

Three PSFs were selected based on their relevance to the context of the selected task. The factors selected were complexity, stress, and uncertainty (Moyle et al., 2017). These factors can be defined in a variety of ways. PSFs were selected based on a set of criteria, such as the ability to implement them in AVERT, their relevance to real-life situations, and whether they could be controlled to minimize the effects of variables other than the chosen PSF. (Moyle et al., 2017). Figure 6 shows the evolution of the PSFs from selection to implementation in AVERT. The process started with a high level election of the PSFs, and then each PSF was focused to determine variables that could be varied in each scenario.

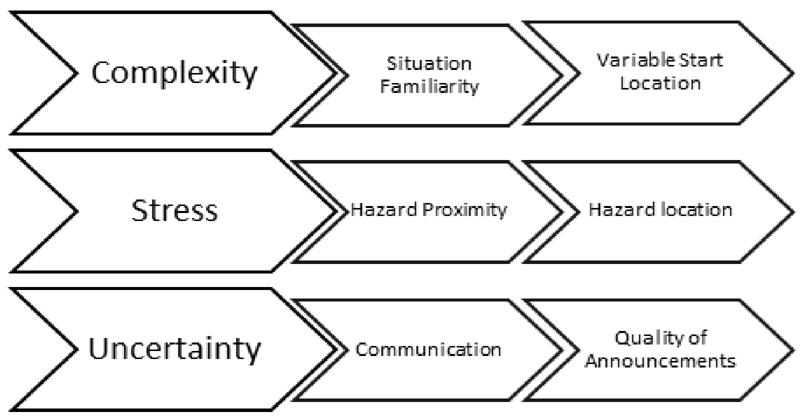

Figure 6: PSF Definition

Each PSF was assigned a high and low level of implementation. The high level for *complexity* was set by starting at a familiar location in AVERT. The cabin area of the accommodation was used for this as it was familiar to participants due to the SBML training. The low level of *complexity* was set by starting the scenario at a less familiar location. This was selected to be the Bridge on the navigation deck. *Stress* was varied by changing the proximity of the participant to a hazard. At the high level, the participant had no chance of interaction with a hazard. The hazard locations were in an area that the participant would not have access to. At the low level, the participant could observe a hazard depending on route selection during the scenario. *Uncertainty* was varied by changing the quality of the information communicated in the announcements made during the emergency scenarios. High level provided an announcement having all relevant information for the participant to make an informed decision on how to muster, such as if a route was blocked by a hazard and where the hazard was located. A low level provided less pertinent information about the scenario and just directed the participant to muster (Moyle et al., 2017).

Figure 7 shows the experiment design schematic. This figure demonstrates the link between PSF, task to be completed, and the response assessed from the performance in AVERT.

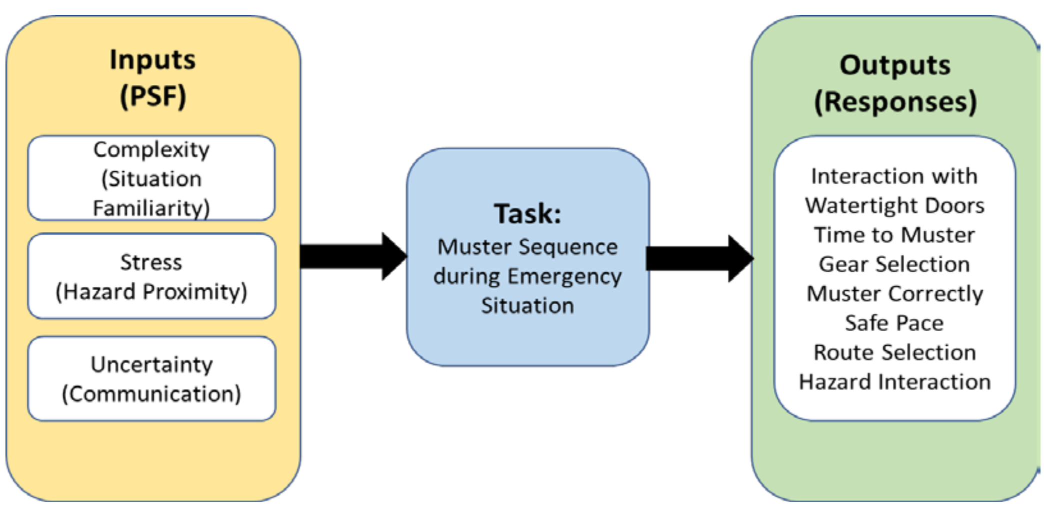

Figure 7 Schematic of Experimental Design

3.5 BN Development

The BN model, as seen in Figure 8, shows the dependency between the PSFs and the task performance, and then between the task performance and the completion of emergency safety procedures.. Tasks were defined as the child node and represent the critical level of the BN. Failure or success of each task depends on the states of the PSFs, which are the root nodes in the BN. The relationship between the PSFs and individual task performance is shown in the upper two levels. The lower two levels show how the probability of individual task performance can be combined to get an overall probability of successful muster. This was done is two steps. Task performance probabilities were first combined to get a dependent probability for two categories: compliance with safety procedures, and spatial awareness. Probabilities of these categories were then combined to get the final probability of successful muster. The reason behind taking an extra step to combine tasks into different categories is to make the computation more manageable (Moyle et al., 2017).

Data needed to fully quantify the BN shown in Figure 8 are the prior probabilities of PSFs, the CPTs of task nodes, the CPTs of intermediate category nodes, and the CPT of the final muster completion node. Two types of data are needed to define the relationship between the PSFs and task performance. First, the prior probabilities of the PSF nodes are needed. Prior probabilities refer to the initial belief about the likelihood of each state of the PSFs. Next, the relationship between the PSF and task performance is needed to populate the CPT tables for each task performance node. This is the data that is collected

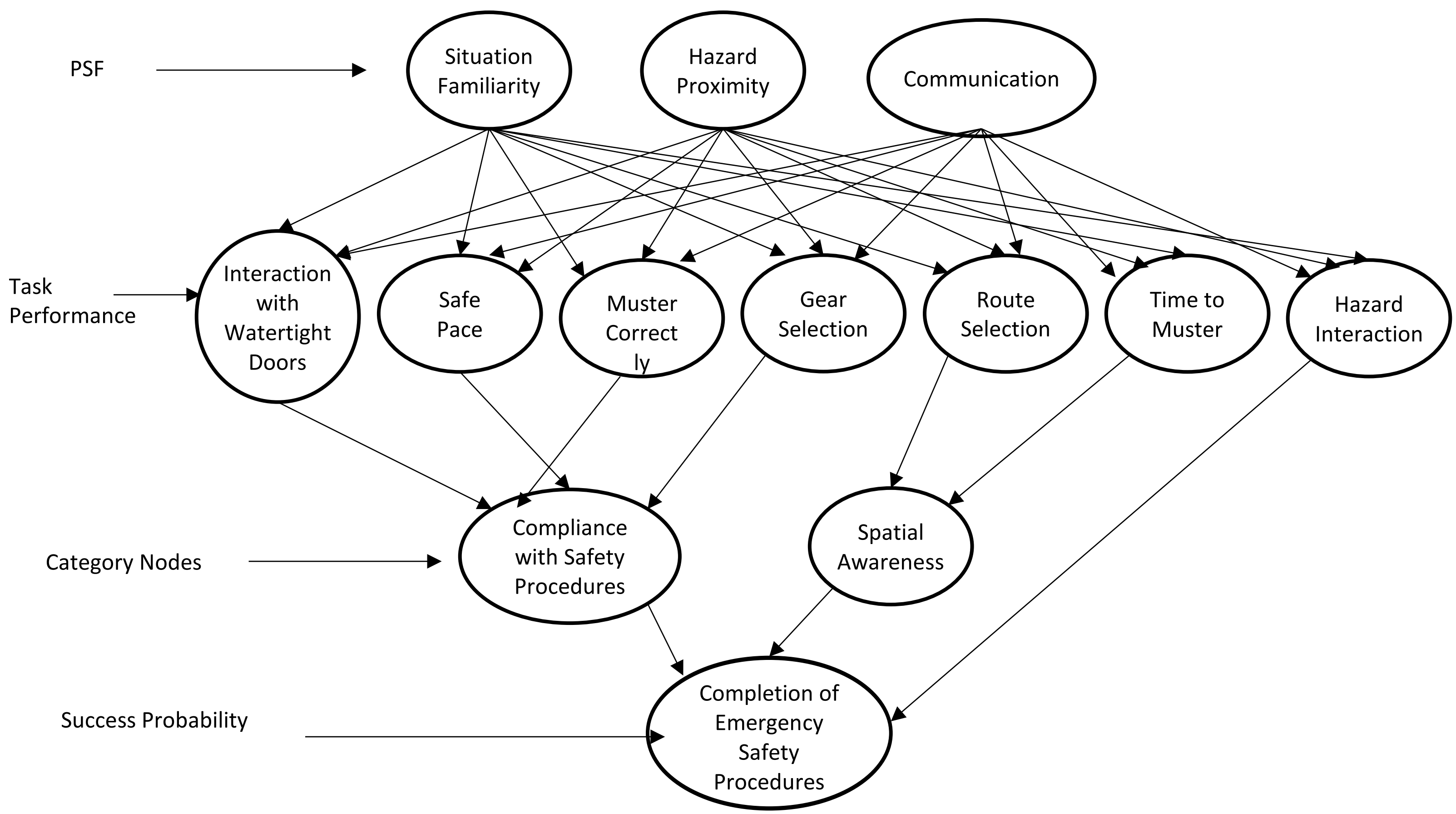

Figure 8: Proposed BN Model

from AVERT during participant performance (Moyle et al., 2017). Detailed information on AVERT data collection is in Section 5.

CPTs of category nodes and the final node are defined by the analyst. These definitions are based on the importance of different tasks to the specified categories. They are context sensitive and may need refinement before being used in a different application (Moyle et al., 2017). Refer to Section 6.2 for further information.

3.6 AVERT Scenario Development

A two-state, three factor experiment was created in AVERT to collect data to investigate the influence of PSFs on task performance. Eight different scenarios were developed that incorporated the high and low level of each PSF and the muster task performance list (Montgomery, 2008). Table 2 presents the list of scenarios and shows how PSFs were defined for each scenario.

The PSF level and the muster performance task list formed the basis for scenario creation. Storyboards on the desired outcome for each scenario were created. Each aspect of the scenario was scrutinized to ensure there were no confounding effects on the factors to be assessed. Once details were finalized, the scenarios were implemented in AVERT for use during the experiment.

Table 2: Scenarios for AVERT

	Communication		Hazard Proximity		Situation Familiarity	
1	+	PA announcement clear, concise, and all relevant information	+	Hazard location in area of no participant access	+	Scenario starts in cabin
2	-	PA announcement lacking relevant information.	+	Hazard location in area of no participant access	+	Scenario starts in cabin
3	+	PA announcement clear, concise, and all relevant information	-	Hazard may be seen in route selection	+	Scenario starts in cabin
4	+	PA announcement clear, concise, and all relevant information	+	Hazard location in area of no participant access	-	Scenario starts in Bridge
5	-	PA announcement lacking relevant information.	-	Hazard may be seen in route selection	+	Scenario starts in cabin
6	-	PA announcement lacking relevant information.	+	Hazard location in area of no participant access	-	Scenario starts in Bridge
7	+	PA announcement clear, concise, and all relevant information	-	Hazard may be seen in route selection	-	Scenario starts in Bridge
8	-	PA announcement lacking relevant information.	-	Hazard may be seen in route selection	-	Scenario starts in Bridge

Note: "+" denotes high level, positive influence; "-" denotes low level, negative influence.

Chapter 4 Experiment Procedure

4.1 Participant Criteria

To be eligible for the experiment, participants were required to meet a set of criteria, including no experience working offshore. The list of criteria is presented in Table 3.

Table 3: Participant Criteria

Question	Eligible Participant Answer
Prior Experience:	
1. Have you completed the Mastery of Learning Training?	**Yes.** Participants must have already completed AVERT Mastery of Learning Experiment.
2. Have you received experience working offshore since the first AVERT study?	**No.** Participants must not have any prior training or experience working offshore.
3. Do you expect to receive training to work offshore in the next 3 months?	**No.** Participants must not be expecting to receive training elsewhere during the course of the experiment.
Background Information:	
1. Are you between the ages of 18 and 65?	**Yes**
2. Do you have normal vision or corrected to normal vision (e.g. wear glasses or contacts)?	**Yes.** You must have normal or corrected to normal vision to be able to participate in this study.
3. Do you have a history of headaches or migraines?	**No.** Participants who have a history of headaches or migraines are not eligible to participate in this study.
4. Do you have a history of seizures or are you prone to seizures?	**No.** Participants who have a history of seizures or are prone to seizures are not eligible to participate in this study.
5. Are you susceptible to motion or simulator sickness?	**No.** The VE may cause symptoms of simulator sickness. Participants who have a high susceptibility to motion or simulator sickness will not be able to participate in the study.

The HRA research experiment was conducted with 37 participants. As described in Section 3.2, participants were already trained to competence in basic offshore emergency procedures using the safety training resources in AVERT.

4.1.1 Ethics

An ethics application was submitted and approved in September 2016 by the Interdisciplinary Committee on Ethics in Human Research (ICEHR) at Memorial University of Newfoundland. An ICEHR review is required under the federal government's Tri-Policy Council, which provides policy on the ethical conduct for research with human participants. Care and consideration of the participants' well-being were monitored and assessed throughout the experiment. Participants were required to sign a consent form highlighting their rights as a participant and stating all information about the experiment in which they participated.

4.2 Arriving at the lab

Participants who contacted the researcher were sent an introductory email explaining the purpose of the experiment. They were also notified of participant criteria (Table 3). If this information was satisfactory to the participant, they were given a scheduled time to arrive at the laboratory where the experiments were to take place.

Upon arrival at the laboratory, participants were introduced to the surroundings, and the workstation and building safety procedures were explained. As part of the required ethics application, participants were required to review and sign a consent form. A copy of the consent form is presented in Moyle and Veitch (2017). The researcher reviewed the consent form with the participant to ensure all parties understood the research and the experiment goals.

4.3 Physiological Data

Another research project was being conducted in conjunction with the HRA study. The theory, methodology, and results of that study are out with the scope of this thesis. However, the experimental set up is an important aspect of the HRA experimental procedure.

Participants were asked if physiological data resulting from their stress response during the HRA experiment could be collected. Three physiological measurements were taken to provide indicators of the participant’s virtual experience. The data is used to the level of stress experienced during the test scenarios. Table 4 outlines the physiological measures and assessment.

Before starting the HRA experiment, sensors were applied to locations on the torso and hand. A five-minute seated baseline of physiological signals was collected prior to the start of each scenario in the HRA experiment.

The details of the physiological related research are out of the scope of this thesis. It is mentioned to fully disclose what each participant was exposed to during the study. Results are not presented in this thesis.

Table 4 Physiological Measurements

Physiological Signal	**Assessment**
Heart Rate (EKG)	A 2-lead EKG will monitor heart activity. An increase in heart rate is an indicator of stress.
Galvanic Skin Response (GSR)	GSR measures sweat gland activity by application of electrodes to two fingers (bi-polar arrangement), or on the palm and forearm (unipolar arrangement). An increase is skin conductance activity is indicative of stress.
Respiration	Respiration will be recorded using a strap placed over the ribcage. In conjunction with heart rate variability, an increase in respiration variability is another indicator of stress.

4.4 Experiment Briefing

After the consent form was signed, the participant was briefed on the HRA experiment and what would be involved in the testing. The research purpose, objective, and goals were relayed to the participant. The experimental setup and process were explained, and the participant's role and expectations were outlined.

Participants were encouraged to envision the session as if it was the first day on an offshore platform and they were competently trained in offshore emergency procedures. The participants were asked to apply knowledge and skills learned during training to complete the emergency scenarios in the HRA experiment. To avoid the effect of additional learning, scenarios were not repeated and feedback on performance during each of the HRA test scenarios was not provided. Participants were also notified that each scenario would be completed in a randomly selected order.

Once all relevant information about the study had been relayed to the participant, testing and data collection could begin.

Chapter 5 Data Collection

During the study briefing, participants were advised that data for the HRA study would be collected in the form of performance metrics, physiological signals, and subjective assessments. Performance metrics are the aspects of participants' performance automatically captured in AVERT, and subjective assessments are answers from questionnaires given to the participant throughout the study. Participants were made aware that this data would be used to further evaluate their abilities in completing offshore safety procedures and potentially make changes to training instruction or platform emergency response procedures and processes.

5.1 Objective Performance Data

5.1.1 Data captured in AVERT

AVERT produces output files that capture the performance of a participant completing a scenario. Data such as scenario time, number of doors opened and closed, and whether a hazard was entered, are examples of such data. AVERT creates an individual report for each scenario. Using a Python script code, each scenario report is combined to a single Excel summary file for analysis. This summary report file is the main data file that is relied upon for participant analysis.

AVERT has the option of replaying scenarios. This is beneficial as it allows the observer to perform a quality assurance check on the AVERT report files and data that are collected by hand (Section 5.1.2). This feature is critical to confirming aspects of participant performance that were not clear from the electronic report files.

5.1.2 Data recorded by researchers

There were limitations to what AVERT recorded. At the time the experiment was completed, AVERT could not track the exact movement of a participant. AVERT can only create an output when a participant crosses a set check point. The actions of a participant in between checkpoints was not recorded. An extra layer of data collection was required to ensure no lost data points. To address these two issues, route maps and movement logs were completed by the researcher. The route maps and movement logs were used to verify participant movements and confirm the outputs of the AVERT report files.

5.2 Subjective data

In addition to the electronic files from AVERT and researcher observed data, participants were asked to complete questionnaires assessing their experience in the virtual environment. The questionnaires that were completed are the simulator sickness questionnaire (SSQ), the post-trial scenario questionnaire, and emotion questionnaire. The emotion questionnaire is related to the physiological study mentioned in Section 4.3 and is not included in this thesis.

The post-trial scenario questionnaire was completed after each scenario was completed. This asked questions about how realistic the simulation was to the participant and whether the experience met the participants' expectations.

The SSQs allow the researcher to monitor the participant's physical conditions. Navigation through the virtual space using a desktop computer configuration may cause some individuals to experience symptoms of visually induced motion sickness (VIMS) or simulator-induced sickness (SIS) (Kennedy et al., 1993). The symptoms of simulator-induced sickness include fatigue, headache, eye strain, difficulty focusing, increased salivation, sweating, nausea, stomach awareness, blurred vision, dizziness, vertigo, and burping. The symptoms of simulator sickness can sometimes occur during, immediately after, or several hours after exposure to the simulator. To ensure a participant did not experience severe symptoms, simulator-induced sickness susceptibility was assessed prior to the study and was monitored throughout the study using the SSQ (Kennedy et al., 1993). The questionnaire allows symptoms to be rated as not present, minimal, moderate, and severe. A SSQ was completed prior to the start of the study, after the first four scenarios, and again at the end of the scenario testing. If any of the symptoms were reported as moderate or severe, the decision about whether the participant should continue with the study or not was made. The post-trial scenario questionnaire and SSQ that was used in the study and results are available in Moyle and Veitch (2017).

5.3 Transfer and Storage of Data

The ICHER policy requires that the identity and personal information of participants are protected. Study policy and procedure was established to protect participant identity and personal information from unauthorized use. This was communicated to the participant in the consent form.

Each participant was given as assigned alphanumeric number. The list that connects the assigned number to personal information is stored on a password protected computer in a locked office, providing two barriers. The participant list, consent form, and recorded data are not stored together.

AVERT report files are stored on a computer dedicated to the study and transferred to an alternate computer for analysis. Hand drawn maps, route logs and participant questionnaires are stored in a locked cabinet, within a locked office space. There is no connection between the participant personal information to the recorded data other than the participant list. This is in line with the requirements under the ICHER ethics application.

Participants were informed that the results and conclusions from the study may be published in peer reviewed journals/conferences and that a formal report will be made available upon request.

5.4 Assessment Concerns

After the start of the experiment, it was noted that five participants had observed a “60 Second” warning. This warning was a part of the AVERT program when the training program was developed, and was not removed during the scenario development for this study. Time stress was not a factor that was studied in this experiment, and there was concern that observing this warning could induce time stress in participants and therefore affect results.

Amongst the five participants, there were twelve occurrences of seeing the “60 Second” warning. The occurrences were assessed and evaluated to determine if the warning had any effect on participants’ performance. Nine occurrences were while the participants were waiting for the scenario to end after all required tasks were completed, or very close to completing all tasks. Participants completed the scenario successfully in these occurrences and it was reckoned that the warning was unlikely to have contributed to the participant’s stress level. Three occurrences (twice with the same participant) happened while the participant was not close to finishing the scenario. The participants appeared confused or unsure of required actions before visually seeing the warning and it was reckoned that the warning was unlikely to have increased any stress already experienced by the participant. Once the assessment was complete, the “60 Second” time was removed from the scenarios so no further participants could observe it.

Chapter 6 Analysis & Results

6.1 Analysis of Performance Data

A rubric was developed to assess the performance results from AVERT and was populated with each participant's performance data. Seven tasks were assessed, as listed in Table 5. Each of the tasks represents a node in the BN shown in Section 3.5. Each task is further defined by individual scenarios. This was done as each scenario is unique and elements had to be specific to accurately assess the performance. The populated rubric was saved separately from the AVERT files to offer a quick snapshot into participants' performance. A copy of the rubric and each participant's performance results are available in Moyle and Veitch (2017).

Table 5: Task Performance Definitions and Locations

Task Performance	Definition	Data Source
Route Selection	Participant selects most efficient route enabling them to muster as quickly as possible	Route Map, Movement Log
Muster Correctly	Participant musters as per alarm type. Procedure to follow is defined per scenario.	AVERT
Gear Selection	Participant selects gear as per location start. Procedure to follow is defined per scenario.	AVERT
Time to Muster	Participant successfully musters and responds to correct alarm fully within 5 minutes	AVERT
Hazard Interaction	Participant does not interact with hazard	AVERT, Route Map, Movement Log
Interaction with Watertight Doors	Participant closes all safety doors they go through	AVERT
Safe Pace	Participant walks throughout scenarios	AVERT

The AVERT report files were assessed and the applicable performance metrics were extracted. As AVERT did not automatically capture all relevant performance metrics that were required for the analysis, route maps and movement logs were used to supplement the electronic files. Table 5 also shows the data source for each task.

From this table, one can see that participant A61 closed all the doors in scenarios 1, 2, 3, 4, 5, 6 and 8 (as indicated by the shaded boxes). In scenario 7, the participant left one or more doors open (as indicated by the shaded box).

6.2 BN Analysis

GeNIe Modeller (Anonymous, 2017) was used for the quantitative analysis presented in this thesis. GeNiE incorporates the equations presented in Section 2.3 to calculate the probability of task performance, specified categories, and finally the successful completion of the required safety procedures during emergency situations.

6.2.1 Performance Shaping Factors Nodes

For this study, prior probabilities of different states of each PSF were assumed to be 50%. This means that in the experimental setup, for each participant, four out of the eight scenarios had a high level of each PSF and four had a low level of each PSF. Each participant was exposed to a combination of high and low levels of each PSF during each scenario as shown in Table 2. The probability of exposure to a level of PSF during a scenario was 50%.

6.2.2 Task Performance Nodes

The CPT for the task performance values is populated from the experiment data that were collected using eight different scenarios. Table 7 shows the CPT at the node “Interaction with watertight doors” for Participant A61. As shown in the table, probabilities of closing fire doors in eight different conditions are needed in the CPT. The information in the rubric of Table 6 is transferred to the CPT. In this example, the participant closed all fire doors in seven scenarios, and did not close all fire doors in one scenario. As the probability of the event occurring is binary (i.e. the event fully happened, or it did not), the values 1 and 0 are used.

Table 6: CPT data collected for "Interaction with watertight doors"

Communication	*High*				*Low*			
Hazard Proximity	*High*		*Low*		*High*		*Low*	
Situation Familiarity	*High*	*Low*	*High*	*Low*	*High*	*Low*	*High*	*Low*
Close All Doors	1	1	1	0	1	1	1	1
Did Not Close All Doors	0	0	0	1	0	0	0	0

Table 7 shows the CPT of the node “Interaction with watertight doors”. CPTs of the other task performance nodes were populated in the same way from the participant’s performance data. Full Performance data and rubric for Participant A61 are available in Moyle and Veitch (2017).

After integrating all experimental data, the marginal probability of each task element occurring was calculated by the GeNie Modeller program.

6.2.3 Category Nodes

CPTs of category nodes and the final node are defined by the analyst. These definitions are based on the importance of different tasks to the specified categories. They are context sensitive and may need refinement before being used in a different application.

For this study, the CPT was populated based on the analyst's belief of what errors contributed to the failure of each category. For example, for "Spatial Awareness" category, the node values were "0" for any case that the task node was not in compliance. Table 8 shows the CPT table for the category node "Spatial Awareness". In this CPT, the node values are "Yes" or "No" indicating compliance was achieved or not. As shown in the table, when "Spatial Awareness" is "Yes", the node values are 1. This relates to when a participant selected the most efficient route or a deviated route, indicating the participant reached the muster location, and completed the muster sequence in an acceptable time defined by the analyst. For the category "Compliance with Safety Procedures", the values of the node are also "Yes" or "No". For this CPT, the "Yes" element reflects that all task performance nodes connected to the category are "Yes", or in the positive status, which is the same for the CPT for the final node "Completion of Emergency Safety Procedures". All category and final node CPTs are the same for all participants.

Table 7: CPT for Category Node: “Spatial Awareness"

Time to Muster	*Less than 5 Min*			*More than 5 Min*		
Select Efficient Route	*Most Efficient*	*Deviated Route*	*Does not Muster*	*Most Efficient*	*Deviated Route*	*Does not Muster*
Spatial Awareness – Yes	1	1	0	0	0	0
Spatial Awareness -No	0	0	1	1	1	1

6.3 Success Probability Analysis

Once data was input into each CPT and the prior PSF values were given, the probability of successful completion of emergency procedures was calculated for each participant. This value indicates the strength of a participant’s knowledge of safety procedures and the ability to act appropriately in an emergency situation. This is valuable to the analysis as it is a prediction of how well a participant is likely to perform during an emergency situation.

6.3.1 Success Probability - Individual Participants

Performance results from AVERT can help determine how a participant is likely to perform in a real-life emergency situation (Musharraf et al., 2016). By using the performance results as data input into a BN, a numerical value of probability of success of an individual in a variety of offshore emergency situations can be calculated. To illustrate, Figure 9 shows the completed BN that quantifies the egress success probability for participant A61. Participant A61 is shown to have an 87.5% chance of successfully completing an emergency situation.

To demonstrate, an example calculation is presented below using the equations in section 2.3. Performance of participant A61 is used here as the example. Prior probabilities of the status of the PSFs are known to be 50%. The CPTs for the relationships between PSFs and each task node are populated from the participant's performance data in AVERT. CPTs along with the prior probabilities are used to calculate the marginal probability of the task nodes. For example, using the CPT table (Table 7) the marginal probability of the node "Interaction with Watertight doors" can be calculated as below. The other marginal probabilities for each task node are shown in Figure 9.

$$P\ (Interaction\ with\ Watertight\ doors\ =\ close\ all)$$

$$= P\ (Interaction\ with\ Watertight\ doors\ =\ close\ all|\ Com. = High, HP = High, SF = High) \times P(Com. = High) \times P(HP = High) \times P(SF = High)$$

$$+ P\ (Interaction\ with\ Watertight\ doors\ =\ close\ all|\ Com. = High, HP = High, SF = Low) \times P(Com. = High) \times P(HP = High) \times P(SF = Low)$$

$$+\ P\ (Interaction\ with\ Watertight\ doors\ =\ close\ all|\ Com. = High, HP = Low, SF = High) \times P(Com. = High) \times P(HP = Low)\ \times P(SF = High)$$

$$+ P\ (Interaction\ with\ Watertight\ doors\ =\ close\ all|\ Com. = High, HP = Low, SF = Low) \times P(Com. = High) \times P(HP = Low) \times P(SF = Low)$$

$$+ P\ (Interaction\ with\ Watertight\ doors\ =\ close\ all|\ Com. = Low, HP = High,$$
$$SF = High) \times P(Com. = Low) \times P(HP = High) \times P(SF = High)$$

$$+ P\ (Interaction\ with\ Watertight\ doors\ =\ close\ all|\ Com. = Low, HP =$$
$$High, SF = Low) \times P(Com. = Low) \times P(HP = High) \times P(SF = Low)$$

$$+ P\ (Interaction\ with\ Watertight\ doors\ =\ close\ all|\ Com. = Low, HP =$$
$$Low, SF = High) \times P(Com. = Low) \times P(HP = Low) \times P(SF = High)$$

$$+ P\ (Interaction\ with\ Watertight\ doors\ =\ close\ all|\ Com. = Low, HP =$$
$$Low, SF = Low) \times P(Com. = Low) \times P(HP = Low) \times P(SF = Low)$$

$$= (1 \times 0.5 \times 0.5 \times 0.5) + (1 \times 0.5 \times 0.5 \times 0.5) + (1 \times 0.5 \times 0.5 \times 0.5) +$$
$$(0 \times 0.5 \times 0.5 \times 0.5) + (1 \times 0.5 \times 0.5 \times 0.5) + (1 \times 0.5 \times 0.5 \times 0.5) +$$
$$(1 \times 0.5 \times 0.5 \times 0.5) + (1 \times 0.5 \times 0.5 \times 0.5)$$

$$= 0.875$$

Therefore,

$$P\ (Interaction\ with\ Watertight\ doors\ = did\ not\ close\ all) = 1 - 0.875 = 0.125$$

Once the task nodes are populated, the probability of the category nodes are calculated. Conditional probability tables are required to complete the calculation.

The “Spatial Awareness” node has input values from the “Route Selection” and “Time to Muster” nodes. To calculate the success probability of the “Spatial Awareness” node, the following equation is calculated using the CPT information from Table 8:

$$P(Spatial\ Awareness = Yes)$$

$$= P(Spatial\ Awareness = Yes | Route\ Sel. = correct, Muster\ time = less\ than\ 5\ min) \times P(Route\ Sel. = correct) \times P(Muster\ time = less\ than\ 5\ min)$$

$$+ P(Spatial\ Awareness = Yes | Route\ Sel. = correct, Muster\ time = more\ than\ 5\ min) \times P(Route\ Sel. = correct) \times P(Muster\ time = more\ than\ 5\ min)$$

$$+ P(Spatial\ Awareness = Yes | Route\ Sel. = deviated, Muster\ time = less\ than\ 5\ min) \times P(Route\ Sel. = deviated) \times P(Muster\ time = less\ than\ 5\ min)$$

$$+ P(Spatial\ Awareness = Yes | Route\ Sel. = deviated, Muster\ time = more\ than\ 5\ min) \times P(Route\ Sel. = deviated) \times P(Muster\ time = more\ than\ 5\ min)$$

$$+ P(Spatial\ Awareness = Yes | Route\ Sel. = incorrect, Muster\ time = less\ than\ 5\ min) \times P(Route\ Sel. = incorrect) \times P(Muster\ time = less\ than\ 5\ min)$$

$$+ P(Spatial\ Awareness\ =\ Yes|Route\ Sel.\ =\ incorrect, Muster\ time\ = more\ than\ 5\ min)\ \times\ P(Route\ Sel.\ =\ incorrect)\ \times\ P(Muster\ time\ = more\ than\ 5\ min)$$

$$= (1\ \times\ 0.625\ \ \times\ 1.00) + (0\ \times\ 0.625\ \text{x}\ 0.00) +\ (1\ \times\ 0.375\ \times\ 1.00) + (0\ \times\ 0.375\ \times\ 0.00) +\ (0\ \times\ 0.00\ \times\ 1.00) +\ (0 \times\ 0.00\ \times\ 0.00)$$

$$= 1.00$$

Therefore,

$$P(Spatial\ Awareness = No) = 1 - 1.00 = 0.00$$

The results for the other category node "Compliance with Safety Procedures" is shown in Figure 9.

After the category nodes are calculated, probability of "Completion of Emergency Safety Procedures = Yes" is calculated. This node has inputs from the "Compliance with Safety Procedures", "Spatial Awareness" and "Hazard Interaction" nodes. The equation used for calculation is:

$$P(Success\ =\ Yes)$$

$$= P(Success\ =\ Yes\,\big|\,Hazard = does\ not\ interact, Comp. with\ Safety = Yes, Spatial\ Aware. = Yes) \times\ P(Hazard = does\ not\ interact)\ \ \times\ P(Comp. with\ Safety\ =\ Yes) \times\ P(Spatial\ Aware. =\ Yes)$$

$$+ P(Success = Yes \mid Hazard = does\ not\ interact, Comp.with\ Safety = Yes, Spatial\ Aware. = No) \times P(Hazard = does\ not\ interact) \times P(Comp.with\ Safety = Yes) \times P(Spatial\ Aware. = No)$$

$$+ P(Success = Yes|Hazard = does\ not\ interact, Comp.with\ Safety = No, Spatial\ Aware. = Yes \times P(Hazard = does\ not\ interact) \times P(Comp.with\ Safety = No) \times P(Spatial\ Aware. = Yes)$$

$$+ P(Success = Yes|Hazard = does\ not\ interact, Comp.with\ Safety = No, Spatial\ Aware. = No) \times P(Hazard = does\ not\ interact) \times P(Comp.with\ Safety = No) \times P(Spatial\ Aware. = No)$$

$$+ P(Success = Yes|Hazard = See - No\ interact, Comp.with\ Safety = Yes, Spatial\ Aware. = Yes \times P(Hazard = See - No\ interact) \times P(Comp.with\ Safety = Yes) \times P(Spatial\ Aware. = Yes)$$

$$+ P(Success = Yes \mid Hazard = See - No\ interact, Comp.with\ Safety = Yes, Spatial\ Aware. = No) \times P(Hazard = See - No\ interact) \times P(Comp.with\ Safety = Yes) \times P(Spatial\ Aware. = No)$$

$$+ P(Success = Yes \mid Hazard = See - No\ interact, Comp.with\ Safety = No, Spatial\ Aware. = Yes) \times P(Hazard = See - No\ interact) \times P(Comp.with\ Safety = No) \times P(Spatial\ Aware. = Yes)$$

$$+\ P(Success\ =\ Yes\,\big|\,Hazard\ =\ See - No\ interact), Comp.with\ Safety\ =$$

$$No, Spatial\ Aware. =\ No\ \times\ P(Hazard\ =$$

$$See - No\ interact)\ \times\ P(Comp.with\ Safety\ =\ No) \times\ P(Spatial\ Aware. =\ No)$$

$$+\ P(Success\ =\ Yes\,\big|\,Hazard\ =\ See - Interacts, Comp.with\ Safety\ =$$

$$Yes, Spatial\ Aware. =\ Yes) \times\ P(Hazard\ =$$

$$See - Interacts)\ \times\ P(Comp.with\ Safety\ =\ Yes) \times\ P(Spatial\ Aware. =\ Yes)$$

$$+\ P(Success\ =\ Yes|Hazard\ =\ See - Interacts, Comp.with\ Safety\ =$$

$$Yes, Spatial\ Aware. =\ No)\ \times\ P(Hazard\ =$$

$$See - Interacts)\ \times\ P(Comp.with\ Safety\ =\ Yes)\ \times\ P(Spatial\ Aware. =\ No)$$

$$+\ P(Success\ =\ Yes\,\big|\,Hazard\ =\ See - Interacts, Comp.with\ Safety\ =$$

$$No, Spatial\ Aware. =\ Yes) \times\ P(Hazard\ =$$

$$See - Interacts)\ \times\ P(Comp.with\ Safety\ =\ No) \times\ P(Spatial\ Aware. =\ Yes)$$

$$+\ P(Success\ =\ Yes\,\big|\,Hazard\ =\ See - Interacts, Comp.with\ Safety\ =$$

$$No, Spatial\ Aware. =\ No) \times\ P(Hazard\ =$$

$$See - Interacts)\ \times\ P(Comp.with\ Safety\ =\ No) \times\ P(Spatial\ Aware. =\ No)$$

$$=\ (1\ \times 0.75 \times 0.875 \times\ 1.00) +\ (0\ \times\ 0.75\ \times\ 0.875\ \times\ 0) +\ (0\ \times\ 0.75\ \times\ .125\ \times$$

$$1.00) +\ (0\ \times\ 0.75\ \times\ 0.875\ \times\ 0) + (1\ \times\ 0.25 \times 0.875 \times\ 1.00) +\ (0\ \times\ 0.25\ \times$$

$$0.875 \times\ 0) +\ (0\ \times\ 0.25\ \times\ 0.125 \times\ 1.00) +\ (0\ \times\ 0.25 \times\ 0.875\ \times\ 0) +\ (0\ \times$$

$$0 \times 0.875 \times 1.00) + (0 \times 0 \times 0.875 \times 0) + (0 \times 0 \times 0.125 \times 1.00) + (0 \times 0 \times 0.875 \times 0)$$

$$= 0.875$$

Therefore,

$$P(Success = No) = 1 - 0.875 = .125$$

Therefore, as stated above, Participant A61 has a 87.5% chance of successfully completing an emergency situation. In comparison, the results for Participant A4 are shown in Figure 10. As shown in Figure 10, this participant has a 37.5% chance of successfully completing an emergency situation. This probability is calculated in the same way as Participant A61 as described above.

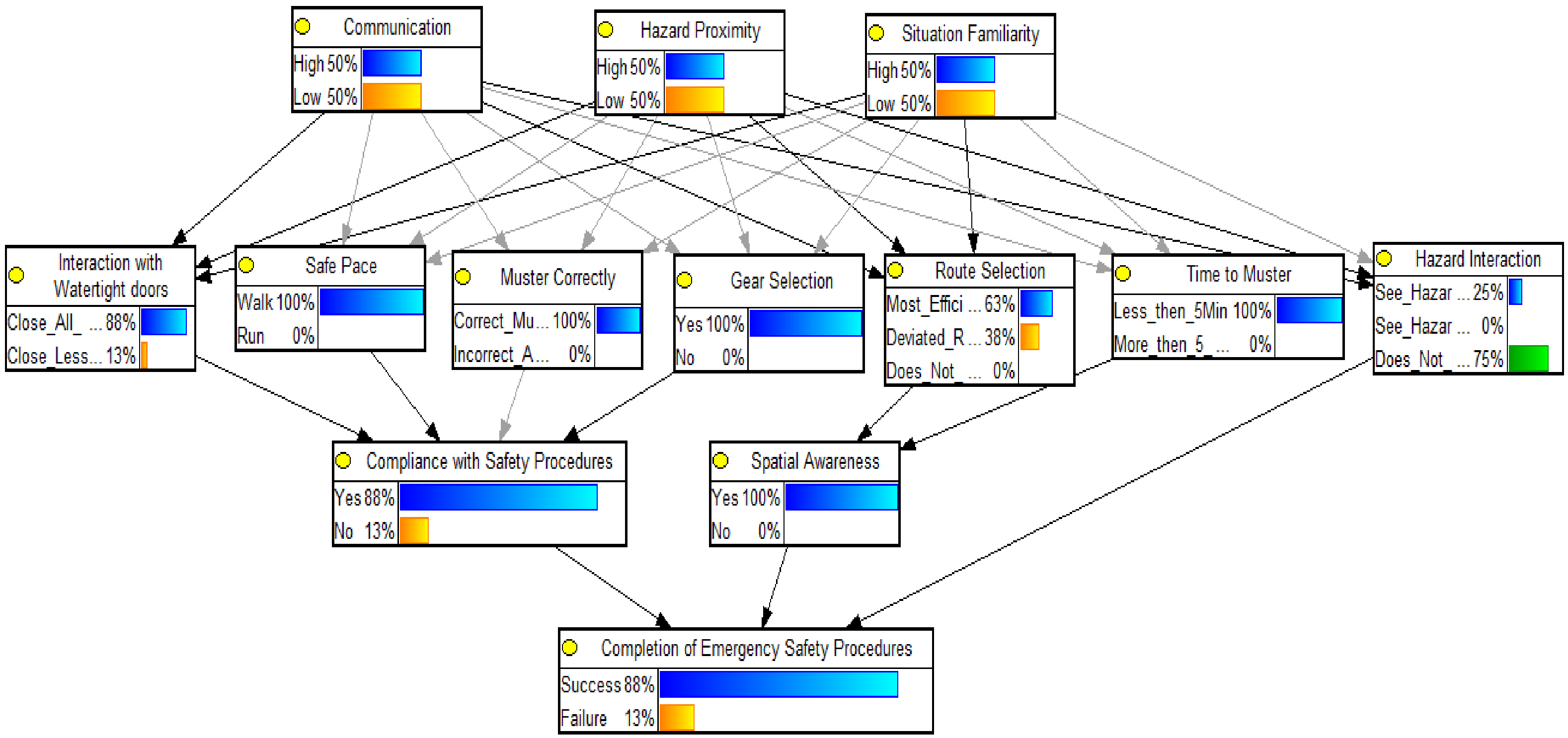

Figure 9: BN Results for Participant A61

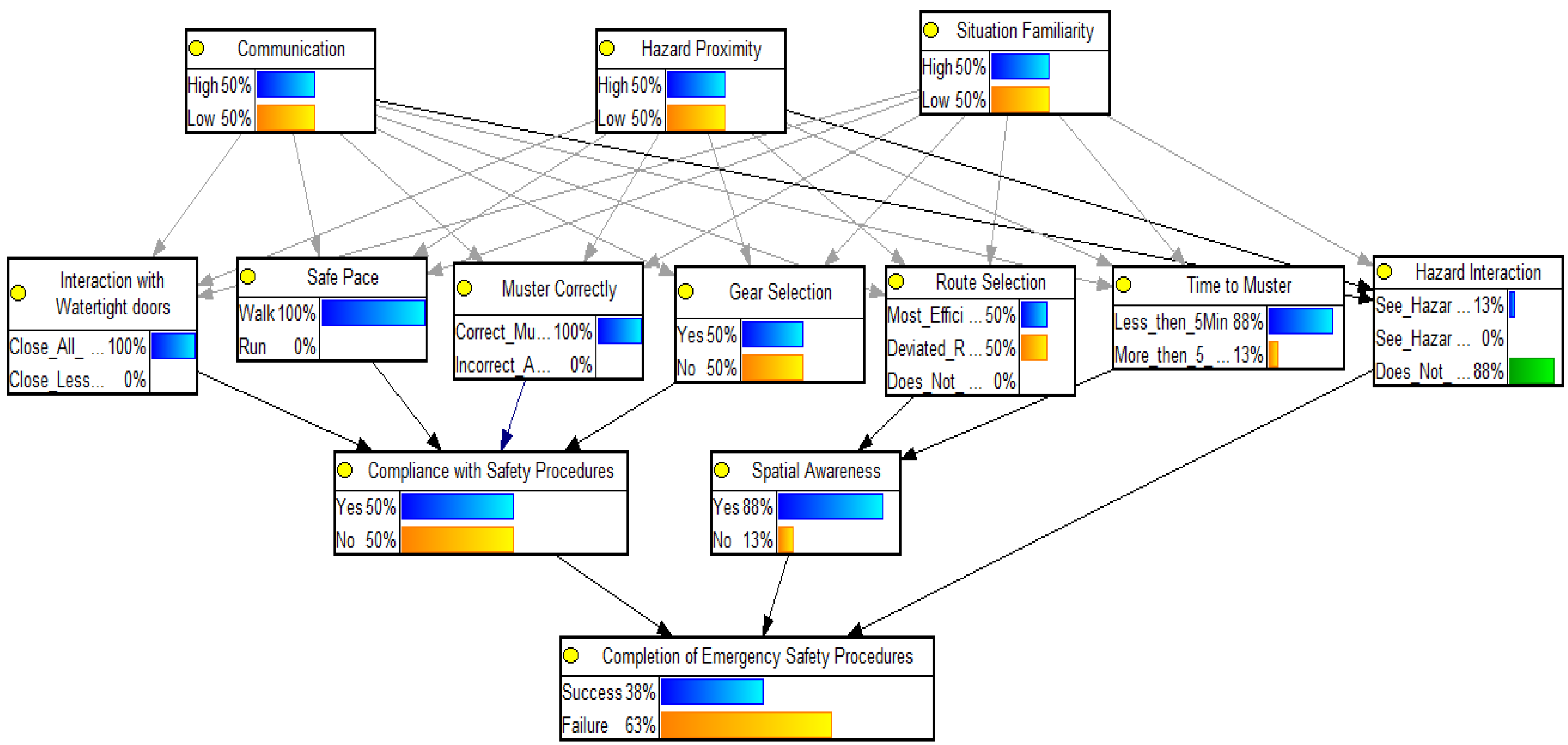

Figure 10: BN Results for Participant A4

6.3.2 Success Probability –Cohort Summary

The probability of successful completion of emergency safety procedures was calcuated for 37 participants using the method described in section 6.3.1. The results for the cohort are summarized in the histogram in Figure 11. All participants achieved some level of success. Twenty four participants had an 80% or higher probability of success. These participants retained the information provided in the training and could apply the knowledge and skills they acquired.

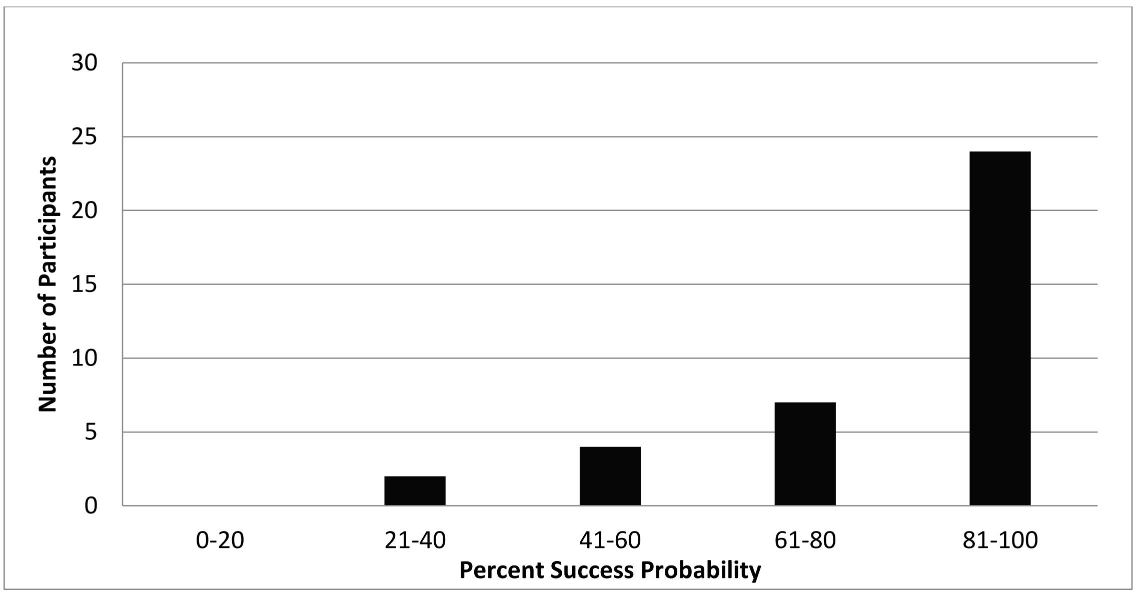

Figure 11: Overall Success Probability Results

6.4 Root Cause Analysis

A benefit of using a BN to model participant performance is the flexibility to do a root-cause analysis, which can be used to investigate the strength of influence of different PSFs on successful completion of emergency procedures (Musharraf et al., 2017).

Manipulation of the BN within GeNie Modeller can determine the effect each level of PSF has on the overall success probability.

6.4.1 Root Cause Analysis - Individual Participants

To demonstrate the benefit of the flexibility to manipulate a BN, a root-cause analysis was completed on all participants to investigate the strength of influence of different PSFs on success probabilities, and to identify the root causes of failure, given that the participant has failed.

First, the effect of 100% failure occurring is demonstrated. Given that a failure in completing safety procedures has happened, probabilities of PSFs for each participant will change accordingly. The change in PSFs show how each PSF contributes to failure. For example, Figure 12 demonstrates the updated BN for Participant A61 given the evidence that a failure has happened. To demonstrate failure occurring, the value of failure in the node “Completion of Emergency Safety Procedures” is set to 100%.

Table 9 shows the prior probability, posterior probability, and change in probability percentage point change of the PSFs when failure is 100%. Two PSFs, hazard proximity and situation familiarity, change to 100% (low). This indicates that when participant A61 is in a relatively unfamiliar area and sees a hazard during mustering, they have a high probability of not completing emergency procedures successfully. This knowledge is beneficial as it highlights that additional training might usefully focus on conditions defined by the low level of these PSFs. In contrast, the communication PSF changes to 100% (high). Despite having all relevant information to make an informed decision on

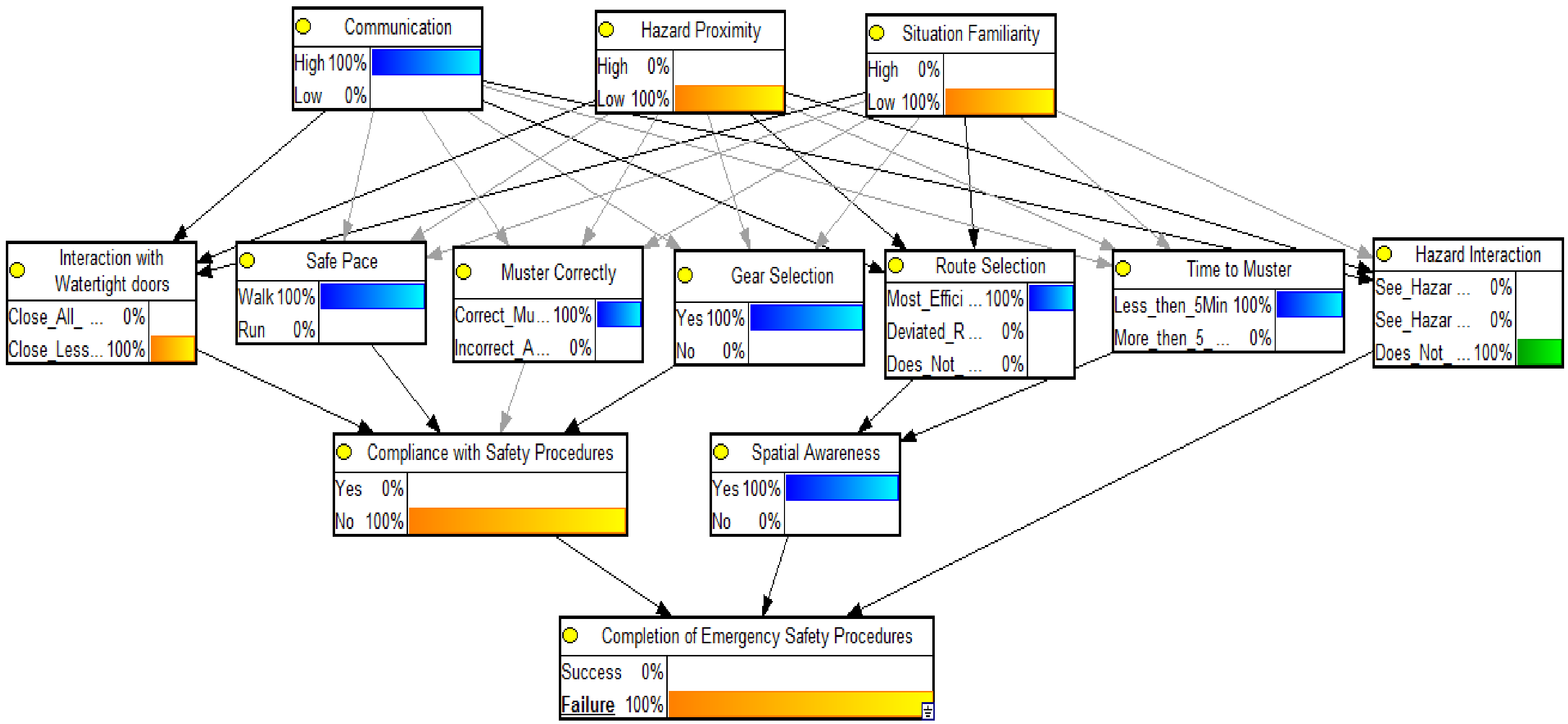

Figure 12: BN for Participant A61 after Evidence Update

how to muster, the participant still has a high probability of not completing emergency procedures successfully, which indicates that the other two PSFs, hazard proximity and situation familiarity, may strongly affect the participants' decision making more than the communication PSF.

Table 8 : Prior Probability and Posterior Probability for Participant A61

	Hazard Proximity (Low)	**Hazard Proximity (High)**	**Communication (Low)**	**Communication (High)**	**Situation Familiarity (Low)**	**Situation Familiarity (High)**
Prior Probability (%)	50	50	50	50	50	50
Posterior Probability (%)	100	0	0	100	100	0
Change in Probability (percent points)	+50	-50	-50	+50	+50	-50

In comparison, Figure 13 demonstrates the updated BN for Participant A4 given the evidence that a failure has happened.

Table 10 shows the prior probability, posterior probability, and percent change in probability of the PSFs when failure is 100%. The largest change is for situation familiarity. This indicates that when in a relatively unfamiliar area, Participant A4 has a high probability of not completing emergency procedures successfully. As before, this knowledge is beneficial to the sample participant as it highlights where additional training might be most usefully focused.

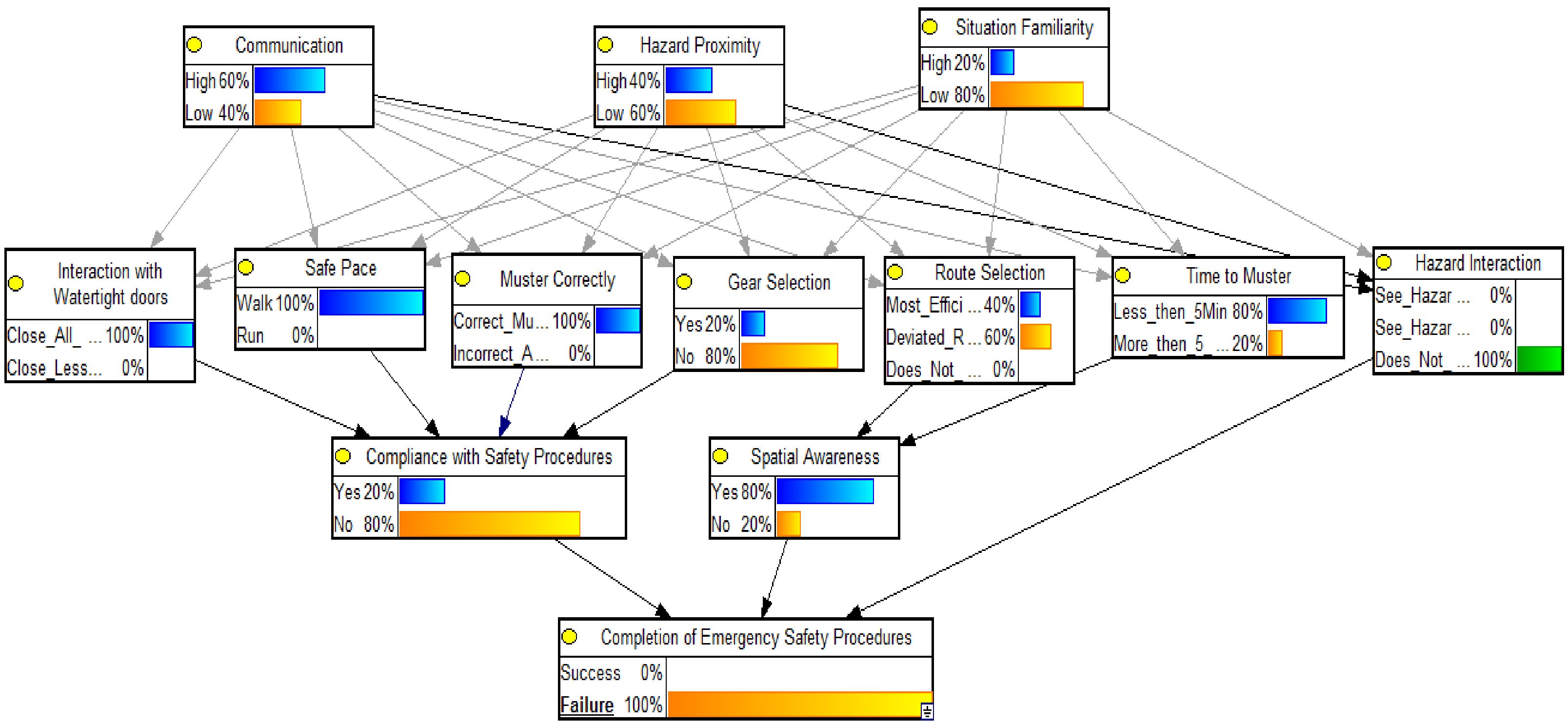

Figure 13: BN for Participant A4 after Evidence Update

Table 9: Prior Probability and Posterior Probability for Participant A4

	Hazard Proximity (Low)	**Hazard Proximity (High)**	**Communication (Low)**	**Communication (High)**	**Situation Familiarity (Low)**	**Situation Familiarity (High)**
Prior Probability (%)	50	50	50	50	50	50
Posterior Probability (%)	60	40	40	60	80	20
Change rate of Probability (%)	10	-10	-10	10	30	-30

Another application of the BN is that it can be used to assess the effect of each PSF on the success probability. To do so, each level of PSF is set to 100% one at a time and an assessment on how the success probability changes is made.

Figure 14 shows how each PSF affects Participant A61's probability of success. A tornado graph is used to demonstrate the change in success probability. From the graph, when the situation familiarity and hazard proximity PSFs have a high value of 100%, the success probability increases to 100% as well. When the communication PSF has a high value of 100%, the success probability decreases to 75%. This difference in value is due to the performance results in AVERT combined with manipulation of the PSF value. The results indicate Participant A61 had low performance results in scenarios when the communication PSF had a high value, and had high performance results in scenarios when the situation familiarity PSF and the hazard proximity PSF had a high value. The magnitude of the effect of each PSF on probability is the same, and it can be either positive or negative.

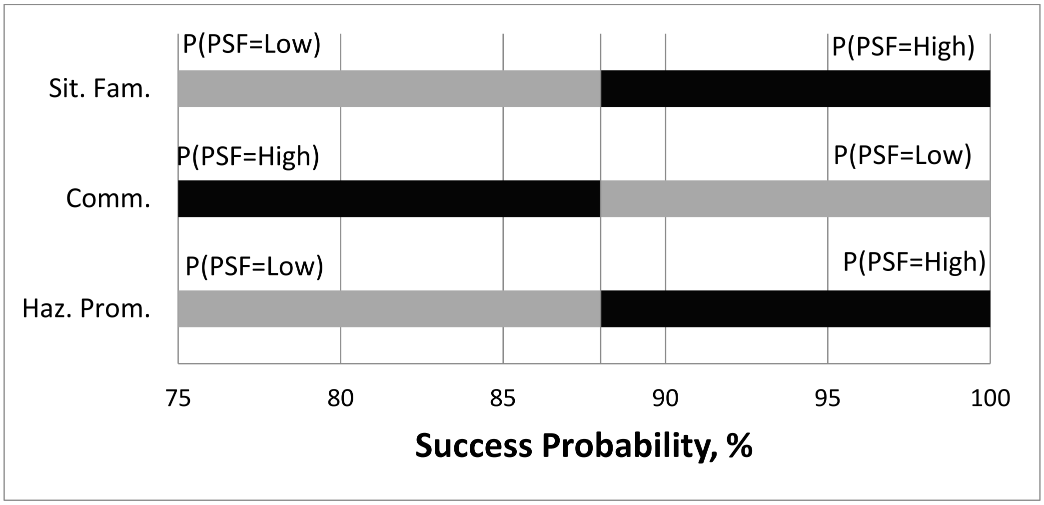

Figure 14: Tornado Graph for Participant A61

Figure 15 shows how each PSF affects participant A4's probability of success. From the graph, the situation familiarity PSF affects participant A4 the greatest. The range of success probability is from 0% to 75%.

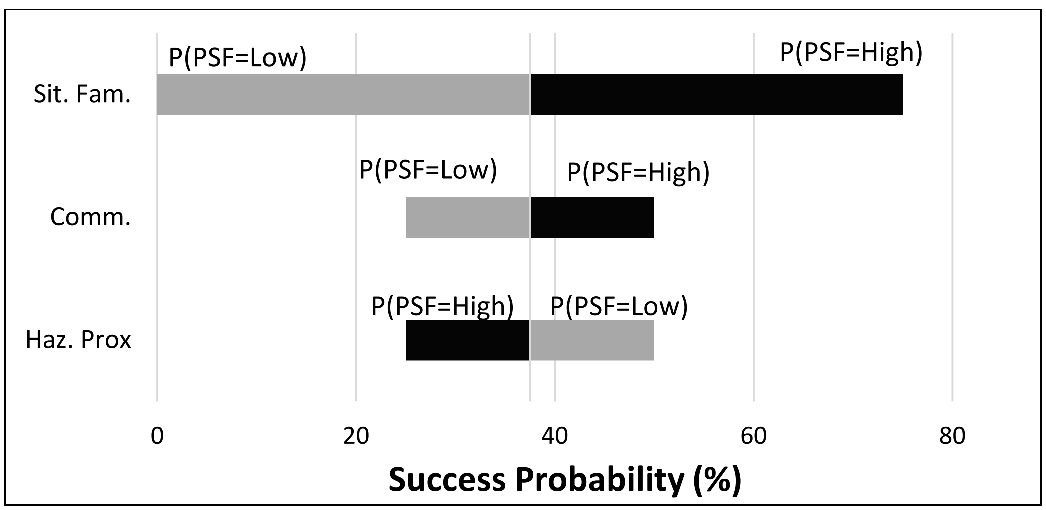

Figure 15: Tornado Graph for Participant A4

As illustrated by comparing Participant A61 to Participant A4, it is shown that individuals are affected by PSFs in different degrees. Therefore, it is important to assess skills and knowledge individually.

6.4.2 Root Cause Analysis - Overall Cohort

A root-cause assessment to determine the PSF that most affected the probability of success was done for all thirty-seven participants. Figure 16 presents a histogram that shows the results. As shown in the figure, sixteen participants were not adversely affected by any change in PSF. These participants retained the information provided in the training and could apply the knowledge and skills they acquired regardless of the change in PSFs. As the PSF probabilities changed, the overall success probability remained at 100%. Nine participants were most affected by situation familiarity. Additional training relating to situation familiarity may be useful for these participants. Eleven participants were equally affected by all PSFs. This indicates that regardless of the change in PSF, success probability changed by the same amount. These participants may need to undergo further training in all areas relating to the PSFs.

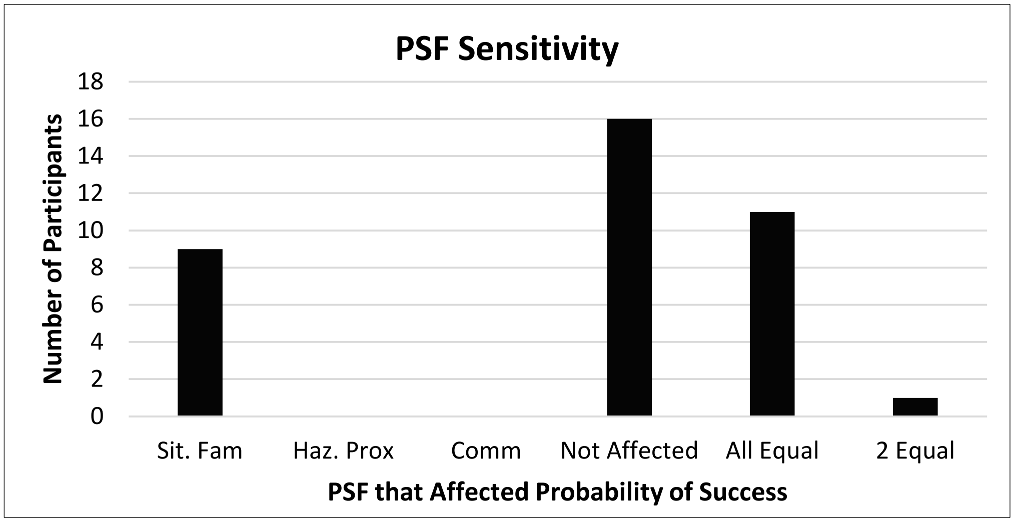

Figure 16: Root Cause Analysis for 37 Participants

6.5 Task Performance Errors

In addition to assessing the likelihood of successful completion of emergency procedures, how successful the participants were in completing specific tasks was assessed. Task performance results contribute to the overall success probability. By assessing the task performance results, we can identify specific areas of the initial training program, or physical layout of the platform, that can be improved.

Table 11 shows the performance tasks and the percentage of how successfully each task was completed across all scenarios and participants.

Table 10: Task Performance - Percentages

Task Performance	Requirement	Percentage of Successful Completion
Route Selection	Participant selects most efficient route enabling them to muster as quickly as possible	97%
Muster Correctly	Participant musters as per alarm type. Procedure to follow is defined per scenario.	96%
Gear Selection	Participant selects gear as per location start. Procedure to follow is defined per scenario.	89%
Time to Muster	Participant successfully musters and responds to correct alarm fully within 5 minutes	94%
Hazard Interaction	Participant does not interact with hazard	99%
Interaction with Watertight Doors	Participants closes all safety doors they go through	96%
Safe Pace	Participants walk throughout scenarios	99%

The task performance node with the lowest percentage of success was “Gear Selection”. There are four scenarios that started in a platform location that was unfamiliar to the participant. The process during an emergency situation is to muster, and retrieve safety gear at the muster location as needed. From evaluation of the data, it appears that some participants did not retain the information that there was additional safety gear present at the muster station. This resulted in these participants returning to their cabin to select safety gear, which is against safety procedures. Additional training information could be given to enhance this knowledge on an individual basis, or the overall training program could be adjusted for everyone. As well, additional physical reminders on the platform of safety gear locations could be installed.

The next performance task node that had a lower success percentage was “Time to Muster”. From a review of the data, it was noted that the majority of errors occurred when the situation familiarity PSF was in the low level. The scenarios that had a low level of situation familiarity PSF started in the control room. It is assumed participants were not confident when in this location on how to find the primary route to the muster station, or could not locate an area as requested in an acceptable time. This error is tied to the performance task node “Route Selection”. More often than not, participants selected the secondary route no matter what information was available from the other PSFs. This indicates that route designation of primary or secondary was not made clear in the training, so participants were not able to make a good decision on route selection based on priority. Initial training could benefit from further testing to increase spatial awareness of other locations in the platform.

Chapter 7 Conclusion

This thesis presents a methodology to investigate the effect of PSFs on offshore emergency egress and muster actions using a virtual environment. The results show that a person's response during an emergency situation can be probabilistically quantified using the BN approach. Manipulation of the BN enables a review of which PSF(s) affected individuals during emergency scenarios. The results allow for adaptive training to enable individuals to improve on the areas where they are less than competent. It also shows areas that participants are stronger in, thus indicating the potential for taking on more leadership roles in an emergency situation. This benefit addresses the points made in Flin Flin & Slaven (1995) that state understanding how participants make decisions in a high pressure simulated environment can provide insight into the reactions they may experience in a real life situation and determine if they are competent to command an emergency.

The results also show other information that can lead to an improvement in initial safety training. Most participants were affected by the situation familiarly PSF, indicating participants could not apply the knowledge and skills learned to an area they were not familiar with. As the situation familiarity PSF was the PSF that affected participants' success probability the most, it indicates additional training in spatial awareness would be required to improve the probability of successful completion of emergency procedures.

The PSFs selected for this experiment were based on the experience and opinion of the researcher and the research group. It is noted that other PSFs can be selected and developed for different scenarios. This may result in different results for each individual. The results presented are specific to the selected experiment PSFs. To determine if other factors affect an individual's results, the experiment would need to be repeated with new PSFs.

This study also primarily focuses on individual human factors. Future work would benefit from inclusion of organizational and teamwork performance shaping factors. The addition of such factors would demonstrate the importance of a company`s safety management program and how well an individual performs in a group situation. Examples of how to implement these factors would be to create a training program where participants can interact with other participants or agents to complete the required tasks during an emergency situation.

References

Anonymous. (2017). *Academic Use of our Products.* Retrieved from BayesFusion, LLC: http://www.bayesfusion.com/

Bell, J., & Holroyd, J. (2009). Review of human reliability assessment methods. *Health and Safety Laboratory, United Kingdom*.

Boring, R. L. (2010). How many performance shaping factors are necessary for human reliability analysis? In *Proceedings of 10th International Conference on Probabilistic Safety Assessment and Management,*. pp. 1479-1487.

Chang, Y. H. J., & 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.

Cullen, W. (1990). The public inquiry into the Piper Alpha disaster. H.M. Stationary Office, London.

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.

Doody, K., Smith, J., and Veitch, B. (2017). Retention of skills attained through simulation based mastery learning in the All-hands Virtual Emergency Response trainer (AVERT). *Ocean Engineering Research Centre Report 2017-006,* 75p.

Flin, R. H., & Slaven, G. M. (1995). Identifying the right stuff: Selecting and training On-Scene emergency commanders. *Journal of Contingencies and Crisis Management, 3*(2), 113-123.

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.

Groth, K. M., & Mosleh, A. (2012a). A data-informed PIF hierarchy for model-based human reliability analysis. *Reliability Engineering and 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., 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 and System Safety, 128,* 32-40.

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 and System Safety, 115*, 33-42.

Haldar, A., & Mahadevan, S. (2000). *Probability, reliability, and statistical methods in engineering design*. John Wiley.

Hänninen, M., Banda, O. A. V., & Kujala, P. (2014). Bayesian network model of maritime safety management. *Expert Systems with Applications*, *41*(17), 7837-7846.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, *3*(3), 203-220.

Kim, J. W., & Jung, W. (2003). A taxonomy of performance influencing factors for human reliability analysis of emergency tasks. *Journal of Loss Prevention in the Process Industries, 16*(6), 479-495.

Kirwan, B. (1994). *A guide to practical human reliability assessment*. CRC Press.

Kirwan, B. (1996). The validation of three human reliability quantification techniques – THERP, HEART and JHEDI: Part 1 – technique descriptions and validation issues. *Applied ergonomics*, *27*(6), 359-373.

Kirwan, B. (1997). The validation of three human reliability quantification techniques – THERP, HEART and JHEDI: part iii – Practical aspects of the usage of the techniques. *Applied Ergonomics*, *28*(1), 27-39.

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.

Montgomery, D. C. (2008). *Design and analysis of experiments*. John Wiley and Sons.

Moyle, M., Musharraf, M., Smith, J., Veitch, B., Khan, F. (2017). Human Reliability Analysis Using Virtual Emergency (VE) Scenarios via a Bayesian Network Model. *Proceedings, C-RISE 3rd Workshop & Symposium.*

Moyle, A., Veitch, B. (2017). Human Reliability Analysis using Virtual Emergency Scenarios via a Bayesian Network model. *Ocean Engineering Research Centre, 2017-007.*

Musharraf, M., Hassan, J., Khan, F., Veitch, B., MacKinnon, S., & Imtiaz, S. (2013). Human reliability assessment during offshore emergency conditions. *Safety Science*, *59*, 19-27.

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 and System Safety, 132*, 1-8.

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. (2018). Incorporating individual differences in human reliability analysis: An extension to the virtual experimental technique. *Safety Science*, *107,* 216-223.

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

Norazahar, N., Khan, F., Veitch, B., & MacKinnon, S. (2014). Human and organizational factors assessment of the evacuation operation of BP Deepwater Horizon accident. *Safety Science*, *70*, 41-49.

Rangra, S., Sallak, M., Schön, W., & Vanderhaegen, F. (2015). Human reliability assessment under uncertainty – towards a formal method. *Procedia Manufacturing, 3*, 3230-3237.

Smith, J., & Veitch, B. (2019). A better way to train personnel to be safe in emergencies. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering*, *5*(1), 011003.

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