Using Dynamic Nonparametric Bayesian Belief Nets (BBNs) to Model Human Influences on Safety

A Potential Tool for Safety Management

Master Thesis

by

Wiebke Jäger
Faculty of Technology, Policy Analysis and Management
Delft University of Technology

Student no:	4125576	
Program:	Engineering and Policy Analysis	
Graduation date:	February 2013	
Graduation section:	Safety Science	
Examination committee:	Prof. dr. B. J. M. Ale, Safety Science	Chairman
	Dr. D. M. Hanea, Safety Science	First supervisor
	Dr. J. H. Slinger, Policy Analysis	Second Supervisor

Acknowledgments

First I would like to thank dr. Anca Hanea, dr. Dorota Kurowicka and dr. Daniela Hanea who supported me in my ambition to find a topic that has enough subject matter for two theses, one for the Engineering and Policy Analysis program and one for the Applied Mathematics program.

Further, I would like to thank my thesis committee consisting of Prof. dr. Ben Ale, dr. Daniela Hanea and dr. Jill Slinger for administering my thesis. Prof. dr. Ben Ale's enthusiasm for the topic has been a great source of inspiration. Dr. Daniela Hanea, my first supervisor, was always available and had an open mind for my worries, no matter how busy she was herself. I am especially grateful for her guidance and encouragement when I felt a bit "drifting" in the first stages of the research. I would also like to thank dr. Jill Slinger for all her helpful comments and her exceptional involvement in the experiment. Without her help, the thesis might not have been in the form it is now. Of course, I am also thankful to her students, who were willing to participate in the experiment.

Special thanks go also to dr. Bert Enserink who made my impossible wishes for this double degree program possible.

Finally, I would like to thank Laurence who had great ideas for suitable games and all my friends who served as first guinea pigs.

Wiebke Jäger

Summary

In potentially hazardous industries, human and management influences are widely recognized as more fundamental causes to accidents, than random technical failures. Often, engineering systems have been drifting towards a high risk state for a long time, before an "unsafe act" triggered the accident. For safety management the variable of interest is the human error probability (HEP) and the goal of safety management systems (SMS) is to design management actions that keep the HEP within acceptable bounds. However, models representing the intricate and dynamic relations between the human and technical components of the system are in their early stages.

Common probabilistic risk analysis (PRA) tools for modeling risks associated with the technical components of the system are event trees (ETs) and fault trees (FTs). However, they have restricted capacities to incorporate uncertainties and to account for common cause failures. For this reason they are inappropriate to model human performance. Bayesian belief nets (BBNs) seem to be a promising tool to integrate human and technical influences on safety in a single model. They are a type of probabilistic graphical model consisting of nodes (variables) and arcs (influences). Their main use is to make inferences about uncertain states when information is limited and additionally ETs and FTs can be converted in BBNs so as to build a homogeneous model. However, a main short coming is the generally poor reflection of dynamic relationships, which makes it difficult to observe the implications of management actions on the future. Up to now the applications of dynamic BBNs have been very limited and the advantages and disadvantages of this methodology still have to be explored. This leads to the following research question:

How can dynamic nonparametric BBNs support the design of improved SMS by enabling a monitoring of the effect of management actions on safety?

However, it is very difficult to realistically define and quantify human factors and their dynamic influences, because they are not measurable. In general so-called proxies have to be used: operational definitions that capture a part of the human factor, which is considered to be important for a particular application. Since a real application requires a wide literature search including accident reports and human reliability data bases as well as consultation with experts, it is beyond the scope of this research. Therefore the adopted approach is to quantify a delimited demonstration case with an experiment and to explore, on the basis of this case, the potential of dynamic nonparametric BBNs to (1) realistically reflect human factors and (2) to design and monitor management influences on them.

The demonstration model represents, in a strongly demarcated system, the qualitative and quantitative relationship between the management action *training* and the human factor knowledge over time. The experiment constitutes the system: Students play four rounds of the online puzzling game *3Dlogic* and the level they were assigned to complete is interpreted as training, while a normalization of the required time to complete this level is interpreted as knowledge. This

interpretation accentuates how difficult it is to measure these kind of factors and how case specific operational definitions are.

Based on the data gathered from the experiment, the relationship between training and knowledge has been modeled for four time steps. In principle the model could be extended to arbitrarily many time steps, but the nodes and arcs of each step need to be quantified individually, which sets practical limitations.

The software Uninet enables the analyst to easily quantify the model based on the raw data in a mathematically sound way. However, several structures may be statistically valid and it is up to the analyst to decide which nodes should be connected and in what direction the influence is. The "best" of all (statistically) valid structures is the one suiting the interests and needs of the user most. For the case treated in this thesis, there are few grounds for making such a choice. Instead, it has been investigated what the consequences for the model output for two different plausible structures.

BBNs can be used in two ways to design management actions (1) by using predictive reasoning and (2) by using diagnostic reasoning. The fact that the model structure depicts cause - effect relationships in an intuitive manner facilitates the design process. For the first approach, a hypothetical management action is translated to the context of the experiment and implemented in the model. This is done by conditioning on a certain variable, e.g. conditioning on high training. The model is updated and the resulting distributions for all other variables are shown. Using this approach the potential effect of a management action on a human factor can be explored. For the second approach the desired value of the human factor is prescribed to the model and from the updated distributions of the other variables it can be inferred what management actions are necessary to achieve this value. For instance, when conditioning on "high knowledge" the distributions of training shift to higher values as well as leading to the (for this demonstration case obvious) conclusion that one should train individuals more to make them more knowledgeable.

The output of dynamic nonparametric BBNs is very suitable for the purpose of risk analysis. The representation of variables in terms of probability distributions reflects the inherent variability in human performance. It includes an indication of the mean and the variance, which immediately shows the trend of the influence: does it increase or lower the value of the effect - variable and how certain is that (roughly)? But moreover, the whole range of possible values is expressed, which makes extreme values with very low probability visible. Especially in hazardous industries, it is important that risk models represent events that almost never occur, because their consequences may not be acceptable and/or affordable and need to be mitigated by all means. This characteristic of BBNs enables the designing of management actions that are especially targeted at preventing low-probability-high consequence events.

The contribution of the thesis is two-fold. On the one hand the results described above add to the limited literature of human performance models in risk analysis. On the other hand, it is described at length how a dynamic nonparametric BBNs can be quantified and validated from data. To our knowledge, this has not yet been discussed in the literature and may also be of interest to various other applications in research and the industry.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The starting point of this research was that human factors have widely been cited as the fundamental causes for accidents in high hazard industries demonstrating the need for a wider reaching risk analysis, which incorporates them in the risk models. Well known cases are e.g. the accident of the chemical plant in Bhopal, India (1984) [Marais et al., 2006], the deepwater horizon blow-out in the gulf of Mexico (2010) [National Commission on the BP Deepwater Horizon Oil Spill and Offshore Drilling, 2011] or the Fukushima Daiichi nuclear disaster (2011) [Los Angeles Times, 2011],

Before describing the problem and objectives of the research in more detail, the following section introduces the concepts of safety, risk and uncertainty and indicates the position of the research within the field of safety science.

1.1 Safety, Risk and Uncertainty

Safety is the "condition of being protected from [hazards], harm or risk" [Oxford University Press, 2005]. *Risk* is a closely related concept. Throughout the literature the term is used in many different ways, but (probably) all definitions include the two dimensions *extent of consequence* and *probability of occurrence*. In the most basic version it is the product of the two.

Risk management is the process of keeping hazards under control, i.e. reducing or eliminating the probability for events with unacceptable consequences. However, society's perception of risk generally does not agree with the basic definition above. Besides on consequence and probability of a risk event, the willingness to accept certain risks seems to depend on additional, often non-quantifiable aspects. Uncertainty, potential extent of damage, involuntariness or uncontrollability have been identified as major factors. For example, the risk associated with a fatal accident of a nuclear power plant is much smaller than the one associated with a fatal traffic accident. Nonetheless, it is not necessarily more accepted by society, which leads to the conclusion that the notion of risk is more complex. [Ale, 2009]

For (political) decision-making concerning potentially hazardous industries, such as nuclear or oil and gas, it is important that all actors have the same understanding of the risks. However, various actors, e.g. companies, governmental agencies, environmental organizations and residents, are likely to have different interests and therefore have strong motives to question the information presented by other parties. In particular they may not have the same opinion on what is acceptable or not and including this rather intangible concept in the definition of risk makes it vulnerable to contesting. The various actors are likely to question the estimated risk presented by the opposing party for the activity under discussion, because it serves as a validation for the claim of safety or to demonstrate the need for further safety improvement, which is often the controversial subject. This is a classical dilemma of multi-actor decision making and a solution is to let the parties negotiate on the "correct" knowledge [de Bruijn and ten Heuvelhof, 2008]. Consequently, whether an activity is acceptable or not and what the role of risk and other factors is needs to be discussed and agreed upon by the actors involved.

Risk analysts can contribute by providing objective and incontestable information regarding risks, which are the extent of consequence and the probability of occurrence, and leave other aspects open for debate. However, for complex technological system, the computation of consequences and their associated probabilities is not trivial. The system's behavior is usually determined by an intricate combination technical, scientific and inter-personal components and complex mathematical models are required to reflect that, although it is not possible to accurately quantify all the relevant elements and relationships [Enserink et al., 2010]. Admittedly, all models are "false" and their predictive quality depends on how uncertainties about the elements as well as the structure are treated.

The notion of *uncertainty* is tight-knit with the notion of risk. Uncertainty "is that what disappears when we become certain" [Bedford and Cooke, 2001, p.19]. In a practical context, certainty is achieved through observation or, said differently, by learning about the system. However, not all kinds of uncertainty can be observed; there is a distinction between *aleatory* and *epistemic* uncertainties. Aleatory uncertainties refer to the natural variability of the system; they cannot be learned. In contrast, epistemic uncertainties refer to lack of knowledge of the system. A quantitative measure of uncertainty is probability. Having absolutely no clue about the outcome of an event can be interpreted as a uniform distribution. Knowledge about the event is reflected by shape and range of the distribution. [Bedford and Cooke, 2001]

Human factors, the topic of this thesis, include both types of uncertainty. At least for the purpose of modeling human factors, the following categorization is useful[1]. Predicting the performance of a random individual in a random situation is impossible. Indeed, a significant part of human behavior is cryptic to bystanders: genetic endowments and previous experiences of life are probably the main determinants. This is the aleatory uncertainty. However, we could learn about certain preconditions of the individual that affect his/her performance, e.g. knowledge, stress level or fatigue.

For models in the field of safety management the variable of interest is the probability of human error and how it can be controlled brings us back to the research problem.

1.2 Research Problem

Risk management of hazardous industries quests for "no-accident-is-tolerable" strategies [Rasmussen, 1997]. In spite of all efforts, low probability high-consequence accidents, such as the recent blowout in the Gulf of Mexico and the nuclear disaster in Fukushima, do not appear to occur less frequently, while their consequences appear to increase. This indicates that current approaches to risk management do not render the goal of zero accidents possible and improved safety management systems (SMS), i.e. systems designed to manage risks [Ale, 2009] are of great social relevance.

[1] cf. [Bedford and Cooke, 2001, section 2.6] for an example on how the same uncertainty may be classified differently for different models with different goals.

It is recognized that a complex combination of human and organizational factors are often more fundamental causes of major accidents than random technical failures. Hence, there is a need to integrate those factors with the technical aspects of risk models in order to achieve a wider-reaching risk analysis and assessment (e.g. [Ale et al., 2012, Mohaghegh et al., 2009]). Moreover, management actions have an influence on human and organizational factors contributing to fatal accidents. However, it is extremely difficult to identify underlying failures in management, since they may not be closely related in time and space to the event and a systematic model. To see the broad picture, it is indispensable to monitor the effectiveness of management actions to improve safety [Lin, 2011].

Popular probabilistic risk assessment (PRA) methods, such as event trees (ET) and fault trees (FT), are very useful when modeling risks associated with technical systems, but have very limited potential to model human and organizational factors [Mohaghegh et al., 2009]. Bayesian belief nets (BBNs) seem to be a promising tool to integrate human and technical influences on safety in a single model, because the variables are quantified with probability distributions which allows to explicitly model uncertainties. A recent application is CATS [Ale et al., 2009a], a model for air transport safety. However, this model is static and therefore not able to reflect certain important influences. For instance, the effect of management actions to control risks and the personal risk perception of an operator are inherently dependent on time. Management actions may only have visible outcomes after a couple of months and the personal risk perception is generally higher shortly after an accident.

A support tool that allows monitoring of management actions on safety and captures human and organizational influences more realistically is still lacking, which leads to the following problem statement:

> *As yet, the possibilities to model human and organizational factors as part of an integrated risk model are limited, which makes it difficult to monitor the effect of specific management actions on safety.*

Recent mathematical developments theoretically allow for dynamic BBNs [Kurowicka and Cooke, 2006]. Up to the present, applications of dynamic BBNs are not well-known and the advantages as well as disadvantages of this methodology still have to be explored. Further, guidelines or consideration for the validation process of such a model are needed. Besides these rather mathematical issues, the knowledge gaps also include taxonomies of human factors. A realistic set of human factors, including their dependences in possibly very different time frames, as well feedback loops has not yet been established. The identification of the various factors from different levels in the organization, external pressures on the company or individual career planning on the safety culture is needed to establish such a list. Additionally, it is not clear if and to what extent the model can be used to design management actions and to test the effect of management actions.

Based on the problem statement and the identified knowledge gaps, the following research objective is formulated:

> *To introduce the methodology of dynamic nonparametric BBNs as a useful tool to design improved SMS through monitoring the effect of management actions on safety*

The project aims to deliver a list of considerations for building and using a realistic model that emphasizes human and organizational influences as well as the effect of management actions on safety. This list is based on the analysis on a review of previous models and of a demonstration model that exemplifies how dynamic nonparametric Bayesian belief nets can be used to monitor the effect of management actions on human performance. The following section describes the research questions and the research approach.

1.3 Research Questions and Approach

From the above discussion and formulation of the research objective, the main research question is:

How can dynamic nonparametric BBNs support the design of improved SMS by enabling a monitoring of the effect of management actions on safety?

To be able to answer this question, four sub-questions are determined and guide the research process. They are stated below and the research methods as well as data requirements are described for each of them.

1. Why is monitoring the effect of management actions relevant to improve safety of engineering systems?

   This sub-question is answered by means of desk research. As a first step, the impact of human and organizational factors on system risk and how management actions can influence them is analyzed.

2. How can human factors and management influences be incorporated in technical risk models?

   Also this sub-question is answered through desk research. Two existing models are known to emphasize the human and organizational influences on safety: a causal model for air transport safety (CATS) and a socio-technical risk analysis (SoTeRiA). Both models are reviewed, their main characteristics highlighted and their limitations pointed out. As a second step, taxonomies of human factors are described and existing time dependences and feedback loops between management actions and human factors are explored. Finally, it is discussed why human factors are particularly difficult to operationalize and to model in a quantitative way.

3. What is the added value of dynamic nonparametric BBNs when modeling human factors and management actions?

   As a first step, the literature is reviewed to explore the current state of the art of dynamic nonparametric BBNs. Then, on the basis of an experiment with 24 university students, a demonstration model is built consisting of one management action (training) and one human factor (knowledge). This model is used to discuss the potential of dynamic nonparametric Bayesian belief nets to realistically represent human factors.

4. How can management actions be developed with dynamic nonparametric BBNs?

   This last sub-question is answered by analyzing the demonstration model and exploring its output when implementing hypothetical management actions. Furthermore, the limitations of the methodology are discussed and it is indicated what has to be taken into account when extending the demonstration model to a realistic model. Finally, guidelines on how to validate a realistic model are given. This is exemplified on the demonstration model.

Reflecting on the method and the deliverables of this proposal, the following limitations should be considered. This project aims to put forth a new scientific methodology to improve SMS by facilitating monitoring of management actions. Hence, it suggests steps that could be taken to design improved safety system, but it cannot give advice on immediate actions.

Finally, following the proposed considerations to build and use a management risk model based on the introduced methodology may not be sufficient to construct a good realistic model. Elaborate industry specific research and thorough verification and validation of the realistic model will be necessary. Moreover, the development of successful management actions is not a guaranteed result.

1.4 Structure of the Thesis

The reminder of this thesis is organized as follows. Chapter 2 carries out a review on human error and its role for safety in high-hazard engineering systems. Chapter 3 provides the theoretical background for the risk analysis methods that are mentioned and used in this project. Most importantly, dynamic nonparametric Bayesian belief nets (BBNs) are introduced in this chapter. In chapter 4, hitherto existing achievements in human factor modeling from other projects are described. Challenges that arise when attempting to operationalize human factors and to find causal relations between them are discussed in detail. Chapter 5 motivates the need for an experiment to quantify a demonstration model. The data generated by the experiment are presented and discussed. Then, chapter 6 engages in the building process of a dynamic nonparametric Bayesian belief net and illustrates how the model can be used. Finally, conclusions are given in chapter 7.

Chapter 2

Human Error in Engineering Systems

High-risk industries are complex socio-technical systems where humans have been identified as the 'weakest link' [Turner and Pidgeon, 1978]. Unsafe acts by individuals can be categorized broadly as either errors or violations [Reason, 1990]. Both may occur in various tasks, such as regulation, administration, management, design, installation, maintenance, inspection or operation. When attempting to control high-hazard systems to err may be fatal. However, to err is human and, to a certain extent, inevitable. Therefore, a sensible approach to control risks is not only targeted at avoiding errors, but also at early error detection and consequence mitigation. The next sections present a framework of aberrant behavior proposed by [Reason, 1990].

2.1 Classification of Human Error

2.1.1 Intended and Unintended Actions

There are many working definitions of human error. The classification of human error described here is based on the definition by Reason, who links the notion of error to the notion of intention, that is the intention to do something right: Human error is "a generic term to encompass all those occasions in which a planned sequence of mental or physical activities fails to achieve its intended outcome, and when these failures cannot be attributed to the intervention of some chance agency" [Reason, 1990, p.9] . This definition includes two kinds of failures: (1) Slips or lapses are failures to reach a predefined goal (e.g. a calculation error). They are unintended and occur during the execution of a task. Usually the actual execution deviates from the intended execution, if a largely automated task is to be performed in a very familiar setting while being distracted from the job. For this reason slips and lapses are also termed as execution failures. (2) Mistakes, or planning failures, arise from inadequate planning. In other words, they appear when an individual correctly carries out an intended action, while this intended action is inappropriate to achieve the desired outcomes. Mistakes originate from higher-level cognitive processes than execution failures. More precisely, they are failures in the judgmental and inferential processes to select an objective or the appropriate means to achieve it. Consequently, these kind of errors are much harder to detect that execution failures. [Reason, 1990]

Contrary to errors, violations are deliberate actions that deviate from prescribed procedures, codes of practice, rules and the like, but generally do not have malevolent intentions [1]. Of course there are violations that have no prior intentions, so-called erroneous violations, and which can be attributed to the class of unintentional errors. More important are so-called routine violations and exceptional violations. Routine violations have two underlying causes. On one hand it is in human nature to take paths of least effort. On the other hand production targets may be set unrealistically high, so that they cannot be met while complying with all safety procedures at the same time. As a result, safety procedures will be violated routinely, if they look rather trivial and their transgression is rarely tracked and punished. Exceptional violations occur when the system is in an exceptional state creating double binds making it inevitable for the operator to violate, even though he has only good intentions, at least one procedure. In this case , the operator deems the violation necessary to ensure safe operation or to perform damage control. [Reason, 1990]

The above classification rests upon the different ways the process to achieve a desired outcome may fail. It indicates that errors are related to individual cognitive processes, more precisely to information processing processes, whereas violations have a social context. The differentiation between violations and the few error types is summarized in figure 2.1.

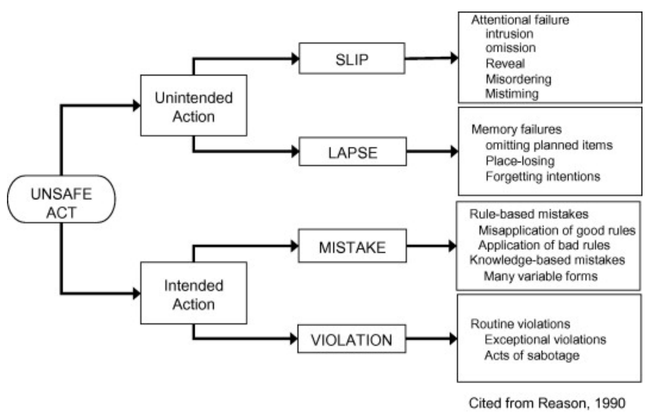


Figure 2.1: Classification of Unsafe Acts (source: [Reason, 1990])

2.1.2 Active and Latent Errors

A different way to distinguish errors is according to their appearance in the system: as active or as latent errors. This distinction is helpful to understand how accidents arise as concurrence of many human faults.

Active errors have immediate, visible consequences and are typically committed by front line operators, e.g. control room operators. Latent errors, on the contrary, are primarily associated with high-level decision makers and managers. They are dormant for a long time before an accident sequence begins and only become evident when they coincide with other faults. Normally, large scale accidents are released by active errors and or local triggering events (e.g. weather), but the root causes have long been present. In other words, active errors are only the tip of the iceberg; the latent errors are the underlying causes of man-made disasters. Latent errors may drive systems in exceptional states and contribute to the incidence of active errors.

[1] Violations that intent to damage the system are sabotage. Sabotage did not play a role in large-scale accidents and for this reason is not included in the classification.

Accidents rarely evolve along predictable lines, even though they may begin in a conventional way. If they would evolve along predictable lines, there would be no difficulty to prevent them. Every accident is unique and absolutely novel. For this reason, the operator's hitherto existing experience cannot help him to cope with the out-of-tolerance system state and he is bound to make (active) errors. This means that active errors, which eventually release an accident, are "in large part the delayed effects of system design failures" [Reason, 1990, p.183]. Evidently, many latent errors may be categorized as organizational errors [Stewart and Melchers, 1997], i.e. failures arising primarily at the managerial and organizational level and combining adversely with local triggering events as well as the active failures of individuals at the execution level [Reason, 1997].

In order to improve system safety it is thus more effective to neutralize latent errors, than active errors: They are much longer "alive" and attackable than active errors, which furthermore may be impossible to eliminate when significant latent errors are present. Where within the organization latent errors arise and how they propagate is elucidated in the following section.

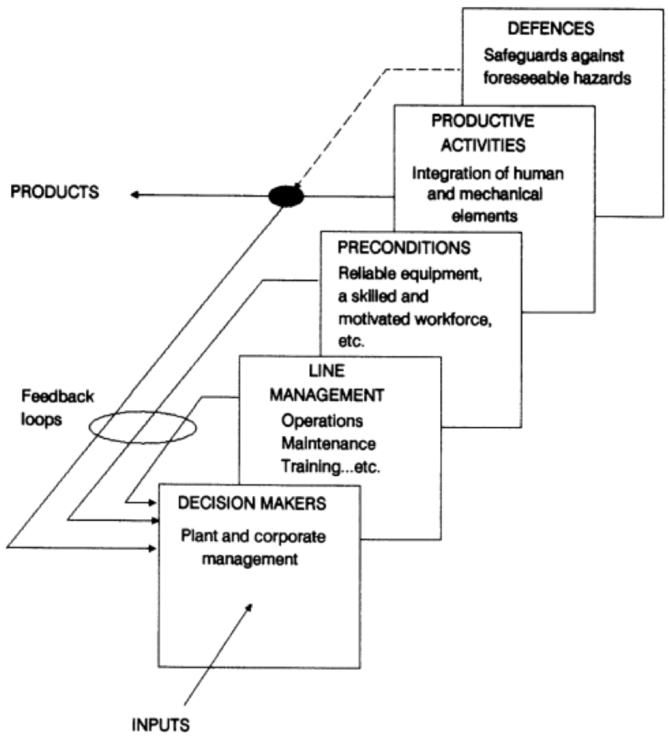
2.2 Error creation and propagation in organizations

There are two well-known, high-level conceptual models of error propagation in engineering systems. The first is put forward by Reason [Reason, 1990, Reason, 1995], the second by Rasmussen [Rasmussen, 1997]. While the models show considerably many similarities, there are a few differences. Reason's model focuses on organizational levels within the company, but also incorporates possibly external factors with psychological influence on the individuals. Rasmussen's model has broader scope in the sense that it includes organizational levels up to regulators and the government. Nevertheless, it only contains the organizational levels and no other elements. The two models are not contradictory; if anything, they complement each other providing a wider understanding of error propagation and its consequences. First, we describe the elements of both models, then we discuss some dominant error propagation mechanisms.

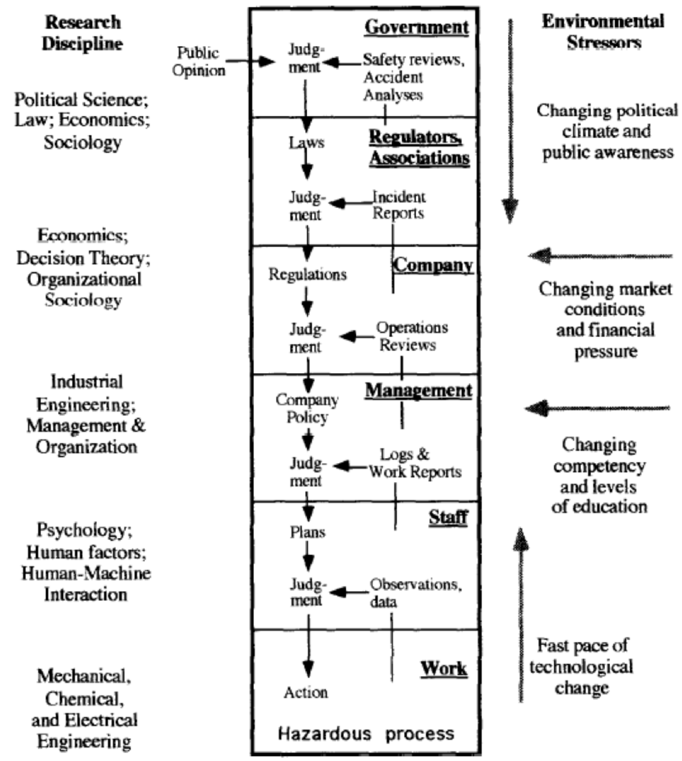
2.2.1 Conceptual Framework of Error Propagation

Reason demonstrates how errors propagate through an organization using five basic 'building blocks' which are present in any complex system. The building blocks are displayed as five individual planes in figure 2.2a: decision makers, line management, preconditions, productive activities and defenses. The decision makers set the company goals and choose the means to reach them. They allocate (finite) resources, such as money, equipment, people or time in order to maximize productivity as well as safety. On the next level are the line managers, who are for example responsible for operations, training or maintenance and implement the strategies formulated by the decision makers. The next plane constitutes preconditions, such as reliable equipment, attitude, schedules, motivation or environmental conditions. The productive activities are the interactions of the people at the front line with the technical components of the system creating the product of the company. Finally, potentially hazardous industries have a number of (automated) defenses to protect individuals and machines against any foreseeable damage.

Rasmussen claims that safety depends on the control of the work-processes on all levels of the socio-technical system, as depicted in figure 2.2b. Half of the levels can be found again in Reason's model: The company level corresponds to the decision maker plane, the management level resembles the line management plane, and the work level is the same as the productive activities. Additionally, Rasmussen considers the government as the top level, where society seeks to



(a) The "elements of production" (source: [Reason, 1990])



(b) The socio-technical system involved in risk managemengt (source: [Rasmussen, 1997])

Figure 2.2: Different Organizational Levels Influencing Failures

oversee safety through the legal system and by setting boundaries of acceptable human and societal conditions. On the next level authorities, industrial associations and workers' unions interpret the legislation and implement it in rules for specific groups of employees. On the company level, these rules are once more interpreted and implemented, this time in the context of the company and its goals. High level decision makers choose suitable risk control measures and allocate resources. The management, on the subsequent level, has insight in local conditions and processes and make the company's objectives operational, including both its production goals as well as its safety goals. As described above for Reason's model, management is for example responsible for operational procedures, training or maintenance. Furthermore, Rasmussen's scheme includes the level of the staff between the management and the work level. The staff contains the individuals who directly interact with the technical components of the system and perform the 'actions' on the work level. They execute the standard operating procedures keeping the working process within the safe limits of operation. Moreover, they have to cope adequately with events that were unthought of at higher levels and have not been included in procedures or rules [Lin, 2011]. The work level describes all the actions or processes that directly involve the technical system.

2.2.2 High Level Decision Makers and Management

Some of the decisions high level decision makers and managers on various levels take proof to be amiss later on. This is an unavoidable part of the management process, since it is human to be mistaken. Identifying all kinds of random errors that may be committed on the management level is, just as modeling them, hardly possible; the set of possible errors affecting safety may be infinite. It is, however, worthwhile to consider the context in which decisions are made.

Managers on all levels have to unite two competing goals, safety and high productivity [Reason, 1990], while their performance of doing so is monitored. Commonly, key performance indicators (KPI) are used to observe and guide management practices. KPIs are performance measures of the system under control, and such, indirectly of the managers in charge. However, KPIs for safety are mostly probabilistic making it hard to prove if management has met the required safety goal. This is particularly true for large-scale disasters, whose probability is extremely low so that they are high unlikely to occur during the period a manager is in charge [Ale et al., 2012]. For this reason, managers may be more motivated to control short-term risks and direct efforts towards productivity, as they are immediately visible in KPIs, but fail to mitigate the risk of rare disaster. Being able to monitor all management actions that aim to control risks, can be a crucial incentive for managers to mitigate low probability risks. Moreover, financial pressure inevitably creates a bias in decision making in favor of time- and cost-savings at the cost of safety [National Commission on the BP Deepwater Horizon Oil Spill and Offshore Drilling, 2011] and can, together with other organizational factors, drive engineering systems systematically towards a high risk state where almost any abnormal behavior could release an accident [Marais et al., 2006], [Rasmussen, 1997].

2.2.3 Lower level Management, Preconditions, Unsafe Acts and Defenses

The relations between lower level management, preconditions, unsafe acts and defenses are much more direct and concrete than the influence of higher level management.

Failures in lower level management affect for example the quality of training or the suitability of procedures and schedules or other preconditions for the staff. The impact of "bad" line management on preconditions may have various consequences; there is a so-called many-to-many mapping. Deficiencies in training can manifest themselves in undue time pressure, inappropriate perception of hazards, ignorance of the system, motivational difficulties etc. Likewise, undue time pressure can be due to poor procedures, poor scheduling, deficiencies in skills, knowledge or maintenance inadequacies).

However, unfavorable preconditions need not necessarily result from deficient line management. They may be out of the company's reach, but be brought from outside to the workspace, e.g. being stressed, motivational issues, depression, harmful psychological effects or negative personal life effects (separations, bereavements, etc.). Between preconditions and the actual human faults exists a some-to-many mapping: one bad precondition may open the pathway for many unsafe acts. In this way it can be regarded as a *common cause* for various failures.

Due to the some-to-many mapping between, remedial efforts upon preventing the recurrence of unsafe acts are hardly of avail. Some may fall into recognizable classes, e.g. not wearing safety glasses and can be subjected to targeted safety programs and trainings, but others may be absolutely unforeseeable. It is more effective to improve the preconditions. Against foreseeable unsafe acts a number of defenses are generally in place. A defense may simply be personal safety equipment, e.g. safety glasses, or consist of many redundant automated safety systems, "defenses in depth".

2.2.4 Emergence of Accidents

It is widely acknowledged that large scale accidents result from a number of errors, which independently would not have been sufficient to cause them. Rasmussen [Rasmussen, 1997] suggests they are not caused by a number of independent failures and human errors, but by "a systematic migration of organizational behavior toward accident under the influence of pressure toward cost-effectiveness in an aggressive, competitive environment" [Rasmussen, 1997, p.189].

Decisions made by several different actors in the organization are likely to be under the same kind of competitive stress and are thus not independent of each other. Individuals in the company make decisions at different times and in different parts of the organization, which together prepare an accident. The trigger can be a single human act, especially because staff at the front end can hardly overlook the total picture during their daily operational decision making and are incapable to judge the state of the multiple defenses conditionally depending on decision taken by other people in other departments. [Rasmussen, 1997]

2.3 The Detection of Human Errors

The efficacy of error detection is crucially dependent on the immediacy and validity of feedback information. At low level (execution error), feedback is supplied directly, but at higher levels the information is unavailable or at least open to many interpretations. According to [Reason, 1990] there are only three detection modes for human errors: (1) Self-monitoring, (2) environmental clues and (3) detection by other people. Self-monitoring is based on feedback control: The deviations of output from an ideal state are fed back to the controlling agency of the brain, which tries to minimize these discrepancies. This is most effective on the skill level. The most obvious environmental clue is, if the environment blocks onward processes until the error has been corrected. This is called a forcing function. Detection by other people is crucial for mistakes. It appears to be the only way to detect mistakes. Mistakes are hardly detected by the people failing, since there is no discrepancy between action and intention. A pair of fresh eyes is needed to expose the inappropriateness of the plan.

2.4 Chapter Conclusions

This chapter reviewed error classification, error propagation in organizations and the resulting emergence of accidents. It has been established that various actors at different levels in the organization are likely to experience the same kind of managerial pressure towards cost-effectiveness. Hence, their decisions and actions are not independent and may disable several defenses at once. Managers may be more motivated to control short-term risks and direct efforts towards productivity, as their success becomes immediately visible, and ignore the risk of rare disaster. Since it is "rare", such an event is very unlikely to happen during their time in charge and also their efforts are not visible and may not be rewarded. A model that monitors the effects of all kinds of management actions on safety may make these efforts more visible and motivate managers to more intensely control low-probability risks.

Less abstract, there is an influence from higher level management on the line management, which in turn affects the preconditions for the staff directly interacting with the system. Because of the some-to-many mapping of preconditions and unsafe acts by the staff, e.g. inappropriate training can manifest itself in lack of knowledge, undue time pressure or ignorance of hazards, it is more effective to attempt to manipulate preconditions instead of trying to prevent individual unsafe acts. Additionally, resulting unsafe acts may even be completely unforeseeable.

Even though, "bad" preconditions surely influence human performance they need not manifest themselves in human error; they affect the error probability. Because human error has only the two outcomes "present" and "absent", which are random, it is not an optimal concept for risk modeling. Instead, the term *human performance* with an associated human error probability (HEP) is used and preconditions are performance shaping factors which may modify the HEP.

Chapter 3

Methods in Risk Analysis

This chapters reviews the methods commonly used in risk analysis and risk modeling and provides the theoretical background for the subsequent chapters. First, event and fault tree analyses are discussed, which are the classical techniques in probabilistic risk analysis (PRA) of technical components. Then, System Dynamics (SD) described and Bayesian belief nets are introduced. As opposed to event and fault trees, these two approaches are not especially designed for risk analysis, but are used to support decision making in various disciplines. Examples of their application in risk modeling are given in chapter 4.

3.1 Event and Fault Tree Analysis

Event tree (ET) and Fault tree (FT) analysis are the basic tools to analyze and model technical failure of a system. The first uses forward logic, whereas the latter is based on backward logic. For this reason they are applied in different situations, or more commonly combined in one model. This section is based on [Bedford and Cooke, 2001]. Alternative references are given if used.

Event Trees

Accidents can have a wide range of possible outcomes and consequences depending on the course of events following an abnormal incident [Hanea, 2009]. In other words, an abnormal incident is associated with a number of risk scenarios resulting from different chains of events. A tool to identify the full range of risk scenarios are event trees[1] (ET).

An event tree is a pictorial representation of the possible sequences of events arising from an *initiating event*. An initiating event indicates a deviation from normal operation and poses a hazard on the system. As stated before, event trees use forward logic. Starting with an initiating event, they "propagate this event through the system [...] by considering all possible ways in which it can effect the behavior of the (sub)system" [Bedford and Cooke, 2001, p.99]. This results in a series of branches from the initial event. Each branch represents a possible chain of events characterized by the occurrence or non-occurrence of various intermediary events and defines a risk scenario (which is the end event). The intermediary events are called pivotal events and represent the functioning or malfunctioning of a (sub)system.[2] If all (sub)systems are functioning,

[1] Event trees (ET) are also called Event Sequence Diagrams (ESD). They are structurally similar to Decision Trees.

[2] Generally, more than two outcomes are possible. They have to be distinct and mutually exclusive [Hanea, 2009].

the system will return to its normal state.

Figure 3.1.shows an exemplary event tree for an explosion in a building. The occurrence probability of one outcome is calculated by multiplying the frequencies[3] following the paths leading to that outcome. For example, a *controlled fire with no alarm* results, if a fire starts, the sprinkler system functions properly and the fire alarm is not active. Its probability of occurrence (per year) is $10^{-2} \cdot 0.80 \cdot 0.99 \cdot 0.001 = 8.0 \cdot 10^{-5}$.

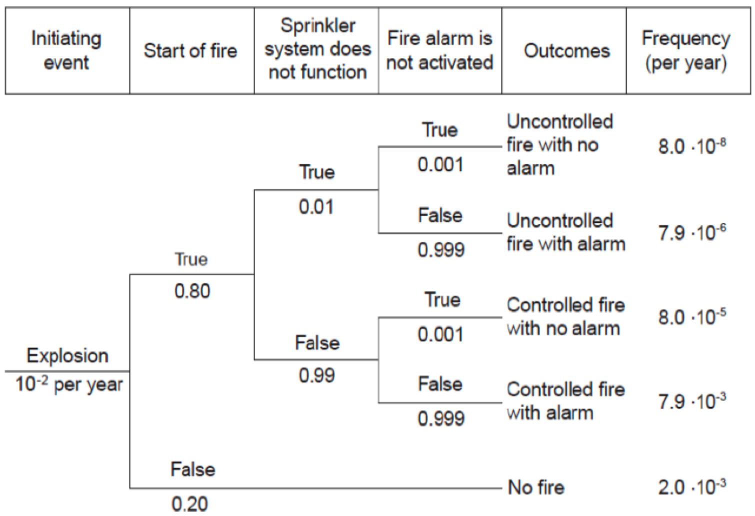


Figure 3.1: Example of an Event Tree for an Explosion in a Building (source: [Kurowicka, 2011] probably adapted from [Frantzich, 1998b])

Fault Trees

Fault tree analysis attempts to estimate the reliability of a system by analyzing how its failure results from a combination of individual (independent) faults. Fault trees (FT) are a pictorial representation of a system in boolean logic. As opposed to event trees, fault trees use a backward logic. They relate the occurrence of a so-called *top event* to its root causes, which are the occurrences or non-occurrences of *basic events*. The top event is described as a combination of intermediate events through logic gates, which in turn are described further in terms of lower level events until the *basic events* are reached. A tree structure results (figure 3.2). When the top event is the system failure, the basic events are the system components. The boolean logic requires that all events are binary, that is, true or false. This imposes restrictions when modeling situations where more than two states are important. Most common gates are the AND gate (all the causes have to be present to induce the event on the next higher level) and the OR gate (at least one of the causes is present to induce the event on the next higher level) [Hanea, 2009]. An example is the system power failure (figure 3.2). The power is usually provided by a generator. If the generator fails then a backup system starts operating, it switches over to a battery. The fault tree for this example is shown in 3.2. The system power fails if both generator and backup system fail. The backup system fails, if either switch or battery fail or both.

[3] The terms probability and frequency are interchangeable. In the frequentist interpretation, the probability of an event is defined as the limit of its relative frequency in a large number of trials.

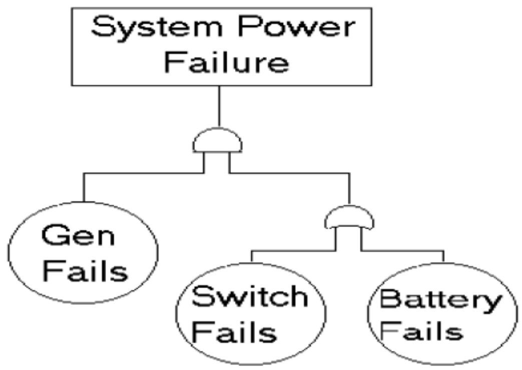


Figure 3.2: Example of a Fault Tree for System Power (source: [Frantzich, 1998a])

The probability of the top event can be computed by applying the laws of boolean algebra and basic probabilities to the basic events. The probabilities of the basic events are established from reliability data. For the example in figure 3.2, the probability of system failure can be computed as follows:

$$
\begin{aligned}
&P(\text{system power failure}) \\
&= P(\text{generator fails } \cap \text{ backup system fails}) \\
&= P(\text{generator fails}) \cdot P(\text{backup system fails}) \\
&= P(\text{generator fails}) \cdot P(\text{switch fails } \cup \text{ battery fails}) \\
&= P(\text{generator fails}) \cdot (P(\text{switch fails}) + P(\text{battery fails}) - P(\text{switch fails}) \cdot P(\text{battery fails}))
\end{aligned}
$$

Limitations of Event and Fault Tree Analysis

Many uncertainties cannot be included in event and fault tree analysis [Hanea, 2009], the most salient limitations are listed in this section. First, uncertainty in the model or the input data cannot be taken into account using fault trees. If all components (AND gate) fail or at least one component (OR gate) fails, the top event certainly fails. In reality, the occurrence of certain events may only increase the probability of the occurrence of another event. Second, the failure probability of a component is fixed. Uncertainty about this value cannot be incorporated in the tree. Third, human behavior can only be factor into the trees on a global level, that is not comprising a variation from person to person, and merely as the failure or success of performing a specific predefined activity.

More generally, it is difficult to model common cause failures with logical trees [Frantzich, 1998b]. A common cause affects several components, which means that their respective failures are not independent. In most analyses, the component failures are assumed to be independent to render computation feasible, even though common cause failures can be very important contributors to system failure, because they can disable several redundant safety systems at once [Bedford and Cooke, 2001].

These limitations imply that event and fault tree analyses are not sufficient tools to account well for the influence of human performance on system reliability. Attempts to combine event and fault trees with other techniques to overcome these problems is described in chapter 4.

3.2 Bayesian Belief Nets

Bayesian belief nets (BBNs), also called Bayesian Belief Networks, Bayesian Networks or simply Bayes Nets, are a type of probabilistic graphical model to represent a high-dimensional probability distribution [Bedford and Cooke, 2001] They are a tool for making inferences about uncertain states when the available information is limited. BBNs are used for a wide range of applications, such as diagnosis in medical science as well as various engineering disciplines, prediction, risk management, modeling of ecosystems, sensor fusion or monitoring and alerting [Norsys Software Corp., nd].

3.2.1 Intuition and Introductory Example

Many relationships between variables of a system are not deterministic, but probabilistic. For example, being a smoker increases the probability of having lung cancer, but it need not cause it. Probabilistic reasoning may be counterintuitive and requires some calculations with probabilities. For instance, if we assume that 0.05% of the population have lung cancer, 40% of the population smokes and the smokers account for 85% of all lung cancer cases, what would be the probability of having lung cancer when being a smoker? Applying Bayes rule yields

$$P(L|S) = \frac{P(S|L) \cdot P(L)}{P(S)} = \frac{0.85 \cdot 0.0005}{0.4} \approx 0.001 = 0.1\% \tag{3.1}$$

where S denotes being a smoker and L having lung cancer. Such or similar problems can easily be estimated using basic knowledge of probability theory, but assessing larger problems on the basis of intuition is hardly possible.

Bayesian belief nets (BBNs) are a computational tool to reason probabilistically. A BBN is a directed acyclic graph whose nodes represent random variables and whose arcs indicate a probabilistic influence from the *parent node* to the *child node*. It reflects all the possible states of the system in form of the joint probability[4]. The random variables can be discrete or continuous and their relation rests upon conditional probabilities.

Asia, an exemplary BBN first used by [Lauritzen and Spiegelhalter, 1988], is displayed in figure 3.3. It demonstrates the influences of smoking and traveling to Asia on a patient's lung conditions. Each of the eight variables in *Asia* has two states, yes or no, corresponding to the conditions of a patient, e.g. being a smoker or not.

The main use of BBNs is updating: Once new evidence on one or more variables is obtained, its effect can be propagated through the network using Bayes theorem. Evidence can be propagated both, forward and backward, which allows for predictive as well as diagnostic reasoning. Predictive reasoning updates the probability for a child node, based on new evidence of a parent node. For *Asia* the probability of having lung cancer can be recomputed, if it becomes known that the patient is a smoker, as in equation 3.1. The other way around, diagnostic reasoning updates the probability of a parent node, based on new evidence for a child node. The probability of having lung cancer can be recomputed, if a patient has an abnormal X-ray. For a more extensive introduction to BBNs the reader is referred to appendix A.

[4]The joint probability is the probability of events defined in terms of all random variables of the system. It entails the probabilities of all possible combinations of the states of the variables.

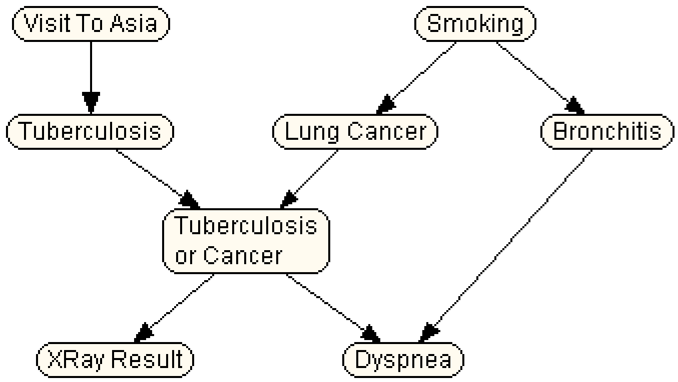


Figure 3.3: Nodes and arcs of *Asia* (source: [Lauritzen and Spiegelhalter, 1988])

3.2.2 Mathematical concepts of BBNs

The mathematical theory of discrete and nonparametric-continuous BBNs is summed up in this section. [Kurowicka and Cooke, 2006] contains most of the concepts presented here. If not, alternative references are given.

Bayesian Belief Nets are directed acyclic graphs. Figure 3.4 shows a simplistic BBN in which X influences Y. The joint distribution is specified by the marginal[5] distribution of X and the conditional distribution of Y given X as

$$f(x,y) = f(x)f(y|x). \tag{3.2}$$

Figure 3.4: A simplistic BBN

More generally, for n variables, the chain rule allows us to decompose the joint distribution using only conditional probabilities:

$$f(x_1, x_2, ..., x_3) = f(x_1)\prod_{i=2}^{n} f(x_i|x_1...x_{i-1}). \tag{3.3}$$

This expression is usually not very compact, since it requires specifying values of an n-dimensional function. Under conditional independence[6] assumptions however, the expression can be simplified. BBNs represent the joint probability more compactly by specifying a set of conditional independence assumptions in the form of a directed acyclic graph as well as a set of probability functions. The directed graph postulates that each variable is conditionally independent of all predecessors given its parents. Figure 3.5 shows a BBN on 4 variables where X_1, X_2 and X_3 are the parent set

[5] individual

[6] Consider three random variables X, Y and Z. We say that X and Y are *conditionally independent* given Z, if we know the value taken by Z, then no information given about Y would change our uncertainty about X [Bedford and Cooke, 2001]. This implies that $P(X|Y,Z) = P(X|Z)$. Note that it does not imply that X and Y are independent.

$pa(X_4)$ for node X_4. Node X_4 is a child for each of the parents. The conditional independence property allows us to simplify the conditional probability functions, so that we can write

$$f(x_i|x_1...x_{i-1}) = f(x_i|x_{pa(i)}) \tag{3.4}$$

for each i. If x_i does not have predecessors it is called a source node and $f(x_i|x_{pa(i)}) = f(x_i)$ [Hanea, 2008]. This yields

$$f(x_1, x_2, ..., x_3) = f(x_1) \prod_{i=2}^{n} f(x_i|x_{pa(i)}). \tag{3.5}$$

Note that we only have to specify functions with dimension not greater than the maximal number of parents of any node.

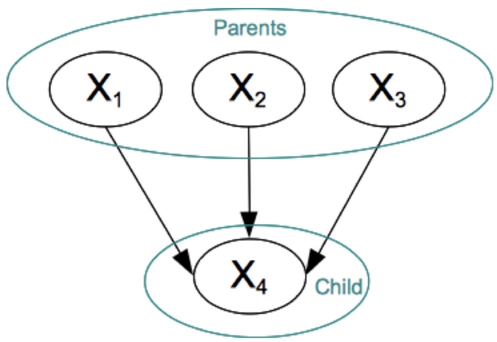


Figure 3.5: Parents and child of a BBN on 4 variables

Discrete BBNs

Discrete BBNs consist of discrete random variables. Each variable can be in one of a number of different states. The states should be exclusive (it is not possible for more than state at once to hold) and exhaustive (all possible states are specified) [Bedford and Cooke, 2001]. If we want to model continuous factors, they have to be discretized. The joint distribution is specified by the marginal distributions of the source nodes and the conditional distributions of all child nodes in the form of a conditional probability table (CPT). These distributions can be obtained from data, if available, or alternatively, from expert judgment.

If we consider the BBN from figure 3.5 to have discrete random variables the marginal distributions of X_1, X_2 and X_3 and the conditional distribution of X_4 given X_1, X_2 and X_3 have to be specified. Table 3.6 shows the CPT for variable X_4 assuming each node can take k values, denoted $x_i^j, i = 1, ..., 4, j = 1, ..., k$. The resultant CPT contains k^4 probability entries. For four binary variables, e.g. with values *yes* and *no*, 16 values need to be specified. For four variables with 10 states each 10000 values need to be found. This exponentially increasing assessment burden is one of the weakest points of discrete BBNs. If a model contains random variables of discrete nature, these have to be discretized. Considering the very high assessment burden, only a very limited number of discrete values can be allowed or the model has to be simplified drastically in order to make the quantification of the values feasible. A second serious shortcoming is the maintainability of discrete BBNs. They are very inflexible with respect to changes in modeling. For instance, if one parent node is added, all the previous quantification for the children of this node have to be redone.

Such disadvantages motivate the use of continuous BBNs for the management risk model. A promising type of continuous BBN, the non-parametric continuous BBN, is described in the following section.

As shown in the introductory example (appendix **??**), the main use of BBNs is to update distributions after observations. If some variables have been observed, we want to infer the probability distributions of the other variables, which have not yet been observed. Updating is based on Bayes Theorem and quite complex. An algorithm which allows fast updating for large BBNs has been proposed by [Lauritzen and Spiegelhalter, 1988].

Commercially available software that does all the calculations are, e.g. Netica or Hugin.

X_1	X_2	X_3	$P(X_4 = x_4^1 \mid X_1, X_2, X_3)$	$P(X_4 = x_4^2 \mid X_1, X_2, X_3)$	...	$P(X_4 = x_4^k \mid X_1, X_2, X_3)$
x_1^1	x_2^1	x_3^1	?	?	...	?
x_1^1	x_2^1	x_3^2	?	?	...	?
...	...	...	...	...	...	...
x_1^k	x_2^k	x_3^k	?	?	...	?

Figure 3.6: *Conditional Probability Table* for X_4 (source: [Hanea, 2008])

Non-parametric continuous BBNs

Shortcomings with regard to discretization, maintainability and flexibility give reasons to use non-parametric continuous Bayesian belief nets (NPBBNs) to model management risk. They have been introduced by [Kurowicka and Cooke, 2005] as an approach to continuous BBNs using *vines* together with *copulae*. Suffice to say here that vines are a graphical model closely related to BBNs [Hanea, 2008] and a copula is a joint distribution on the unit square with uniform margins [Bedford and Cooke, 2001] that realizes rank correlations. A freely available software to model NPBBNs is Uninet[7].

In a NPBBN nodes are associated with continuous univariate random variables and arcs with (conditional) rank correlations realized by a chosen copula. Based on the marginal distributions and the rank correlations the joint distributions can be built [Bedford and Cooke, 2001]. The rank correlation constitute the dependence structure of the system and need not be revised of the marginal distributions of one node is changed. Furthermore, if a node is added or removed, then the previously assessed (conditional) rank correlation need not be reassessed. This is a great advantage compared to discrete BBNs. The joint probability distribution can be obtained by sampling. A sampling algorithm using vines and copulae is described in [Kurowicka and Cooke, 2006]. It is a quite general, but comes at the price that these BBNs must be evaluated by Monte Carlo simulation. A second disadvantage is that updating the BBN once new evidence has become available, requires re-sampling of the whole structure.

3.2.3 Dynamic Bayesian Belief Nets

Systems that change over time can be modeled with dynamic Bayesian belief nets (DBBNs). DBBNs can be regarded as a special case of BBNs [Mihajlovic and Petkovic, 2001]. DBBNs consist of a sequence of static BBNs at different points or intervals in time, which are called time slices. They are interconnected with inter-slice arcs representing temporal relations. Figure 3.7 illustrates the principle of various slices. The number of slices corresponds to the number of time

[7] available from http://www.lighttwist.net/wp/uninet

steps. In theory, arbitrarily small and many time steps can modeled. As DBBNs are a special case of BBNs, they can be discrete or non-parametric continuous.

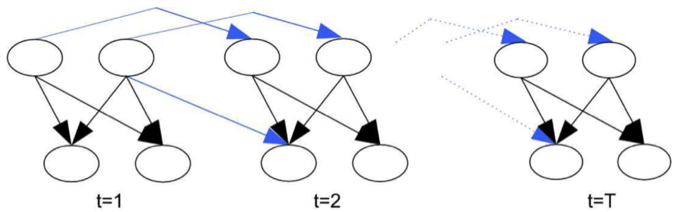


Figure 3.7: The Time Slice Approach for DBBNs (source: [Mihajlovic and Petkovic, 2001])

The inter-slice arcs can be time-invariant or change over time. The same applies to the structure of the time-slices. Usually both, arcs and slices, are assumed to be time-invariant, which simplifies the modeling. An example of a DBBN with dynamic structure changes has been applied to reservoir modeling [Gheorghe, 2010].

The Markov property can be used to simplify models that are time-invariant. For instance, if the first oder Markov property ($P(X_{t+1}|X_t, X_{t-1}) = P(X_{t+1}|X_t)$) is used, the BBN can be defined by unrolling two time-slices until we have T time-slices [Murphy, 2002] using

$$P(Z_{1:T}) = \prod_{t=1}^{T}\prod_{i=1}^{N} P(Z_t^i|Pa(Z_t^i)). \tag{3.6}$$

A model can of course contain Markov properties of different order, which represent different time dependencies of different factors. Generally, the number of slices needed is $order + 1$.

3.3 System Dynamics

System dynamics (SD) is a methodology to facilitate learning in a highly dynamic and complex environment by means of simulation models. A very famous SD model is "the limits to growth", which has been published by the Club of Rome in 1972 [Meadows et al., 1972]. The methodology is motivated by *systems thinking*: the ability to see the world as a complex system in which everything is interconnected. The interactions between various system components, most importantly feedback structures, are often more influential than the components themselves. They create the system dynamics. However, multiple feedback processes have an effect on system behavior that cannot reliably be foreknown and may even be counterintuitive. SD models simulate system behavior over time and, in this way, allow for speed-learning and testing in a "management laboratory" [Forrester, 1961, p.vii]: Managers are able to develop a deeper understanding for the system and to experience long-term (side-)effects of their decisions before taking real actions.

SD is grounded on control theory and applies its principles to socio-technical systems. It represents a system in terms of stock and flow structures, feedback loops and time delays. This representation results in a system of nonlinear differential equations which is solved numerically.

A simple example are *eroding safety goals* as shown in figure 3.8, which is a special case of the archetype *eroding goals* introduced by [Senge, 1990] In this structure, a short-term solution is found by sacrificing long-term goals. The safety gap stems from the difference between the actual safety and the safety goal. Figure 3.8 illustrates that the ambition to reduce the gap has two implications, which together have the negative effect on safety. On one hand, a large gap inspires efforts to improve safety, which leads (over time) to higher safety. On the other hand, the larger gap, the larger is the pressure to adjust the goals to reduce the gap resulting in lower safety goals. In combination this causes perpetually *eroding safety goals* and, consequently, decreasing safety.

To compute how the safety goal and the actual safety evolve over time, the causal loop diagram can be transformed into a stock-flow-diagram and simulated.

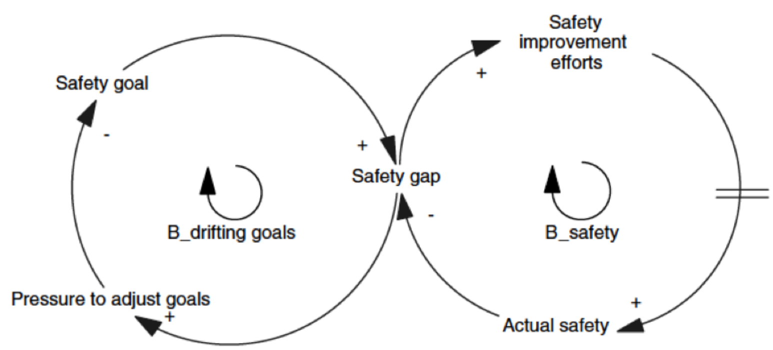


Figure 3.8: Example of a conceptual SD model of Eroding Safety Goals (source: [Marais et al., 2006]

3.4 Chapter Conclusions

This chapter considered methods used in risk modeling, in particular for models that include human and organizational factors. Event and fault trees are the tools commonly used to model failure of technical components. However, they have severe limitations to include uncertainty as well as common cause failures and are alone not suitable to model human performance.

Bayesian belief nets (BBNs) are a type of probabilistic model. They are a very powerful tool when dealing with uncertainties and when information is limited. Moreover, it is possible to convert event and fault trees into BBNs. In principle there is no theoretical reason why Bayesian belief nets should be restricted to static applications. However, not many dynamic applications exist and is it not entirely clear yet how they can be quantified in an efficient way.

System dynamics is methodology to simulate the dynamic behavior of a system. It is a powerful method represent complex systems and feedback loops. As it is based on differential equations, it is a deterministic approach. If sufficient data to quantify the equations is not available, it can have strong impact on the accuracy, and eventually on the explanatory and predictive power, of a risk model.

The following chapter begins with a description of two models that have combined the techniques and their respective advantages in an attempt to model human performance.

Chapter 4

Challenges of Human Factor Modeling

The starting point of this project was to investigate which impact human and organizational factors have on safety. As depicted in the causal diagram (figure 4.1) human factors (HF) are the elements that influence the human error probability (HEP). It can be supposed that there is an intricate system of causal relations that link the human factors with each other and to the error probability. For example, *stress* affects *fatigue* and vice versa and both increase the error probability. Management actions (MA) are means to influence human factors. It is relevant to monitor their effects to assess whether they have been successful or not. Of interest could be for instance, how much *safety awareness* can be increased within three month or at which point it falls below an acceptable level if no actions are taken.

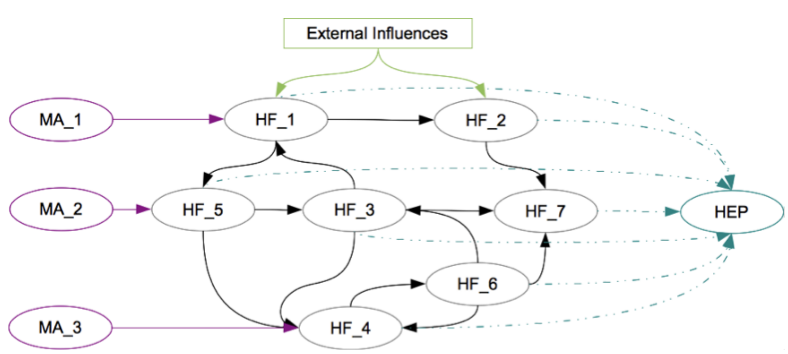


Figure 4.1: Generic scheme illustrating the relations between management actions (MA), external influences, human factors (HF) and human error probability (HEP)

The precedent chapter explained how event and fault trees are used to analyze and model technical failure of the system, but did not reveal how human performance can be linked to the technical components. The first section of this chapter describes two projects, a socio-technical risk analysis (SoTeRiA) and a causal model for air transport safety (CATS), that aim for a wider risk analysis by integrating human and technical factors in one model.

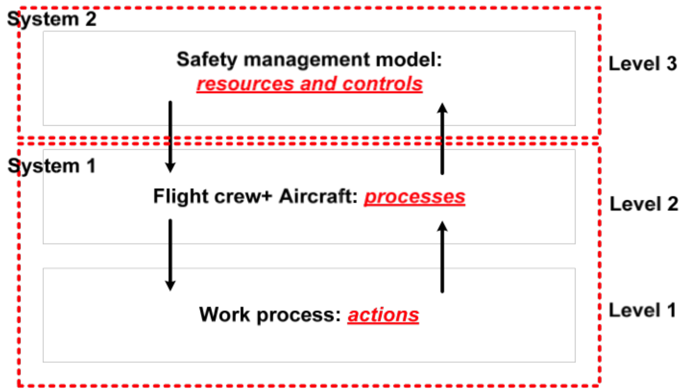


Figure 4.2: The hierarchical concept of CATS (source: [Lin, 2011])

4.1 Previous Attempts to Integrate Human Factors into Risk Models

4.1.1 Causal Model for Air Transport Safety (CATS)

As a high-hazard, low-risk system the aviation industry has developed high level defense system against any single failures, either human or technical [Amalberti, 2001]. The main threat are organizational errors, i.e. failures arising primarily at the managerial and organizational level and combining adversely with local triggering events as well as the active failures of individuals at the execution level [Reason, 1997]. According to [Ale et al., 2006] these kind of failures and interactions can be captured qualitatively by the causal model of air transport safety (CATS), because it takes into account a wider range of accidents, incidents and near misses instead of limiting the scope to catastrophic events and in particular crash analyses, which has been done in previous studies and is not necessarily sufficient to identify systemic problems. Moreover, large parts of the model have quantified and it can serve as a tool to analyze causal chains and quantify risks in air traffic [Ale et al., 2009a, Ale et al., 2009b, Lin, 2011].

Scope

CATS is a risk model for the flight process from gate to gate [Lin, 2011]. The output is accident probability per flight. It aims to explicitly model the first three levels of the socio-technical system involved in risk management as proposed by Rasmussen (cf. figure 2.2b) for the aviation sector. The hierarchical model of CATS is depicted in figure 4.2. Level 1 comprises the observable actions that are executed by the flight crew and the aircraft. Level 2 describes the (hidden) internal processes, the cognitive processes of humans and the internal functioning of hardware, leading to the actions at level 1. Finally, level 3 is the safety management model, which controls the individual factors and internal processes of level 2 in order to ensure that the actions at level 1 meet the objectives set by the management.

General Structure

In order to overcome major limitations of event trees (ETs) and fault trees (FTs), as described in section 3.1 p. 17, CATS combines these methods with Bayesian belief networks (BBNs) to a homogeneous mathematical model [Ale et al., 2009b].

The basic model structure is shown in figure 4.3. As explained earlier, the pivotal events in the ET split where a process can take different directions, faulty or not. In the combined model, the probability that a certain path is followed is determined by a FT which feeds into the pivotal event. The FTs comprise technical failures as well as human errors. The human errors, in turn, are modeled in human performance models (HPMs), represented as BBNs, and are influenced by management models.

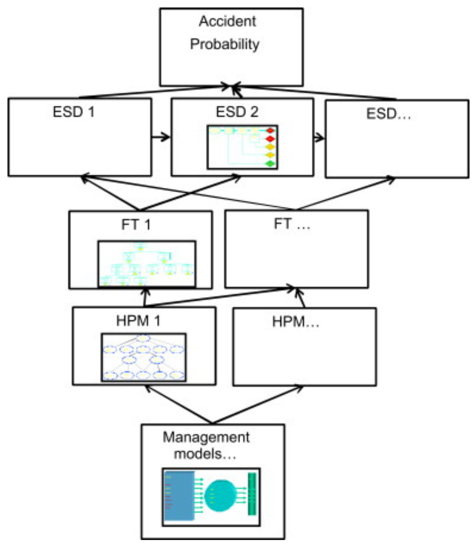


Figure 4.3: Basic model structure (source: [Ale et al., 2009b])

Eventually, the ETs and FTs are converted into belief nets so that the whole structure is a single integrated BBN. A major advantage of this step is that distributions of values rather than point estimates, if appropriate. Further, it removes the artificial transfer points between ETs, FTs and BBNs. It also allows for convenient and consistent handling of dependencies throughout the model. Nevertheless, it is necessary to quantify the parts of the models separately before integrating them into a BBN [Ale et al., 2009b].

Human Performance Models

Flight crew members, maintenance staff and air traffic controllers/managers play an essential role at the execution level of risk-bearing activities in the gate to gate flight process. Hence, there are three human performance models, one for each type of employees. The performance shaping factors have been selected after a literature review, analyses of accidents and incidents, and consultation with experts. Factors that are considered to have a significant influence on the probability of human error and can be quantified are included in the model. Examples are training, fatigue, experience or procedures. The HPM of the flight crew is depicted in figure 4.4 and the definitions given in table 4.1.

Up to the present, the human factors are assumed to be independent, but the BBN methodology allows to introduce dependencies at a later stage, if deemed appropriate. The HPMs are linked to several instances within CATS respecting the role of human error as common cause to multiple events.

Table 4.1: Operationalization of the performance shaping factors for the flight crew in CATS (source: [Lin, 2011])

Node #	Relevant variables	Objectively quantifiable units
1	First Officer Experience	Total number of hours flown by the First Officer
2	First Officer Training	The number of days since the last type recurrent training for the First Officer
3	Fatigue	Stanford Sleeping Scale; where 1 signifies "feeling active and vital; wide awake" and 7 stands for "almost in reverie; sleep onset soon; losing struggle to remain awake"
4	Captain Training	The number of days since the last type recurrent training for the Captain
5	Captain Experience	Total number of hours flown by the Captain
6	Captain Suitability	Likelihood that the Captain fails a proficiency check
7	First Officer Suitability	Likelihood that the First Officer fails a proficiency check
8	Weather	Rainfall rate in mm/hr
9	Intra-cockpit communication	Number of flights in which the pilot and first officer have a different mother tongue
10	Crew Suitability	Likelihood the Captain or the First Officer fails a proficiency check
11	Man-machine interface	Four aircraft generations; 1, 2, 3, or 4
12	Workload	Number of times the crew members have to refer to the abnormal/ emergency procedures
13	Flight Crew Error	Likelihood that the flight crew makes an unrecovered error that is potentially hazardous for the safety of the flight

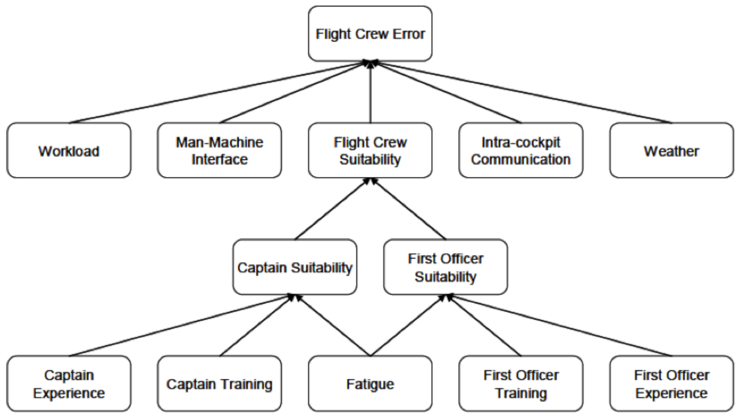


Figure 4.4: Human performance model structure for the flight crew in CATS (source: [Ale et al., 2009b])

Management Models

CATS uses the concept of safety barriers to formulate the link between technical, human and organizational factors. Barriers block the paths between cause and consequence (figure 4.5). The task of the management is to deliver criteria and resources in order to keep the safety barriers intact. Causes for barrier failure are either due to humans and their behavior, technology or external factors like weather and are modeled as base events of one of the fault trees. The criteria and resources are categorized into eight so-called *delivery systems*: (1) *Competence*, (2) *suitability*, (3) *manpower planning and availability*, (4) *workload*, (5) *procedures*, (6) *communication and coordination*, (7) *man-machine interface* and (8) *commitment to safety*. For instance, *competence* refers to "knowledge and skills learned through training and experience" and *suitability* covers both physical as well as mental suitability including factors such as drugs and alcohol, fatigue or general health [Lin, 2011].

The management system consisting of the eight delivery systems is a conceptual model that helps to select PSFs and to design management actions. Selected PSFs that have an influence on the human error probability are directly included in the HPMs. For the model of the crew (figure 4.4), these are experience, workload, training, fatigue, intra-cockpit communication and technology interface. In order to investigate the effect different management actions have on the accident probability, the nodes can be conditioned. For instance, *training*, which is defined as "the number of days since the last type recurrent training" (cf. table 4.1) can be modified in practice and correspondingly in the model. The model is then updated and the influence of this action on the accident probability can be seen.

4.1.2 Socio-Technical Risk Analysis (SoTeRiA)

SoTeRiA, standing for Socio-Technical Risk Analysis, is like CATS a model based on a mixed logic approach. In [Mohaghegh et al., 2009], the authors put forth a methodology for selecting appropriate modeling techniques and their integration in the form of a hybrid approach. The motivation is akin to the CATS project: Since managerial and organizational failures are commonly cited as the root causes of major accidents, there is a need to include these into risk modeling frameworks. SoTeRiA is the proposed framework combining System Dynamics (SD), Bayesian

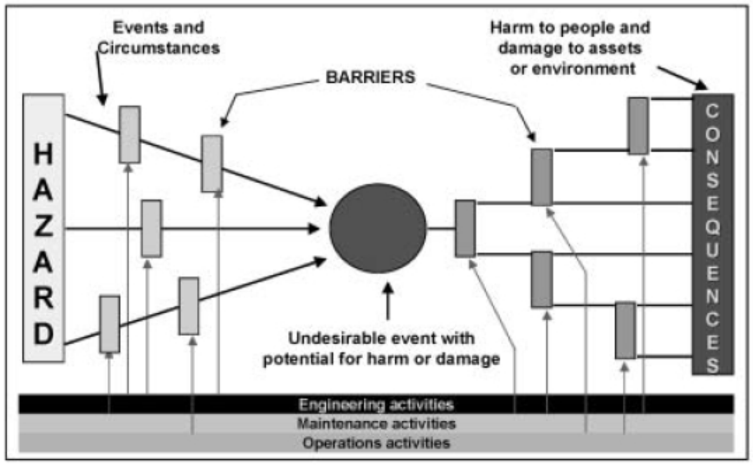


Figure 2 *The Bow Tie (courtesy Shell International Exploration & Production).*

Figure 4.5: Bow tie model illustrating the concept of safety barriers (source: [Lin, 2011])

belief nets (BBNs), event trees (ETs) and fault trees (FTs) are combined. An example on airline maintenance systems intends to show the feasibility and value of this hybrid technique, in particular to analyze the dynamic effects of organizational factors on system risk.

Scope and conceptual framework

The idea behind the conceptual framework (figure 4.6) is comparable to the hierarchical concept of CATS (cf. figure 4.2 and 4.3). The model starts with the *system risk*, where the accident scenarios are development and quantified, on the right side and is built moving leftwards illustrating how organizational root causes propagate. *Safety critical performances (SCPs)* are comparable to level 1 in the CATS hierarchy, the work processes, and *organizational structure & practices* resembles level 3, the *safety management model*. Additionally SoTeRiA includes psychological terms, such organizational climate and group climate, in its general framework and some of them are matchable with level 2, the (hidden) internal processes.

Unfortunately, the corresponding publications [Mohaghegh-Ahmadabadi, 2007, Mohaghegh et al., 2009, Mohaghegh and Mosleh, 2009] do neither reveal the elements within the individual building blocks nor the relationships between them in detail. For this reason, the model components are not directly comparable to the ones of CATS [Lin, 2011]. Nevertheless, general overview of the different modules and what they contain is given in the following:

- The *system risk* model consists of ETs and FTs generating all possible risk scenarios from basic elements.
- *Safety critical performances* (SCP) are human performances (of a group) that directly affect the basic elements in the FTs, e.g. maintenance performance influencing hardware failure. They are the input of the system risk model and the output of the *unit process model*.
- The *unit process model* serves to link different underlying processes with the safety critical tasks of the SCP module. resources, procedures and individual performances are combined group performances, which determine the action, i.e. the SCPs. It has three direct influences:

(1) It depends on *organizational safety practices*, e.g. calibration and test activities affect the resources or procedures may be altered.
(2) It is affected by the *regulatory environment*, which imposes external auditing on organizational procedures.
(3) It is influenced by individual *performance shaping factors* (PSFs). Examples for PSFs are an employee's ability, knowledge or opportunity.

- The direct effect from the individual PSFs is also an indirect effect from the organizational safety structure and practices. There are three influence paths from *organizational practices* to individual *PSFs*:
  (1) There is a direct path to *PSF*s. For instance, training directly affects employee's knowledge.
  (2) There is an indirect influence to *PSFs* through the *organizational safety climate* and the *group safety climate*.[1] Both affect the *psychological safety climate*, which is part of the PSFs. It reflects the employee's believe in their managers' commitment to safety and his/her motivation for the activities in the unit process model.
  (3) There is an indirect influence to *PSFs* through *emerging processes*. Emerging processes include the social interaction process, leadership and supervision as well as the homogeneity in the organization. They determine the "strength" of the shared climate.

- *Organizational safety structures and practices* on the other hand are influenced by *safety Culture*[2], which is part of the *organizational culture* and has influence on the managerial decisions regarding safety. It is also affected by the *regulatory environment*, since it imposes policies and rules on safety practices.

- Organizational culture is influenced by three higher level factors, *industrial & business environment*, *social and political culture and climate* and *organizational vision, strategy and goals*. Moreover it is affected by three feedback effects:
  (1) There is a direct feedback from the *financial outcome*, e.g. imposing financial stress.
  (2) There is an indirect feedback from the *financial outcome* through the *industrial & business environment*
  (3) There is a direct effect from the *system risk*.

- The *financial outcome* also has a feedback effect on the *system risk* and *individual PSFs*, e.g. collapse of morale in the face of possible bankruptcy.

- *System risk* has a feedback effect on the *financial outcome*. For instance, internal costs will increase after an accident.

The second scheme in figure 4.7 clarifies the hybrid character of SoTeRiA. As stated above, *system risk* module is built of event and fault trees. The *unit process model* is a BBN. The two modules are connected by the SCPs. The bottom layers, i.e. *individual PSFs* and various organizational factors, are implemented in a System Dynamics (SD) environment.

[1] "Climate is the perception of "what happens" in the organization and can be described as temporary attributes of an organization." [Mohaghegh et al., 2009] This is a very vague definition...
[2] "Culture is [...] stable, and is related to the employees' ideologies, assumptions, and values" [Mohaghegh et al., 2009]

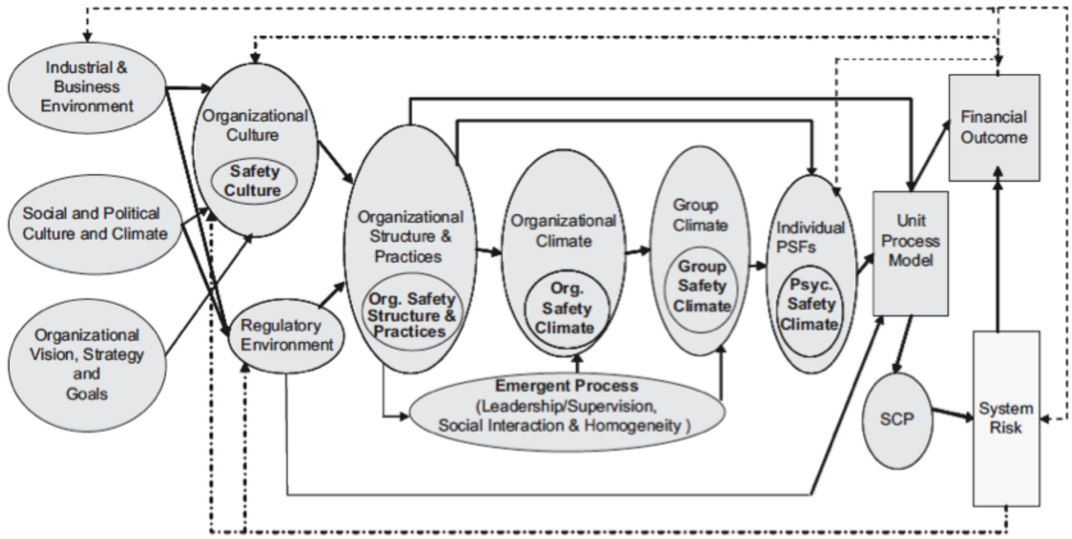


Figure 4.6: Schematic Representation of SoTeRia #1: Overview (source: [Mohaghegh et al., 2009])

Individual Performing Shaping Factors (PSFs) Module

The individual human performance model is limited to those factors which are predominantly influenced by organizational factors. *Motivation*, *ability* and *opportunity* are deemed most relevant and included in the model (cf. figure 4.7 framed module labeled as *Individual-level PSFs*).

- *Motivation* is affected by the so-called *psychological safety climate*, a measure for the individual's perception of safety practices, which in turn is determined by an individual value and the group safety climate.
- *Ability* refers to *physical ability* as well as to *knowledge*.
- *Opportunity* is defined as either *temporal opportunity* or the lack of it, e.g. due to time pressure, or *physical opportunity*, e.g. sufficient light at the workplace.

These three human factors have affect the safety critical actions, which can manifest themselves as violations or as errors, a distinction introduced by Reason [Reason, 1990] (cf. chapter 2), and can be influenced by management modules.

For the final version of SoTeRiA, not all the factors are selected. Table 4.2 summarizes the included factors together with their operationalization. The corresponding SD model "human reliability" is included appendix B.

Management Models

The *organizational structure & practices* module resembles the safety management system of CATS, in that it includes all organizational practices that influence resources, procedures and human actions having a direct impact on safety critical performances. The organizational practices are classified into four groups, as depicted in figure 4.8: human related activities, procedure related activities and resource related activities, all of which are supported by common activities.

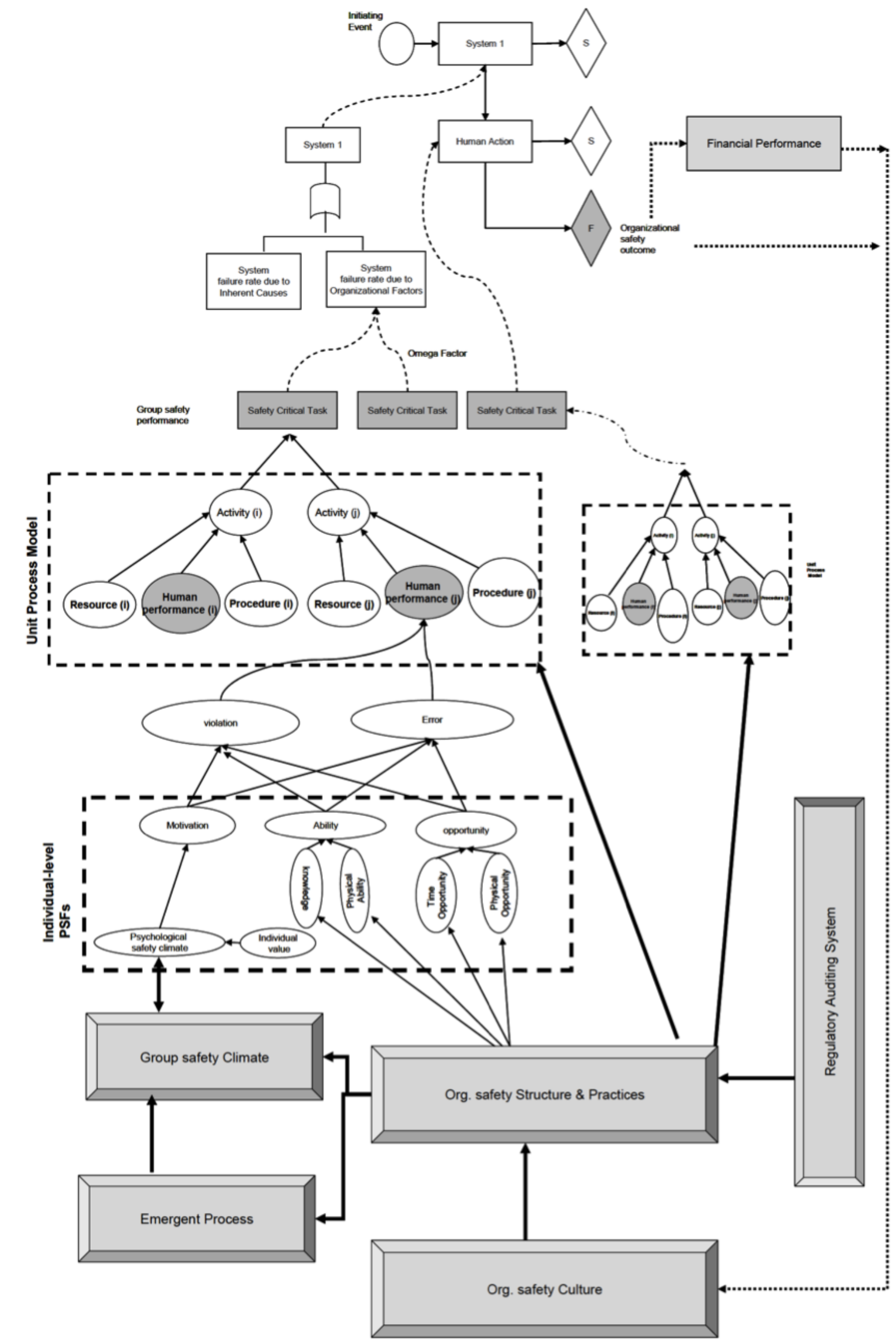


Figure 4.7: Schematic Representation of SoTeRiA #2: hybrid character (source: [Mohaghegh et al., 2009])

Table 4.2: Selected Performance Shaping Factors for a Maintenance Technician in SoTeRiA

Variable	Operational Definition
motivation	morale
psychological climate	-
individual value	-
ability	-
knowledge	level of experience
physical ability	-
opportunity	-
time opportunity	time pressure
physical opportunity	-

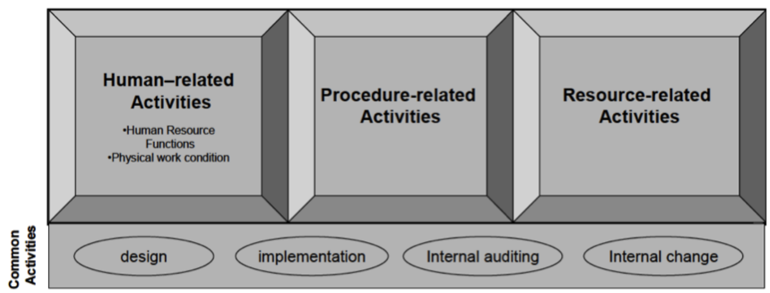


Figure 4.8: Organizational safety practices in SoTeRiA. Y stands for a direct relation, N for no relation and I for an indirect relation, e.g. through (source: [Mohaghegh et al., 2009])

Table 4.3 shows the human related activities designed to influence the individual human performances. However, the relationships are left on this abstract level and the authors point to future research [Mohaghegh et al., 2009]. Moreover, it is indicated that various other modules within SoTeRiA have an impact on the PSFs.

The quantitative example published in [Mohaghegh-Ahmadabadi, 2007, Mohaghegh et al., 2009] is a demonstration case and consists of a strongly simplified structure. All choices are made for illustration purposes and not justified otherwise. For instance, *management commitment* represents *safety culture*. *Training* and *hiring* are the only *organizational safety practices* that have an influence on the commitment. The corresponding system dynamics model can be found in appendix B.

4.1.3 Interim Conclusions

CATS is a realistic model that incorporated selected human factors and management actions to quantify the accident probability of a flight. It combined all modeling methodologies into a (static) nonparametric BBN, which allows for convenient and consistent handling of dependencies throughout the model. Dynamic relationships however cannot be represented. Recent mathematical developments (theoretically) allow for dynamic Bayesian belief nets. Therefore it is, in principle, possible to extent the approach used for CATS and to include dynamic relationships.

SoTeRiA, on the other hand, can be regarded as a proof of concept. In a mixed logic approach,

Table 4.3: Human related activities and their links to individual performance shaping factors (source: [Mohaghegh-Ahmadabadi, 2007])

	knowledge	Physical Ability	Time Opportunity	Physical Opportunity
Selectivity in Recruiting/Hiring (SE)	Y	Y	N	N
Internal Staffing (ST)	Y	N	N	N
Training (TR)	Y	N	N	N
Appraisal (AP)	I	I	N	N
Reward System (RE)	I	I	I	N
Job Analysis (JA)	N	N	N	N
Team Systems(TS)	Y	Y	Y	N
Employee Assistance (EA)	N	N	N	N
Due process (DP)	N	N	N	N
Employee Voice/ Empowerment (EM)	N	N	I	I
Diversity (DI)	Y	N	N	N
Legal Compliance (LC)	N	N	N	Y
Safety (SA)	Y	N	N	Y
union Relations (UR)	N	N	Y	Y
Job Enrichment (JE)	Y	Y	Y	N
Contingent Workforce (CW)	N	N	Y	N
Physical Work Condition (PWC)	N	N	N	Y

dynamics and feedback effects are incorporated through SD modules. However, no method is proposed to quantify the equations that deal with human factors and there is no grounds for using th SD methodology in this thesis.

4.2 Classification of Human Factors

The term performance shaping factor (PSF), which is equivalent to human factor (HF) or also performance influencing factor (PIF), has been introduced as "any factor that influences human performance" by [Swain and Guttmann, 1983]. They differentiated between internal and external factors. Internal factors refer to the characteristics or state of an individual, such as knowledge, experience, stress or fatigue. External factors are external to the individual, e.g. training, procedures or man-machine interface. Performance shaping factors are an integral part of many human reliability analysis (HRA) techniques to describe the human-system interaction [Groth and Mosleh, 2012]. Up to now, however, no standardized rules exist which govern the definition and usage of PSFs. They often served as a "catch-all for explaining less-than adequate human performance" [Lin, 2011, p. 46] mixing contextual and organizational factors. Consequently, the factors are not interpreted consistently across different HRA studies. A first attempt to systematize existing definitions of PSF has been made by [Groth and Mosleh, 2012]. Table 4.4 shows the proposed hierarchical set of PSFs, which organizes them in five categories: organization-based factors, team-based-factors, person-based factors, situation/stressor-based factors and machine-based factors.

This list of HFs is more comprehensive than the list used in CATS, a causal model for air transport safety [Lin, 2011]. Moreover, for every HF in CATS an equivalent can be found in table

Table 4.4: Human factor classification proposed by [Groth and Mosleh, 2012] for use in HRA data collection and causal models

Organization-based	Team-based	Person-based	Situation/stressor-based	Machine-based
Training program	Communication	Attention	External environment	Human-system interface
Availability	Availability	To task	Conditioning events	Input
Quality	Quality	To surroundings	Task load	Output
Corrective action program	Direct supervision	Physical & psychological abilities	Time load	System response
Availability	Leadership	Alertness	Other loads	
Quality	Team coordination	Fatigue	Non-task	
Other programs	Team cohesion	Impairment	Passive information	
Availability	Role awareness	Sensory limits	Task complexity	
Quality		Physical attributes	Cognitive	
Safety culture		Other	Execution	
Management activities		Knowledge/experience	Stress	
Staffing		Skills	Perceived situation	
Scheduling		Bias	Severity	
Workplace adequacy		Familiarity with situation	Urgency	
Resources		Morale/motivation/attitude	Perceived decision	
Procedures			Responsibility	
Availability			Impact	
Quality			Personal	
Tools			Plant	
Availability			Society	
Quality				
Necessary information				
Availability				
Quality				

4.4. Nevertheless, we encounter difficulties when trying to "use" this classification in order to build a causal model, especially when trying to quantify them as variables in a nonparametric BBN.

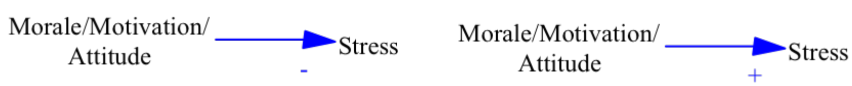
4.3 Problems in Defining Relationships between Human Factors

Next to the classification in table 4.4 [Groth and Mosleh, 2012] propose definitions for all human factors listed. However, problems arise when attempting to operationalize the factors based on those definitions. Consider, for instance, the definition of the factor *morale/motivation/attitude*, which is given by the authors as follows:

> Morale, motivation and attitude (MMA) together refer to style, temperament, personality and intrinsic human variability [...]. These characteristics manifest as willingness to complete tasks, the amount of effort a person devotes to tasks, and the state of mind of the worker. [...] Morale, motivation and attitude can be affected by external factors such as organizational culture, teamwork, and resources, but each person will internalize these factors differently leading to varying MMA even among team members.
>
> [...] Since it is extremely difficult to measure attitude, especially in retrospective analysis it is necessary to include specific work practice behaviors as metrics of [morale/motivation/]attitude. The behaviors identified during review of the data are *problem solving style, information use, prioritization*, and *compliance*.

From this definition it is clear that it is difficult to used the factor *morale/motivation/attitude* as a variable in a causal model.[3] On one hand it is extremely wide, which brings along the danger of wrong causal statements when the causal arcs/arrows refer to different aspects of the factor.

For example, high *morale/motivation/attitude* could refer to someone enjoying his job and being highly motivated to achieve a certain goal. Such a person may feel less stress than one that does not feel joy while working, which leads to the causal diagram in figure 4.9a. Alternatively, high *morale/motivation/attitude* could refer to a feeling of responsibility to complete a task well. In this case, the more someone feels responsible the more stressed he/she is, as depicted in figure 4.9b. The two interpretations lead to opposite causal relations of the same factors. The issue complicates, if one considers both interpretations to be two individual aspects of the same factor that do not overlap (someone enjoying his work may or may not feel responsibility just as someone feeling responsible may or may not enjoy his work). What is the relation of *morale/motivation/attitude* to *stress* then? There are no limits to one's creativity in finding an endless number of aspects for each factor. One could also imagine high *morale/motivation/attitude* includes the will to achieve a certain goal as well as the ability to not get discouraged when confronted with hardships. However, an individual could be both at once: principally very eager to attain a certain goal, but also disheartened from facing a seeming dead-end in his/her approach. This individual would somehow have a high as well as a low *morale/motivation/attitude* at the same time and it is unclear how this can be captured in one variable, let alone how to model causes and effects. Even when trying to make qualitative diagrams as in figure 4.9 difficulties arise.



(a) Negative relation (b) Positive relation

Figure 4.9: Possible causal relations between *morale/motivation/attitude* and *stress*

Of course, this is strongly related to the lack of a metric for the factor, which makes it extremely difficult to operationalize. Observable behaviors, such as *problem solving style*, have to be used as proxies, even though they capture only a very limited part of the factor. This issue arises for most if not all of the factors in table 4.4. Proxies have also been used in the CATS project, e.g. *training* has been approximated as *days since last training* and *intra-cockpit communication* has been quantified as *number of flights in which pilot and first officer have a different mother tongue* [Lin, 2011]. Unfortunately, there are countless possible proxies for each human factor and for each application different choices are made. The aspects of a factor that are deemed most relevant for a particular application and/or are measurable serve as justification for the choices made.

4.4 Chapter Conclusions

This chapter began by reviewing two models from the literature that attempt to incorporate human and organizational factors: a causal model for air transport safety (CATS) and a socio-technical

[3] One might object that *morale/motivation/attitude* is obviously a broad definition, as it includes morale, motivation and attitude in one factor, but most of the listed factors have various facets. Another example is *stress*, which can take the form of pressure, conflict, frustration or uncertainty.

risk analysis (SoTeRiA). Both models combine various modeling techniques: event and fault tree analysis, Bayesian belief networks and System Dynamics.

Whereas CATS is a static, highly specific, quantified model that incorporated a few selected human factors and management influences, SoTeRiA is more of a proof of concept emphasizing the need to include dynamics, in particular feedback, when modeling human and organizational factors. Unfortunately, SoTeRiA does not propose any method to quantify the equations in the System Dynamics modules that deal with human factors.

However, there is not yet a "recipe" how to realistically quantify human factors and their dynamics. As a first step towards finding such a recipe, this chapter reviews how human factors can be classified and discusses the problems that occur when defining relationships between them. It was concluded that it is hardly possible to define causal relations without operationalizing the HFs first. Nonetheless, it is very desirable to identify feedback loops and other nonlinear structures. Intuition says they exists. For instance, *stress* in the form of frustration may decrease the *morale/motivation/attitude* making any of the two diagrams in figure 4.9 a loop. Another hypothetical example is the relation between *workload* and *experience*. High *workload* increases *experience* and in turn high *experience* decreases the time needed to complete tasks and thus the *workload*. Such and more complicated dynamic structures strongly shape the systems behavior over time. They are a good starting point when attempting to design management actions, in particular because they may cause surprising or even unpredictable behavior including negative side-effects of well intended policies.

A meaningful set of human factors and their relations can only be created on the basis of a specific example which justifies the choices made. That is, appropriate operationalizations and influences of HFs can only be selected for a well demarcated system. Still, a real application in high hazard industries requires a wide literature search including accident reports and human reliability data bases as well as consultation with experts, which is beyond the scope of this thesis.

For that reason, this project settles for creating a demonstration model, which represents the qualitative and quantitative relationship between training and knowledge over time. The intention is to put forth dynamic Bayesian belief nets as a suitable method to realistically quantify human factors and their dynamics.

Chapter 5

Demonstration Model Based on a Gaming Experiment

The previous chapter argued for the need of a simple example case that can be used to meaningfully operationalize and quantify human factors (HFs) as well as their mutual influences. For now it has been chosen to consider only the two factors *training* and *knowledge*. According to [Groth and Mosleh, 2012] training "refers to the knowledge and experience imparted" and knowledge (as well as experience) is the accumulated information that has been converted from training. Note that knowledge differs from training in that the information retained from the exact same training will differ among the individuals. It follows that there is a positive causal influence from *training* to *knowledge*. Moreover, *knowledge* is a stock variable implying that the current knowledge depends on the previous knowledge just as the future one depends on the current one. This context is represented as a causal diagram in figure 5.1. Training and knowledge at different time steps are regarded as individual variables, similar to an unrolled dynamic Bayesian belief net.

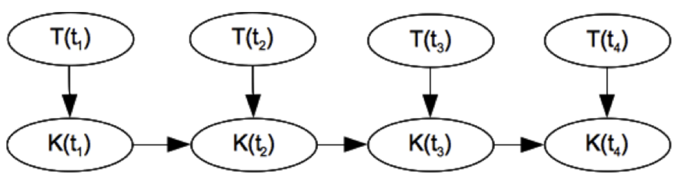


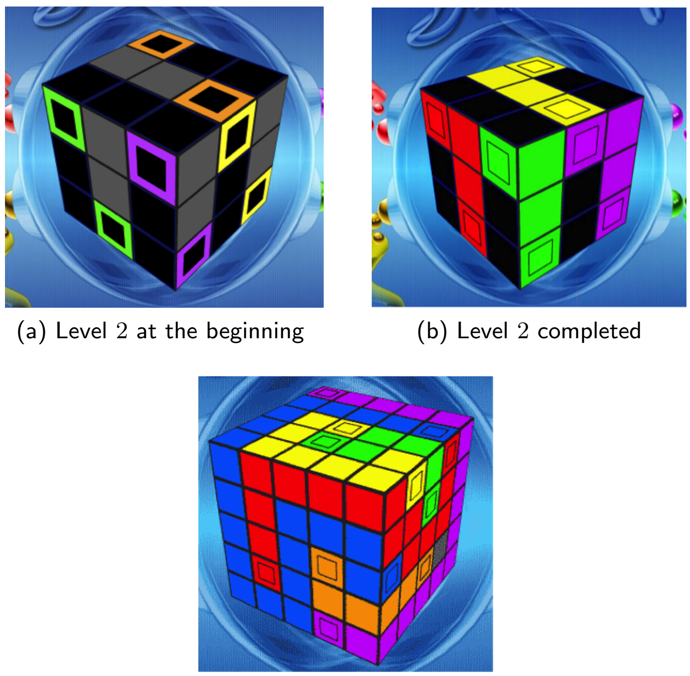
Figure 5.1: Dynamic BBN showing the effect of training on knowledge over four time steps t_1, t_2, t_3 and t_4, where T and K denote training and knowledge, respectively.

The next steps are (1) to find a demarcated system in which most notably the behavior represented by this diagram occurs and (2) to gather data for quantification. The following section describes why the chosen approach is to conduct an experiment, for which participants play the online puzzling game *3Dlogic*.

5.1 Motivation for the Online Puzzling Game *3Dlogic*

In reality not only training has influence on knowledge, but also e.g. intellectual capacity, concentration or interest. A system having a justifiable boundary, which only includes the structure depicted in figure 5.1, is gaming, in particular the online puzzling game *3Dlogic*. As it is self-explanatory, like most online games, the process of playing can be interpreted as training, whereas the knowledge that has been acquired can be determined from the time needed to complete a certain number of levels. Thus, the measurement of the knowledge gained in the first round, is carried out with the training of the second round. Detailed reasoning behind the operationalization of these factors is given in section 5.2.

Before discussing the advantages of *3Dlogic* for this experiment compared to other online puzzling games, its rules are described briefly: The players have to rotate the magic cube (figure 5.2a) and connect the fields with the square inside of the same color. If all color pairs are linked the cube is completed (figure 5.2b). Cells cannot be connected diagonally and black cells are blocked [Mypuzzle.org, nd]. Figure 5.2c gives an idea of how the difficulty in the game increases with higher levels.



(a) Level 2 at the beginning

(b) Level 2 completed

(c) Level 16 completed

Figure 5.2: A few examples of *3Dlogic* (source: [Mypuzzle.org, nd])

An important aspect of *3Dlogic* is that the initial conditions in each level and the solution paths are always the same; the game has no randomness. Playing a specific level for the first time is equivalent to being trained and the gamer retains some information about the solution. The second time the gamer plays that level he/she can apply this knowledge and possibly complete the level faster than the first time. Moreover, he/she receives the same training a second time and may retain more information, which enables him/her to pass the level even faster and so on. This is for instance not the case when playing the famous game *Tetris* [Tetris Holding, 2011], where the building blocks appear in a random order. Each time the gamer plays a particular level, he/she needs to use a different approach, receives slightly different information and consequently a slightly different training. First, this makes a repetition of the experiment under identical conditions impossible, which is a crucial factor to guarantee reliability of the experiment's results, and second, it is difficult, to measure the amount of training the player has received. Further, the gamer cannot make irreversible mistakes: if he/she accidentally makes a "wrong move", e.g. by clicking the wrong button, which is not necessary connected to his knowledge, but to his motor skills, he/she is able to correct it. Finally, there are no time limits in the levels of *3Dlogic*. These last two points argue that this game is indeed a system, in which *training* and *knowledge* are the variables of primary interest and that factors such as motor skills or stress due to time pressure are minimized as far as possible. However, spatial sense and other cognitive factors inherent to the participants cannot be eliminated and will influence the knowledge they acquire. This effect can be reduced by selecting participants with relatively homogeneous educational background such as graduate students. Under this condition, it can even be desired: crew members operating a high hazard technology will have similar differences and it is realistic to include this diversity in a risk model.

If the participants of the experiment play four rounds of *3Dlogic*, data of their *training* and *knowledge* at four consecutive time steps is obtained. Recall that a nonparametric dynamic BBN encodes the joint probability of all variables by specifying their marginal distributions and the (conditional) rank correlation of each arc, which are six and five, respectively, for the BBN in figure 5.1. With the software Uninet these can be learned directly from the data gathered in the experiment.

5.2 Operationalization of the Variables

The terms *training* and *knowledge* were defined in the introductory paragraph of this chapter on page 39. In the following their operationalizations for the experiment are explained and motivated.

Training

In most games, and this applies to *3dlogic* as well, the level defines the difficulty of the game and indicates the progress of the player [Zichermann and Cunningham, 2011]. Each level has a difficulty that needs to be overcome, or in other words, a trick has to be discovered in order to complete the level. This implies that a player who has finished a level has been imparted this trick, which is a piece of knowledge. According to the definition above he has received a specific training. For this reasons the *number of levels completed* is taken as a measurement for *training*.

However, the difficulty of levels does not increase linearly. Most games have a curvilinear relationship between difficulty and level as shown in 5.3. Consequently, training is measured in intervals which are not equally spaced. This has to be taken into account when operationalizing the variable *knowledge*.

In reality training can have various aspects, such as content, frequency or form of the training and not all crew members undergo the same training. During official training session they may receive the same quality and type of training, but they also receive personal feedback or advice from their supervisors while working on their individual tasks (which also differ). This suggests that training is random variable (which is what we want for the BBN anyways). As we have no information about the distribution of training, we assume it to be uniform between level 5 and level 10. Level 5 is chosen, because the first four levels are very easy as they introduce the player to the principles of the game and take very little time to complete. Level 10 is chosen in order to keep the duration of the experiment concise.

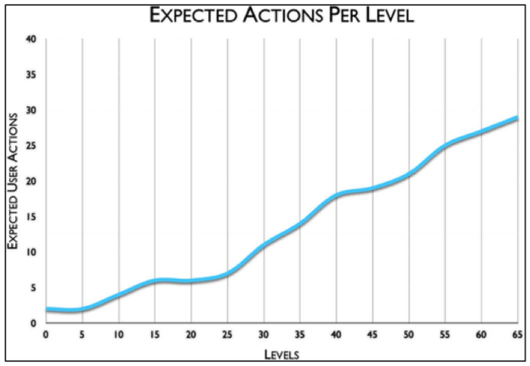


Figure 5.3: An example showing that level progression is not linear or exponential (source: [Zichermann and Cunningham, 2011])

Knowledge

A measure for the "amount" of knowledge a participant has, the time needed to complete a given number of levels is relevant. The idea behind can be illustrated with help of three possible cases.

1. A participant knows the solution, because he received this information in previous training, to a level and is thus able to complete it very fast. Actually, if participants really know the solution (and do not even have to think about it for a moment), their times to completion should be equal and fixed: There exists a minimum time to completion for each level. One could argue that different possible solutions (if existent) could take different times to complete, but in *3Dlogic* the participants have to run their mouses over each square of the cube once no matter which solution they apply. How exactly they choose the paths of color will not make a significant difference in the time it takes to color the entire cube. Of course different participants may have different skills using a mouse leading to minor differences.

2. A participant does not (fully) remember the solution. In this case he will take longer to complete a level than in case 1. The time needed for completion will depend on how much he is able to recall from previous training and how fast he learns while completing the level, i.e. current training.

3. A participant does not know the solution yet. In this case he will take longest to complete a level. His time to completion depends on his ability to convert information into knowledge while completing the level, i.e. current training.

Since all participants complete different numbers of levels, the total time to completion cannot be a measure for *knowledge*. Dividing the time to completion by the number of levels completed is not an option, because the difficulty of levels does not increase linearly as explained earlier. A possibility is to consider the minimum possible time to completion and normalize for each participant:

$$K_i = \frac{t_{n,min}}{t_{i,n}} \tag{5.1}$$

where K_i denotes the knowledge of the participant in round i, t_i, n is the time he/she needed to complete levels 1 to n in round i and $t_{min,n}$ is the minimum possible time to complete levels 1 to n. In this way the *knowledge* of participants is comparable also if they play up to different levels. However, this definition is not global. It only indicates how "well" the participant know level n, knowing even more levels is not reflected by the variable. If a participant has been trained thoroughly on level 5 he might have knowledge 1, if he retains all the information taught to him/her. However, if he/she is subjected to a higher level in the subsequent round, his knowledge, according to this definition, may decrease again. In other words, the knowledge can not be viewed independently of the training. It is not a stock variable that accumulates. This accumulation has to be taken into account by considering all the trainings received. This is also discussed together with the results in section 5.4. Still, this is only an initial effect. After many rounds, it is very likely that most participants have played and memorized all levels.

5.3 The Experiment

5.3.1 Participants

Almost all participants are drawn from the Delft University of Technology student body, mainly from the Faculty of Technology, Policy and Management, but also from the Faculty of Architecture and the Faculty Electrical Engineering, Mathematics and Computer Science. This ensures a relatively homogeneous educational background and comparable initial abilities to solve the puzzle. Because technical staff in engineering companies have the same kind of homogeneity, this reflects the real life situation. All master students and recent graduates who were interested in participating in the experiment were selected for the study. In total 24 students participated, 11 female and 13 male. An anonymous list of the participants and their affiliation is given in Appendix C.

5.3.2 Materials and Procedure

At the beginning of the experiment, participants are informed of the topic of this thesis as well as the purpose of the experiment. A short introduction before the experiment makes the students familiar with the purpose of the experiment and in particular with the content of sections 5.1 and 5.2. Then, each student is given a handout (see Appendix C) and is assigned a university computer. Groups of three students sit in one row, which facilitates the assistance and supervision during the experiment. The hand out includes once more the instructions for the game, part one of a questionnaire and procedure guidelines for the experiment. Part one of the questionnaire is designed to identify the gender as well as the affiliation of the student, if he/she knows *3Dlogic* already and whether performing "well" during the experiment is important to him/her. Part two is be handed out after the experiment and asks about the perceived difficulty of *3Dlogic*, stress or pressure while playing and performance compared to other students. In case some participants seem to underperform, the questions about gender and evaluation of the experiment may give insight. Participants indicating that performing well is important to them, may be threatened if all other participants seem to be finishing faster than them, a factor which can undermine their intellectual performance [Aronson et al., 1999]. There might also be a stereotype threat for women of having less spatial abilities than men (e.g. [McGlone and Aronson, 2006]).

The general sequence of the experiment is as follows. The students play four rounds of the online game *3Dlogic* till the level they have been randomly assigned to for each round and stop the required time to completion themselves, which enables a single experimenter to supervise many students at the same time. The game is played on http://mypuzzle.org/3d-logic and the online stopwatch from www.online-stopwatch.com is used. Each round consists of the following two steps:

1. From a bowl the participant draws one of the numbers between 5 and 10, which marks the level he/she has to complete.
2. The participant plays *3Dlogic* up to and including the assigned level, but **not** further, and measures the required time. He/she leaves the game open and the stopwatch on pause for the experimenter to verify the level and time measurement.

5.4 Results and Discussion

Table 5.1 shows the results of the experiment. The first column indicates the participant. The second, third and fourth column contain the completed level, the required time and the knowledge in the first round, the fifth, sixth and seventh column contain the completed level, the required time and the knowledge in the second round etc. In the following $T1$ denotes the training received in the first round, i.e. the level that has been completed, $t1$ denotes the needed time and $K1$ the knowledge. Likewise defined are $T2$, $t2$, $K2$, and so on. Hence, for the first participant $T1 = 10$, $t1 = 268$ and $K1 = 0.3$. The "knowledge" columns are calculated according to formula 5.1 with the best times achieved by the experimenter (see also table 5.2), e.g. for the first participant $K1 = \frac{81}{268} = 0.30$

In each round equally many participants have been randomly assigned to the same level. For 24 participants and 6 levels this means that each level is played by four participants. The last row shows the average over all participants. Of course, the average level completed is 7.5 for each round. The required time decreases from 299 seconds ($4{:}59$ minutes) to 149 seconds ($2{:}29$ minutes) in the second round to 131 seconds ($2{:}11$ minutes) in the third round and to 114 seconds ($1{:}54$ minutes) in the last round. These results are also plotted in figure 5.4, where *o* denotes the data points and *x* the average of each round. The knowledge advances from a group average of 0.26 to 0.56. The knowledge increase is largest from the first to the second round, then it increases relatively little. It is remarkable that the standard deviation of the times decreases, whereas the standard deviation of the knowledges increases with the rounds.

The results of th questionnaire are included in appendix C. Since no participant indicated to feel extreme stress or pressure and also no one considered himself/herself much worse than the other participants, it is assumed that these factors do not have significant impact on the results. For this reason these aspects have not been further investigated for this thesis.

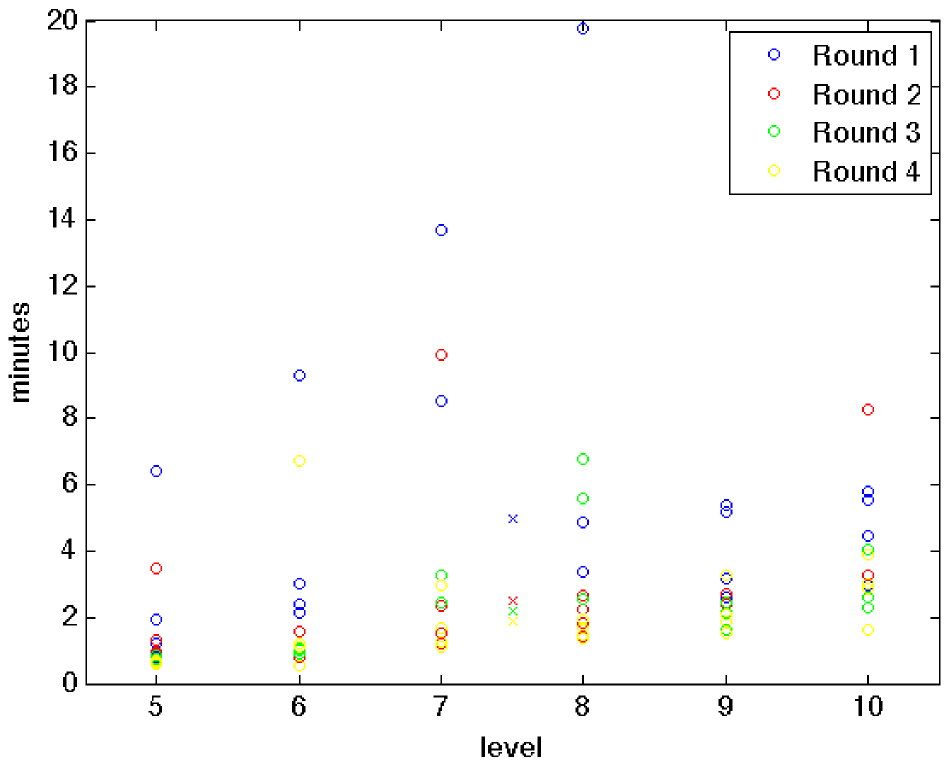
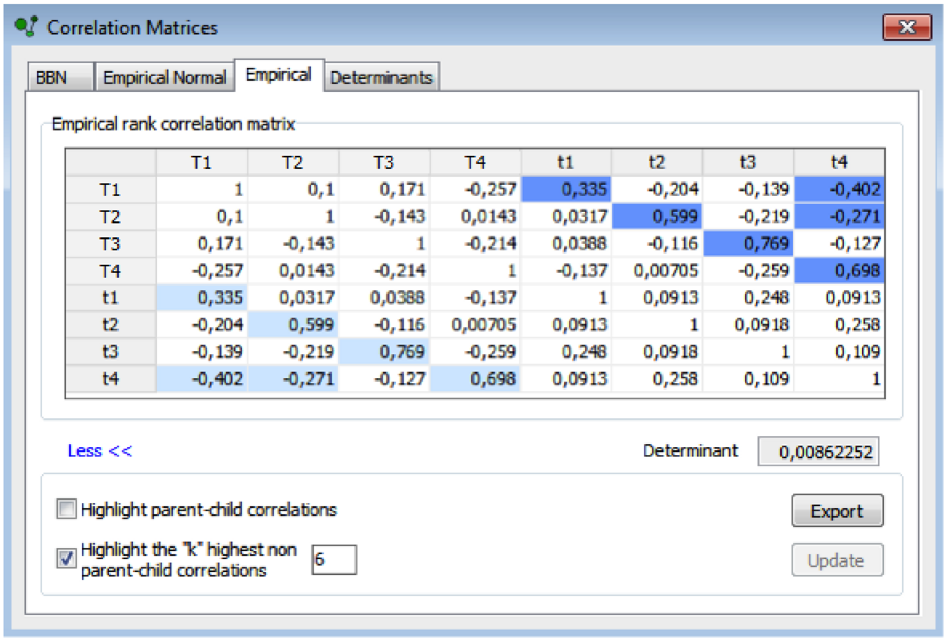


Figure 5.4: Plot of levels against the required times in minutes

Table 5.1: Results of the Experiment. The color coding illustrates that, per round, each level is played exactly four times.

	Round 1			Round 2			Round 3			Round 4		
Participant	**Level**	**Time [s]**	**Knowledge**	**Level**	**Time [s]**	**Knowledge**	**Level**	**Time [s]**	**Knowledge**	**Level**	**Time [s]**	**Knowledge**
1	10	268	0.30	6	59	0.56	10	242	0.33	8	99	0.55
2	9	190	0.35	7	74	0.61	5	36	0.81	9	90	0.73
3	9	324	0.20	7	91	0.49	6	59	0.56	10	173	0.47
4	8	1186	0.05	8	159	0.34	7	101	0.45	5	40	0.73
5	6	129	0.26	6	93	0.35	7	147	0.31	6	74	0.45
6	5	73	0.40	10	138	0.59	7	67	0.67	10	96	0.84
7	10	178	0.46	6	48	0.69	8	81	0.67	6	33	1.00
8	7	99	0.45	8	84	0.64	9	96	0.69	5	35	0.83
9	9	156	0.42	6	49	0.67	9	97	0.68	7	75	0.60
10	6	143	0.23	5	208	0.14	8	337	0.16	6	404	0.08
11	9	312	0.21	10	172	0.47	10	138	0.59	8	87	0.62
12	7	179	0.25	5	56	0.52	10	155	0.52	9	113	0.58
13	8	293	0.18	8	134	0.40	9	148	0.45	7	102	0.44
14	8	135	0.40	9	124	0.53	5	44	0.66	8	82	0.66
15	5	386	0.08	5	54	0.54	8	154	0.35	9	125	0.53
16	5	49	0.59	9	126	0.52	5	44	0.66	7	67	0.67
17	10	348	0.23	9	161	0.41	6	65	0.51	6	68	0.49
18	10	334	0.24	10	195	0.42	9	128	0.52	5	39	0.74
19	7	511	0.09	8	110	0.49	8	406	0.13	5	41	0.71
20	8	204	0.26	9	143	0.46	6	55	0.60	10	182	0.45
21	7	822	0.05	5	80	0.36	5	50	0.58	7	179	0.25
22	5	117	0.25	7	596	0.08	6	64	0.52	10	233	0.35
23	6	559	0.06	10	497	0.16	7	195	0.23	9	196	0.34
24	6	182	0.18	7	140	0.32	10	243	0.33	8	116	0.47
Mean	7.5	299	0.26	7.5	149	0.45	7.5	131	0.50	7.5	114	0.56
Standard deviation	1.74	258	0.14	1.74	131	0.16	1.74	95	0.18	1.74	82	0.21



	T1	T2	T3	T4	t1	t2	t3	t4
T1	1	0,1	0,171	-0,257	0,335	-0,204	-0,139	-0,402
T2	0,1	1	-0,143	0,0143	0,0317	0,599	-0,219	-0,271
T3	0,171	-0,143	1	-0,214	0,0388	-0,116	0,769	-0,127
T4	-0,257	0,0143	-0,214	1	-0,137	0,00705	-0,259	0,698
t1	0,335	0,0317	0,0388	-0,137	1	0,0913	0,248	0,0913
t2	-0,204	0,599	-0,116	0,00705	0,0913	1	0,0918	0,258
t3	-0,139	-0,219	0,769	-0,259	0,248	0,0918	1	0,109
t4	-0,402	-0,271	-0,127	0,698	0,0913	0,258	0,109	1

Figure 5.5: Correlation matrix obtained from Uninet for the data from table 5.1. T refers to the levels played and t, to the time needed

Looking at the correlation matrix in figure 5.5 for the data in table 5.1, we see that, in general, the highest correlation is between the *training* and the *required time* of the same time step. This leads to the expected conclusion that for each round, the level which has been assigned has the greatest influence on the time that has been played. An exception is $T1$, which has a higher correlation with $t4$. The high correlation between $T1$ and $t4$ might be sample specific and simply "bad luck", because the sample size is very small. The reason that the correlation between $T1$ and $t1$ is with 0.333 by more than factor 1.5 smaller than the correlations between the other *training-required time* pairs, may be that their is a great influence of an initial knowledge, for which no measurement has been taken. The correlation between the T and the t of the same round is positive indicating that playing to a higher level takes more time than playing to a lower level. The correlation between T of the current round and the t of future rounds is negative indicating a tendency to complete faster in future rounds if the current level was high. This is reasonable, because a higher level allows the participant to obtain a larger amount of information, which he/she can use int he following rounds. We also observe that the two time vectors of consecutive time-steps have rather small correlations and it is difficult to identify a pattern, which is reasonable since the strongest influence on the time is a randomly assigned level. As discussed earlier in section 5.2, a meaningful interpretation for knowledge can be found when normalizing the required times according to equation 5.1.

Table 5.2 compares the best times achieved by the participants with the one of the experimenter. The best times of the experimenter are regarded as the shortest possible, when knowing the initial set up of the color pairs in the magic cube and solution by heart. It has been measured as follows: The experimenter starts the game, takes time to look at the cube until she has memorized the solution. Then she starts the stopwatch, executes the solution and pauses it after the game has transitioned to the next level. (The game takes almost 2 seconds to transition to the following level.) Again, she takes time to investigate the cube and continues the stop watch for the execution and the transition time. Following this approach the experiment is at least equally fast as the fastest participant. There two reasonable explanations for this fact: (1) it takes time to recognize the cube (design-related) or (2) it takes more time to remember complex solutions. Take for example level 7 (figure 5.6). Reason (1) refers to the fact that the participant recognizes the cube, e.g. realizing that it is the cube for which one "had to go all the way around", whereas reason (2) refers to remembering how exactly this "going all the way around" looked like. This effect is probably strong, because only few rounds are played and the participants hardly have a change to learn the solutions by heart. However, it could also be plain "bad luck", because the sample size of 24 is very small. The fact that the solutions to level 5 and 6 are still rather obvious and "motorically easy" could explain why this is not observed for level 5 and 6.

Figure 5.6: Solution to level 7 (source: [Free Web Archade, nd])

It can be observed that the time differences (of the experimenter as well as the participants) between the various levels is not constant. This can be explained by the fact that the operations, which have to be executed, for the levels are of differently complex (cf. figure 5.3). On one hand the number of squares that have to be colored varies and other the other hand making few long labyrinthine path takes more time than connecting many straight lines, because it requires more caution.

Table 5.2: Best times achieved by the participants and by the experimenter

Level	5	6	7	8	9	10
Best time of the participants	35s	33s	67s	81s	90s	96s
Best time of the experimenter	29s	33s	45s	54s	66s	81s

Table 5.2 leaves us with to choices for $t_{n,min}$ in equation 5.1 to compute the knowledge. We could use the best times of the experimenter, which can realistically be considered the absolute minimum possible time. Or we could use the best times of the participants on the grounds that it is the minimum possible time under the circumstances of the experiment, in particular the few rounds that have been played. We choose the first, because of the risk that the small sample size causes that some levels are never played by the fastest participants. Moreover, it reflects that the participants are able to reach almost knowledge 1 (as in 100 % knowledge) about the lower levels (e.g. participant 8 in round 4), whereas this was not possible for higher levels in the given time frame of four rounds.

Nonetheless, *knowledge* based on participants' best times is used as a comparison to give some insight on the sensitivity of this choice for the normalization parameter. It is not possible to execute an automated sensitivity analysis in Uninet, which computes the correlation matrices, and a full blown sensitivity analysis is therefore beyond scope of this thesis.

The resulting empirical correlation matrices, one by normalizing with the experimenter's times and one by normalizing with the participant's times, are given in figure 5.7. $T1$ to $T4$ refer to the training in the respective rounds, i.e. the assigned level and $K1$ to $K4$ refer to the knowledge measured in each round. In contrast to the matrix in figure 5.5 the largest correlations are between the various knowledge variables, which is expected, as modeling *knowledge* over time was the main purpose of the experiment. A few entries differ notably in the two correlation matrices in figure 5.7b and 5.7a, but both show an akin prevalent structure. The following paragraphs discuss the entries of the matrix in figure 5.7b, as this is the matrix used for the model in chapter 6.

- The correlations between the different knowledge vectors ($K1$, $K2$, $K3$, $K4$) differ in value by up to 0.2. However, a pattern is not recognizable. There are all positive, which shows that the knowledge has a tendency to increase over the rounds. (There may be, however, participants whose knowledge "decreased", e.g. participant 11. This may be because, he/she simply did not understand the game or it could also happen, as for participant 1 whose knowledge varies with the difficulty of the levels he is assigned to.)

- The different training vectors are not uncorrelated, even though they have been designed to be independent. From the first four rows of column one, we can see for example that participants who had been assigned a high level in the first round $T1$, have been somewhat likelier to be assigned a high level again, instead of again a low on, in the second round 2. The converse is true for participants who had been assigned a low level in the first round. This is similar for $T1$ and $T3$, but with a slightly higher correlation, whereas the opposite is true for $T1$ and $T4$. Still, none of the correlations between the training is alarmingly high and they may be neglected for the model (see section 6.2).

- $T1$ is positively correlated with $K1$, $K2$, $K3$ and $K4$. For the correlation with $K1$, which is clearly lower than with the other vectors, no explanation is found. The other correlations may be consequences of the limitations of the *knowledge* definition: As stated above, it includes only how well the participant knows all levels up to the assigned one. How well he knows higher levels, is not reflected by this variable. As a consequence, the knowledge variable is influenced by training from previous rounds. Consider two participants, one having $T1 = 5$ and $T2 = 7$ and the other one having $T1 = T2 = 7$. If the first one completes $T1$ in 60 seconds, and the second one in 94, then both have a knowledge of $K1 = 0.48$. Then in $T2$, the second participant has an advantage, because he has done levels 6 and 7 before. Hence, the knowledge $K2$ depends on $T1$ as well. The positive correlation says, that people that played up to a higher level in one round tend to show a higher knowledge in the subsequent one.

  The equivalent could be argued for the correlations between $T2$ and $K3$ as well as $K4$. It could be that the negligibly small correlation between $T3$ and $K4$ is due to the fact that most of the variance is already explained by $T1$, $T2$ and $T4$ itself.

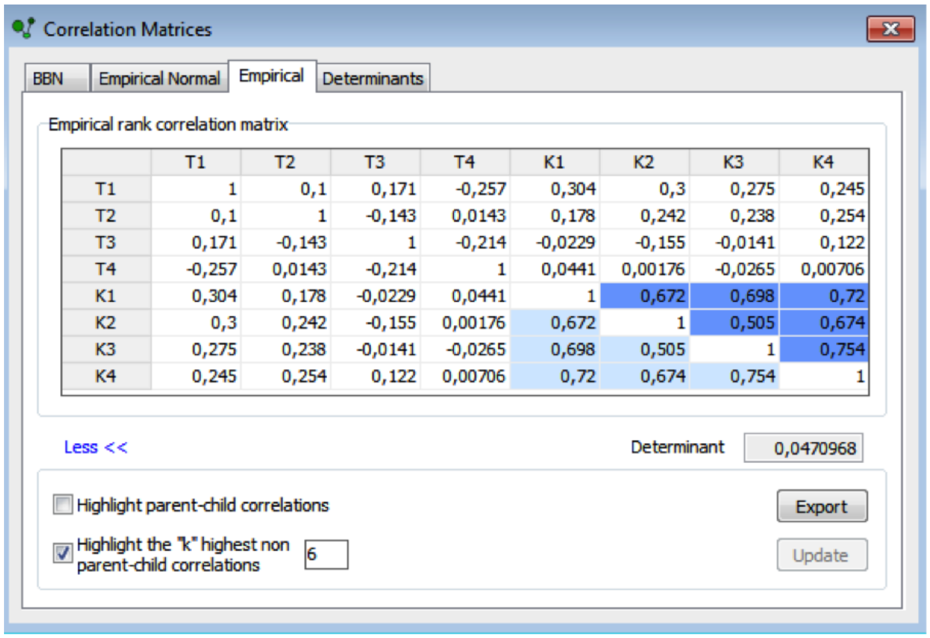
- The correlations are negative between $T2$ and $K2$ (although negligible), $T3$ and $K3$ as well as $T4$ and $K4$. This indicates a tendency of participants that are assigned low T values to show relatively high K values of the same round. Two possible contributing factors are: (1) The lower the level, the higher is the probability that the participant has played it in a previous rounds and has the knowledge to solve it. (2) Finding a solution for the first time takes longer the a higher level is, because the difficulty increases. The time depends on his/her cognitive ability with respect to this specific problem type. (If all solutions had been presented before the experiment these factors would not play a role. It was deemed unlikely, though, that participants are able to follow, let alone remember, solutions to problems they did not experience yet by themselves and consequently decided against it.)

- Finally, the T variables have very low and, hence negligible, correlations with the K variables of the previous rounds. (Exceptions are the correlation between $T4$ and $K3$ as well as between $T3$ and $K2$ for which no explanation can be thought of.) Obvious reason is, that future training cannot influence current knowledge. Neither does knowledge has an influence on training (by design).

As a question remains to what extent the correlations between the T variables influences other correlations. This is one of the reasons why the conclusions of the discussion above are limited. However, in the following chapter it is shown that the order of magnitude of these correlation is not significant with regard to the sample size (cf. section 6.2). A second reason are the couple of entries which did not fit with the explanation. There are two possibilities. Either they are indeed outliers due to the small sample or the correlations in figure 5.7b are much more random and too much meaning has been read into them. The third reason is, that much of this reasoning cannot be affirmed based on the matrix in figure 5.7a.

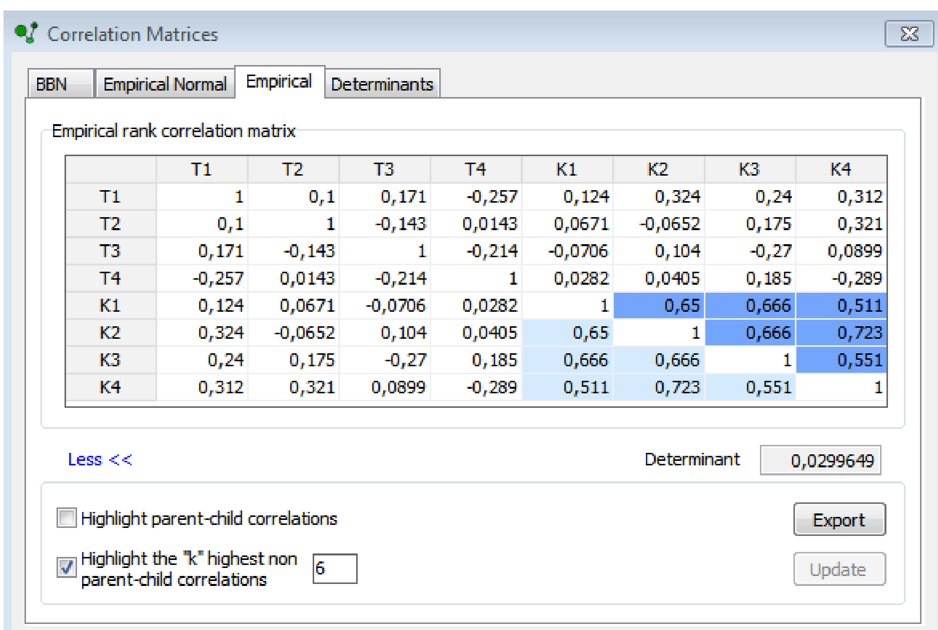
The correlations between the T variables are logically identical. Also the highest correlations are between the K variables, without identifiable pattern. However, the correlations between the T and the K variables differ notably: $T1$ and $T2$ are correlation with all K variables, whereas $T4$ is almost not correlated with any of the $K4$.

One could argue that the two normalization values used for the matrices differ substantially, more than the 10% deviation one conventionally assumes in sensitivity analysis, but together with the unexplained entries, the any reasoning has to be regarded with suspicion.



	T1	T2	T3	T4	K1	K2	K3	K4
T1	1	0,1	0,171	-0,257	0,304	0,3	0,275	0,245
T2	0,1	1	-0,143	0,0143	0,178	0,242	0,238	0,254
T3	0,171	-0,143	1	-0,214	-0,0229	-0,155	-0,0141	0,122
T4	-0,257	0,0143	-0,214	1	0,0441	0,00176	-0,0265	0,00706
K1	0,304	0,178	-0,0229	0,0441	1	0,672	0,698	0,72
K2	0,3	0,242	-0,155	0,00176	0,672	1	0,505	0,674
K3	0,275	0,238	-0,0141	-0,0265	0,698	0,505	1	0,754
K4	0,245	0,254	0,122	0,00706	0,72	0,674	0,754	1

(a) Using the best times of the particpants to compute *knowledge*



	T1	T2	T3	T4	K1	K2	K3	K4
T1	1	0,1	0,171	-0,257	0,124	0,324	0,24	0,312
T2	0,1	1	-0,143	0,0143	0,0671	-0,0652	0,175	0,321
T3	0,171	-0,143	1	-0,214	-0,0706	0,104	-0,27	0,0899
T4	-0,257	0,0143	-0,214	1	0,0282	0,0405	0,185	-0,289
K1	0,124	0,0671	-0,0706	0,0282	1	0,65	0,666	0,511
K2	0,324	-0,0652	0,104	0,0405	0,65	1	0,666	0,723
K3	0,24	0,175	-0,27	0,185	0,666	0,666	1	0,551
K4	0,312	0,321	0,0899	-0,289	0,511	0,723	0,551	1

(b) Using the best times of the experimenter to compute *knowledge*

Figure 5.7: Correlation matrices obtained from Uninet for the two variables *training* (T), which is identical to the number of levels completed, and *knowledge* (K), which results from normalizing the required times to completion.

5.5 Chapter Conclusions

This chapter introduced an experiment to generate data for a *knowledge* and a *training* variable in consecutive time steps. Participants played four rounds the online puzzling game *3Dlogic*. The level they were assigned to complete was interpreted as *training*. A normalization of the required time to complete the level was interpreted as *knowledge*. More precisely, a "minimum possible time for completing a particular level" has been divided by the time a participant needed to complete that level so that the resulting *knowledge* is between 0 and 1. However, this definition is not global. It only indicates how well a participant knows all the levels up to the completed one. If he/she knows further levels, is not reflected. And the other way around, if he/she has been trained to a higher level in the previous than in the current round, the knowledge variable does not reflect how much information he/she retained from the higher levels. This is a limitation of the operational definition of knowledge and causes a correlation between all previous rounds of training and knowledge, a fact not been considered in figure 5.1 in the introduction of the chapter.

Furthermore, it has been investigated how sensitive the dependence structure between the variables is, based on what is taken as a "minimum possible time for completing a particular level". There are a few values which change notably, but the main result is that the *knowledge* variables are significantly higher correlation than *training* with *knowledge*. This indicates that knowledge can be modified by training, but not too abruptly: previous knowledge is a more determining factor. Moreover, the correlations between various variables have been discussed in more detail and (partial) explanations have been provided. However, not all correlations seem reasonable/realistic.

Interpreting the data has not been straight-forward and further limited due to the small sample size of 24 participants. Nevertheless, choices have to be made in order to build a model. On what grounds these choices can be made in order to obtain a model which is suitable to reflect the dependences between the variables and user friendly is discussed in the next chapter. The definition of knowledge is a relatively sensitive assumption, but has not been considered further for the demonstration model. This should be taken into account when attempting to model for more realistic applications. It also emphasizes the difficulty of defining human factor, which are not directly measurable.

Chapter 6

Building and Using the BBN

Specifying the structure of a BBN is a statistical learning process. Statistical learning theory provides a framework for gaining knowledge from a set of data and in particular for constructing a model from this set [Bousquet et al., 2004]. Learning and validating the BBN are carried out simultaneously, since the goal is of course to learn a valid model [Kurowicka, 2010].

This chapter follows [Hanea et al., 2010], who describe a method for mining ordinal multivariate data using nonparametric Bayesian belief nets (BBNs). Alternative references are given when used. Data mining is a computational process of extracting and analyzing information from large data bases [Han and Kamber, 2006]. Even though the amount of data gathered in this research is rather small the same techniques can be applied. However, this technique does not yield a unique model structure. Various structures may be valid and the "best" model is determined based on its potential value for the model user.

The structure of a BBN consists of arcs and nodes. As explained in section 3.2, directed arcs represent a probabilistic influence from *parent node* to *child node*. As in the previous chapter we denote the training of the first round with $T1$, of the second round with $T2$, etc. Analogously, $K1$, $K2$, $K3$ and $K4$ denote the knowledge of each of the four rounds. These 8 variables are the nodes of the BBN. In principle, two variables are connected with a directed arc, if (1) the statistical correlation computed from the data is high and (2) the influence of one variable on the other is meaningful. Not meaningful is, for instance, an influence from $K2$ to $K1$ because it is reversed in time. Also arcs directed from knowledge to training contradict the meaning of the variables. When choosing arcs based on their high correlation, it has to be taken into account that the sample size in this research is rather small, which affects the significance of the computed rank correlations.

The correlation matrices presented in the previous chapter (figure 5.7) show that, to some extent, all variables are correlated with each other. A model connecting each variable to all other variables, called a saturated graph, is most likely to fit the data. However, not all arcs are meaningful; they may represent sample jitter. Small influences are often not beneficial as well, because they distract from the factors which are most relevant to the model outcome. Based on the correlation matrix from figure 5.7b, it is not immediately obvious which arcs to include or not. For some entries, explanations have been found, but others do not seem logical (cf. section 5.4). Considering that the order of magnitude of both kind of entries is similar, it is difficult to justify why one or another correlation should be included in the model, or not. Especially because the sample size is very small, the correlations computed from the data have to be contemplated with caution.

From a user point of view, it is desirable to include as few arcs as possible in the model, because it enhances the transparency of the model. Unnecessary complexity is at the expense of clear representation and has no benefits. More generally, less arcs mean less quantification efforts. For the current research, this argument does not play a role yet, but it might do for subsequent research when extending the model and switching from experimental data to other sources. Especially when using expert judgment, many incoming arcs require the expert to make intricate estimates. The challenges of eliciting (conditional) rank correlations from experts are discussed in detail in [Morales et al., 2008].

Of course, statistical validity is a prerequisite for a useful model. Therefore, the first section presents the necessary statistical considerations and definitions. Section 6.2 describes the learning process taking both the magnitude of the rank correlation and meaningfulness of an arc between two variables into account. Because two valid model structures are found, section 6.3 investigates the effect of either choice on the model output. Also, some examples for model use are presented.

6.1 Statistical Considerations and Preliminaries

From a statistical point of view, it is crucial to validate the following two assumptions:

1. The joint normal copula adequately represents the multivariate data.

   Nonparametric BBNs are graphical models, which build the joint distribution for the data using the so-called joined normal copula (cf. section 3.2 page 21). This first step ensures that the assumption of the normal copula is valid with regard to the present data. It also implies that the saturated graph is a valid model.

2. The BBN with its conditional independence statements is an adequate model of the saturated graph.

   The saturated BBN corresponding to the data from the experiment contains 8 variables and 28 arcs and is depicted in figure 6.1a. But many of these influences are very small and may reflect sample jitter. Especially because the sample size is very small, the relations have to be contemplated with caution. Moreover, several arcs in figure 6.1a can by reason of the causal flow of time not be true, e.g. the arc from $T4$ to $K1$. It is not meaningful either to introduce an arc in the opposite direction, from $K1$ to $T4$. Further, the experiment has been designed such that $T1$, $T2$, $T3$ and $T4$ are independent and identically distributed. Again, the small sample size might be cause for the correlation found in the data.

   In a plausible model "unrealistic" as well as noisy influences are eliminated. Consequently, it has to be ensured that the resulting BBN with the introduced independence statements (that are the deleted arcs) adequately represents the saturated graph in the statistical sense.

These two assumptions can be regarded as null hypotheses are can be validated with a two-sided statistical test: H_0= *assumption is true* and H_1 = *assumption is not true*. To test these hypotheses, the concept "*assumption is true/not true*" needs to be operationalized so that it can be measured. A suitable measure in this case is the determinant of the rank correlation matrix [Hanea, 2008], which attains the maximum value 1 if all variables are uncorrelated and the minimum value 0 if the dependence between the variables is linear. Three different determinants are needed in the validation process:

- Determinant of the empirical rank correlation matrix (DER)

This determinant reflects the dependence structure of the data. It shows to what extend the variables in the sample (not the population) are correlated with each other.

- Determinant of the rank correlation matrix resulting from a transformation of the univariate distributions to standard normals, and then using Pearson's transformation to obtain rank correlations from the product moment correlations (DNR)

  When using the joint-normal copula to build the joint distribution, the data are transformed. This determinant reflects the dependence structure of the transformed data. Generally, it differs from the DER, because the empirical copula need not be normal. If it is "similar" to the DER, the first assumption is validated: The joint normal copula adequately reflects the correlations of the original data.

- Determinant of the rank correlation matrix of the model, the final nonparametric BBN, which uses the normal copula (DMR[1])

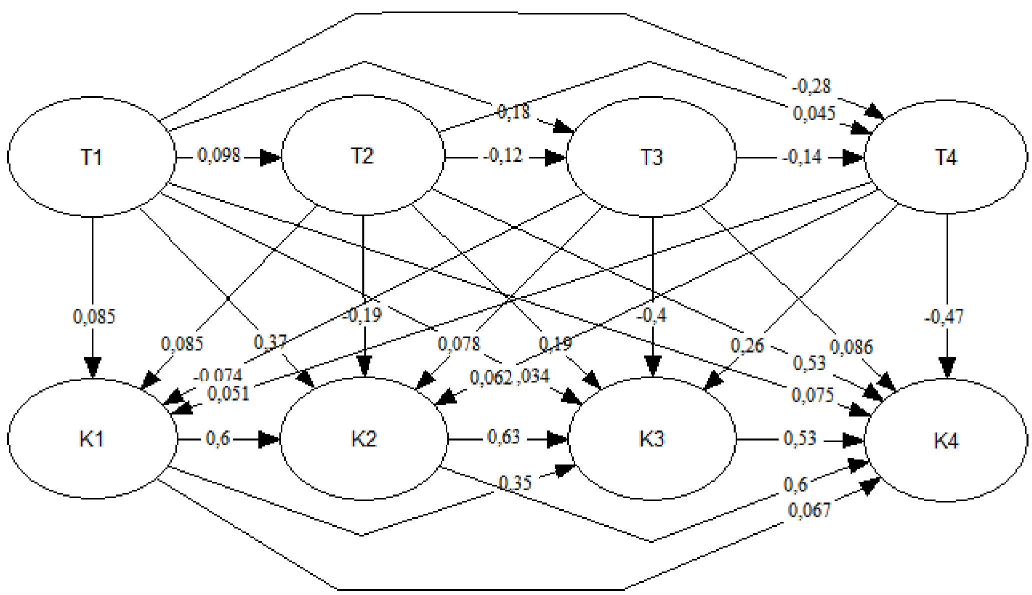
  If the model is the saturated graph (figure 6.1a) the DMR corresponds to the DNR, i.e. $DMR = DNR$. However, when deleting arcs, the corresponding (conditional) correlations are forced to be zero, which introduces conditional independences and, hence, changes the dependence structure. A graph structure, which assumes all variables to be independent, i.e. there are no arcs between the variables (figure 6.1b), has $DMR = 1$. The final model is usually somewhere between this so-called skeletal structure and the saturated structure and therefore its DMR is $DNR \leq DMR \leq 1$. If the DMR is "similar" to the DNR, the second assumption is validated: The BBN with its conditional independence statements is an adequate model of the saturated graph.

Testing whether the DNR is a suitable representative for the DER validates the normal copula assumptions and is described in the following paragraph. Testing the suitability of the DMR is part of the building process of the final BBN and is discussed subsequently. All three determinants can be directly obtained from Uninet[2].

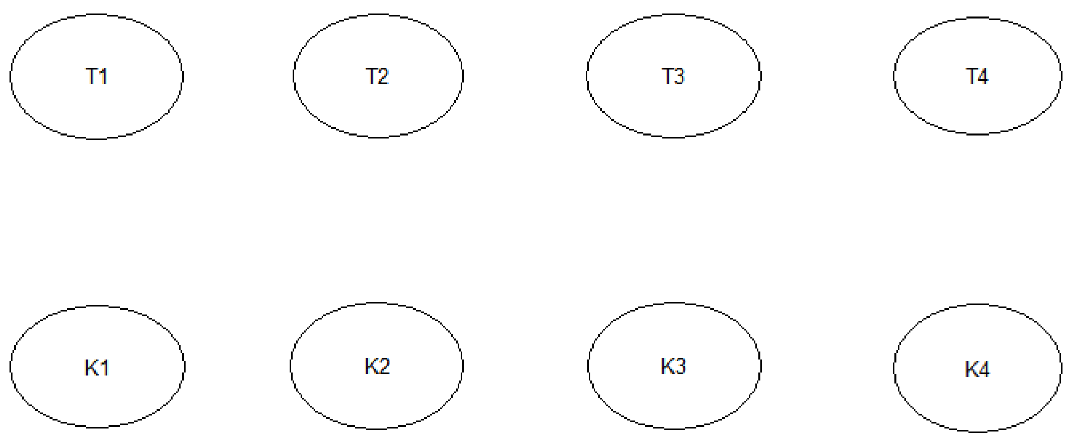
For the eight variables, $T1$ to $T4$ and $K1$ to $K4$, $DER = 0.0300$ and $DNR = 0.0483$. The normal rank correlation matrix is shown in figure 6.2. To check whether the data can be adequately represented by the joint normal copula, we test the hypothesis that DER comes from the sampling distribution of DNR. The sampling distribution of DNR based on 1000 simulations can be calculated in Uninet. The resulting confidence interval is $[0.0035,\ 0.0488]$. DER falls within this interval and the normal copula assumption cannot be rejected on the $\alpha = 5\%$ level. The Uninet output of the test is shown in figure 6.3. This test results also validates the saturated graph, because it has $DMR = DNR$.

[1] In the literature this variable is referred to as $DBBN$, which is already reserved for *dynamic BBN* in this thesis. Hence, the new abbreviation DMR is introduced

[2] available from http://www.lighttwist.net/wp/uninet



(a) Saturated graph of the eight variables.



(b) Skeletal BBN

Figure 6.1: Saturated and skeletal BBN of the eight variables.

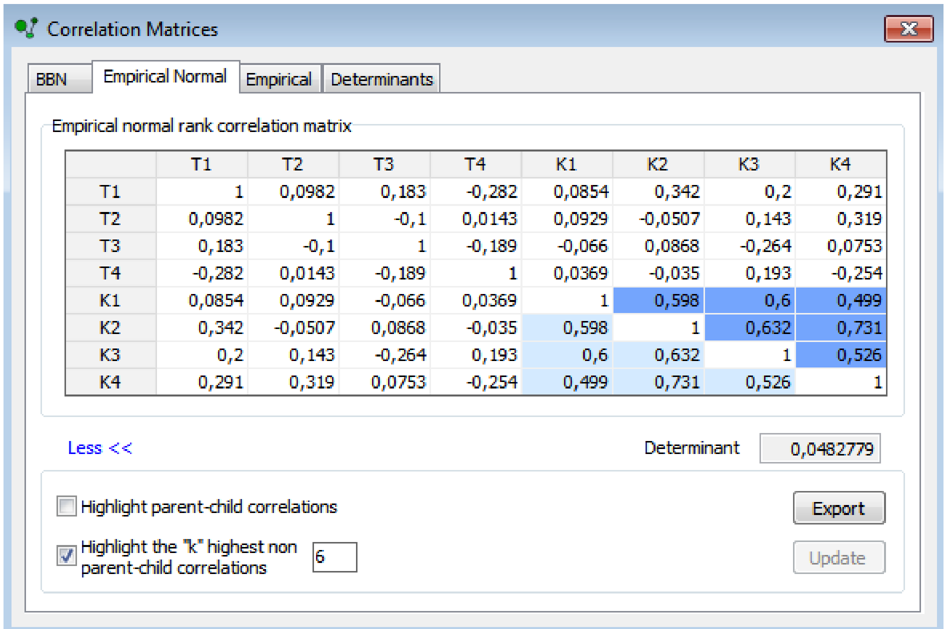


Figure 6.2: Normal rank correlation matrix of the data vectors of the variables *training* and *knowledge*

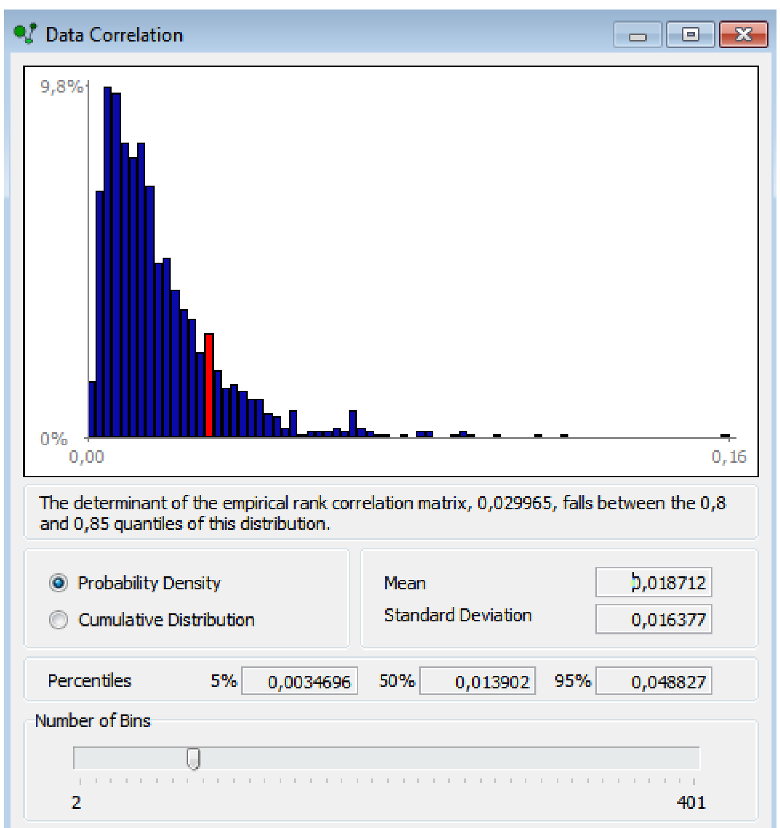


Figure 6.3: Results of the "Validate Normal Copula" function of Uninet

6.2 Learning a Valid BBN Structure

Summarizing the previous sections, the final BBN structure will been "in between" a skeletal BBN and a saturated BBN. Moreover, the arcs are chosen based on the correlation computed from the data and/or their meaningfulness for the application. Intuitively, this yields two approaches to learn the structure of a BBN: One either starts with the saturated graph and deletes individual arcs or one starts with the independent graph and adds individual arcs until a "useful" and valid structure is obtained. The choice to add or delete is based on either correlation or meaningfulness and has to be motivated. Here, the small sample size has to be taken into account, since it affects the significance of the computed rank correlations.

Table 6.1 shows for several α levels the minimum value of a rank correlation necessary for significance (r_c), if the sample size is $N = 24$. The values correspond to a one-tailed test of the null hypotheses that the rank correlation is 0. For instance, for the $\alpha = 0.05$ level only rank correlations $r \geq 0.344$ are significant considering the small sample size in this research. Looking at the normal rank correlation matrix in figure 6.2, we see that only the correlations between the *knowledge* variables are statistically significant on this level.

Table 6.1: Critical values of Spearman's rank correlation for $N = 24$ samples [Ramsey, 1989]

α	0.25	0.10	0.05	0.025
r_c	0.144	0.271	0.344	0.407

Independent from the chosen approach, the learned BBN is not unique. Different structures may be suitable to represent the data. In order to investigate the effect of the structure on the model use, we develop two structures and compare them with each other. First, we start from the skeletal BBN in figure 6.1b trying to add as few arcs as reasonable and second, we start from the saturated graph in figure 6.1a trying to delete only "unrealistic" arcs. As a result, the first BBN will have fewer arcs than the second one.

6.2.1 Adding Few Arcs to the Skeletal BBN

We start by considering the variables independent, as depicted in figure 6.1b. The DNR falls outside the 90% confidence interval of the DMR of this skeletal BBN and, hence, this model is not valid for the data. The correlation matrix for the skeletal BBN and the results of the statistical test are included in appendix D.1.

According to [Hanea et al., 2010], arcs are added between the variables that have the largest rank correlations (in the normal rank correlation matrix). As soon as the DNR is within the 90% confidence interval of the DMR, the procedure of adding arcs is usually stopped. Still additional/different arcs can be added to the model, if it is more realistic and interesting for the user.

The normal rank correlation matrix is presented in figure 6.2. Its values deviate slightly from the empirical rank correlation matrix in figure 5.7b. The arcs with the highest (and significant) correlations are between the various knowledge *variables*. We add one arc from $K1$ to $K2$, one from $K2$ to $K3$ and one from $K3$ to $K4$. These arcs are according to the data not the largest, but these influences are most intuitive and have been expected (cf. figure 5.1). According to the argumentation from section 5.1, the *knowledge* at $t+1$ results from the *training* at t (or possibly $t-1$ and $t-2$). For this reason, arcs from $T1$ to $K2$, $T2$ to $K3$ and $T3$ and $K4$ are added, even though they are not necessarily significant. Figure 6.4 shows the structure, which is valid as shown in appendix D.2.

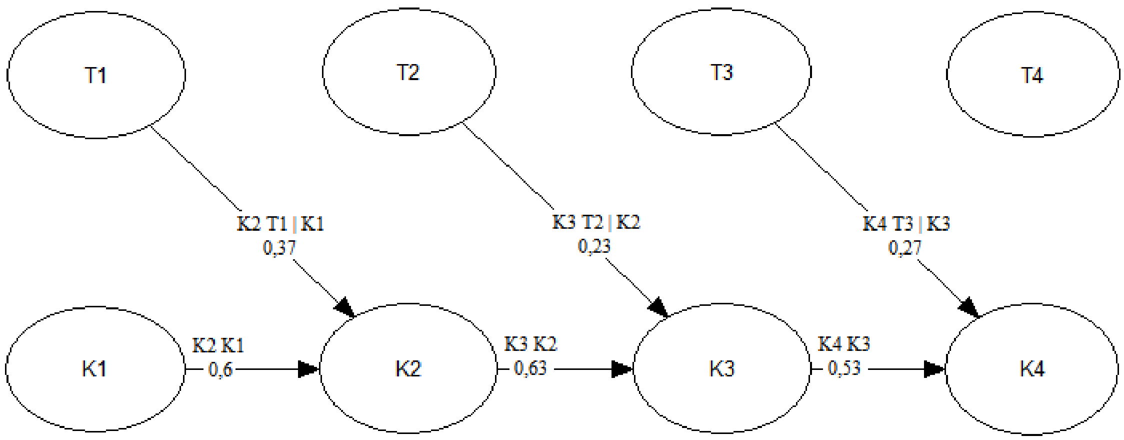


Figure 6.4: BBN with arcs between the consecutive *knowledge* variables and from the *training* variables to the *knowledge* of the next round. The conditional rank correlations are labeled on the arcs.

6.2.2 Removing Few Arcs from the Saturated BBN

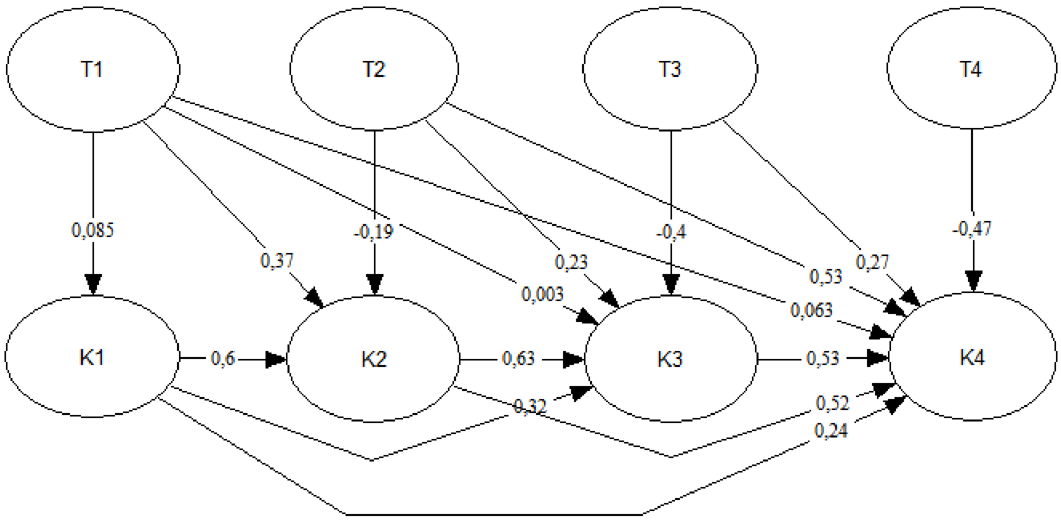
Another possibility is to remove those arcs from the saturated BBN that are very small (close to zero) and/or seem unrealistic and caused by outliers in the data set such that the value of the DNR stays inside the confidence interval of the DMR.

We first remove all arcs that are inconsistent with the causal flow of time from the saturated BBN and the arcs between the four training variables, because the trainings were designed to be independent and identically distributed, namely uniform on $[5,10]$. Not all of these arcs had small correlations. The arcs from $T4$ to $K3$ with value 0.19 and the arc from $T1$ to $T4$ with value -0.28 have the largest ones, but according to table 6.1 they are not significant. Keeping them contradicts the meaning of the variables and would yield invalid model (not from a statistical perspective, but from a user perspective). As long as the structure remains (statistically) valid, it is of course legitimate to remove them. The resulting model is depicted in figure 6.5a and the validation is shown in appendix D.3.

6.2.3 The "Best" Model?

Finding the most suitable model is not straight forward. Besides the two structures introduced in the previous sections it is possible to learn structures that are "in between": having additional arcs to the BBN in figure 6.4, but not yet all the arcs displayed in figure 6.5a. On the grounds of the present data it is difficult to justify which extra arcs should be included. The normal rank correlation matrix (figure 6.2) does not indicate any trends. For instance, $r_{T1,K2} = 0.342$ $r_{T1,K3} = 0.200$ and $r_{T1,K4} = 0.291$ are all more or less of the same order of magnitude, but there is no apparent reason for the fluctuation. Moreover, they are not significant due to the small sample size.

One could argue, for example, that since $r_{T1,K4} = -0.282$ and is set zero on the basis of the assumed independence between trainings, all rank correlations $|r| < 0.282$ should be set to zero



(a) BBN displaying all plausible arcs

(Conditional) Rank Correlation Coefficient	Value
K3 K4	0,52591
K2 K4 | K3	0,59972
K1 K4 | K3 K2	0,067038
T1 K4 | K3 K2 K1	0,074522
T2 K4 | K3 K2 K1 T1	0,52877
T3 K4 | K3 K2 K1 T1 T2	0,085888
T4 K4 | K3 K2 K1 T1 T2 T3	-0,46545

(b) Table with the dependence info of node $K4$. Other nodes are analogue.

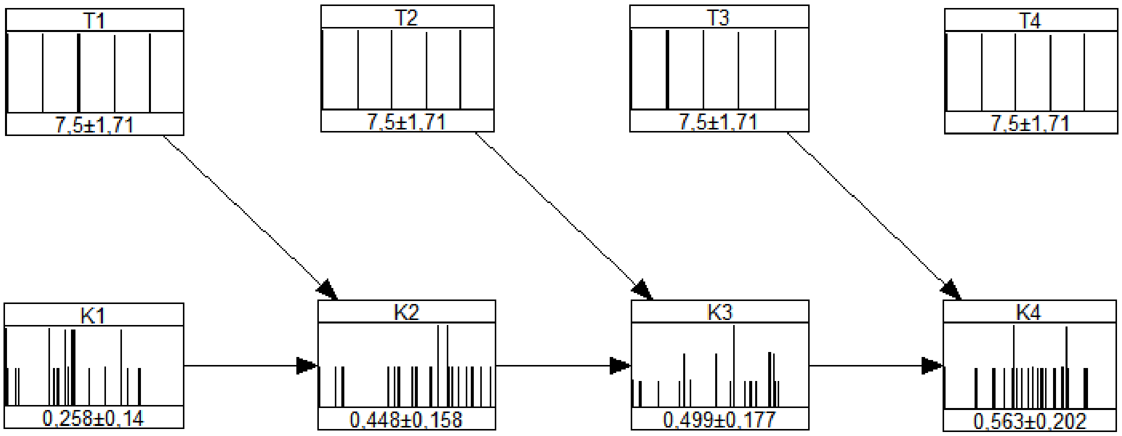
Figure 6.5: BBN displaying all plausible arcs. For the sake of clarity the arcs are not labeled in the graph, but an example for node $K4$ is given in the table

as well in order to be consistent. However, then $T1$ has only influence on $K2$ and $K4$, but not on $K3$. $T2$ would only have an arc to $K4$ and $T3$ and $T4$ would be independent of all other variables. Such a structure is not realistic and is on the basis of a sample size $N = 24$ not defensible.

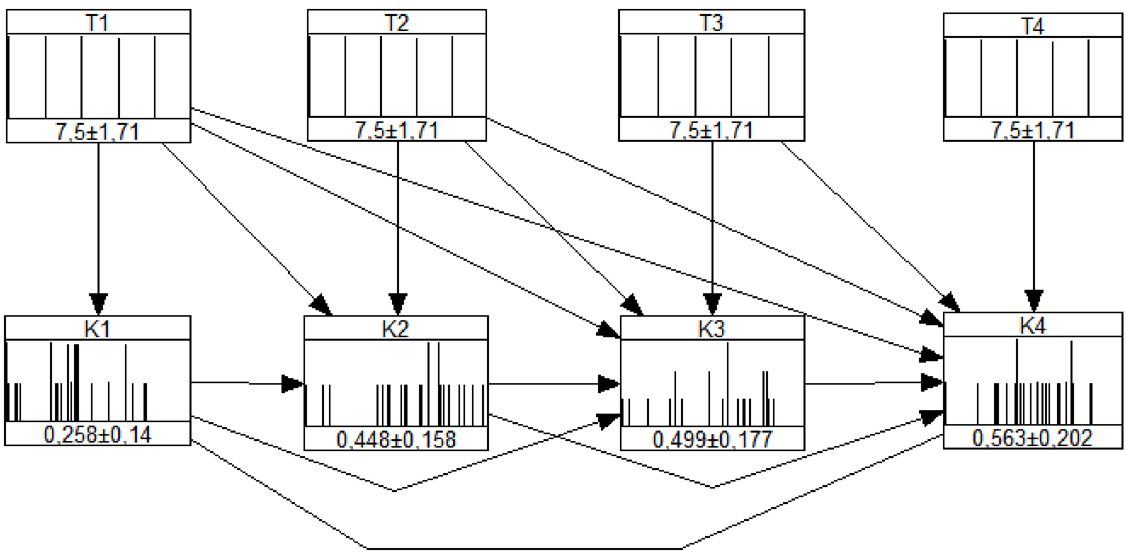
Since, as yet, there is no indication which of the two presented structures, or possibly an "in between" structure, is most suitable, the reminder of this chapter investigates what the consequences on the model outcome are, if one or the other choice is made.

6.3 Comparing the Outcomes of Two Valid Model Structures

Figure 6.6 shows (again) the two models, this time with the nodes as histograms. Unfortunately, there is no feature in Uninet for choosing the number of bins that are displayed in the histograms. This makes it also impossible to display $K1$ to $K4$ with the same ranges to emphasize how the



(a) Model structure 1: Few arcs



(b) Model structure 2: Many arcs

Figure 6.6: The two structures with the nodes as histograms. To see the histograms better, the nodes have been moved close together at he cost that the arcs are not that visible anymore.

knowledge evolves over time.

The distribution of all 4 training variables is identical. The mean is $\mu = 7.5$ and the standard deviation is $\sigma = 1.71$. It is usual that this value slightly deviates from the one in table 5.1, because the distribution are re-sampled in Uninet. The empirical probability mass function of $K1$ has mean $\mu = 0.26$ and standard deviation $\sigma = 0.14$. Both, mean and standard deviation, increase over the rounds. Generally, participants improve their knowledge, but the differences between them become larger.

At this point, the two models are identical, because the distributions are directly taken from the data. Only the dependence structure differs. As soon as some variables are conditioned, the model can be updated and the resulting distributions may differ. Since the model is data mined, it is only possibly to condition on a value that is within the range of the original data after having been transformed to standard normal. For example, it is impossible to condition on $K1 = 1$, because the range of $K1$ is only between approximately 0.06 and 0.59.

In the following $K1_{model1}$ denotes the node $K1$ of the model structure with few arcs and $K1_{model2}$ the one of the model structure with many arcs. The same applies to the other variables.

6.3.1 Conditioning on High Training

A natural management action to increase the knowledge of technical staff members, is to object them to more intense training. In terms of the game, it means to let the participants play to higher levels each round. Figure 6.7 demonstrates the effect on the knowledge. It shows how the knowledge distributions change (according to each of the two models), if a participant played up to and including level 10 in all four rounds. Both BBNs in figure 6.7a and 6.7b exhibit similar results, but the predicted knowledge values from the structure with few arcs are slightly higher.

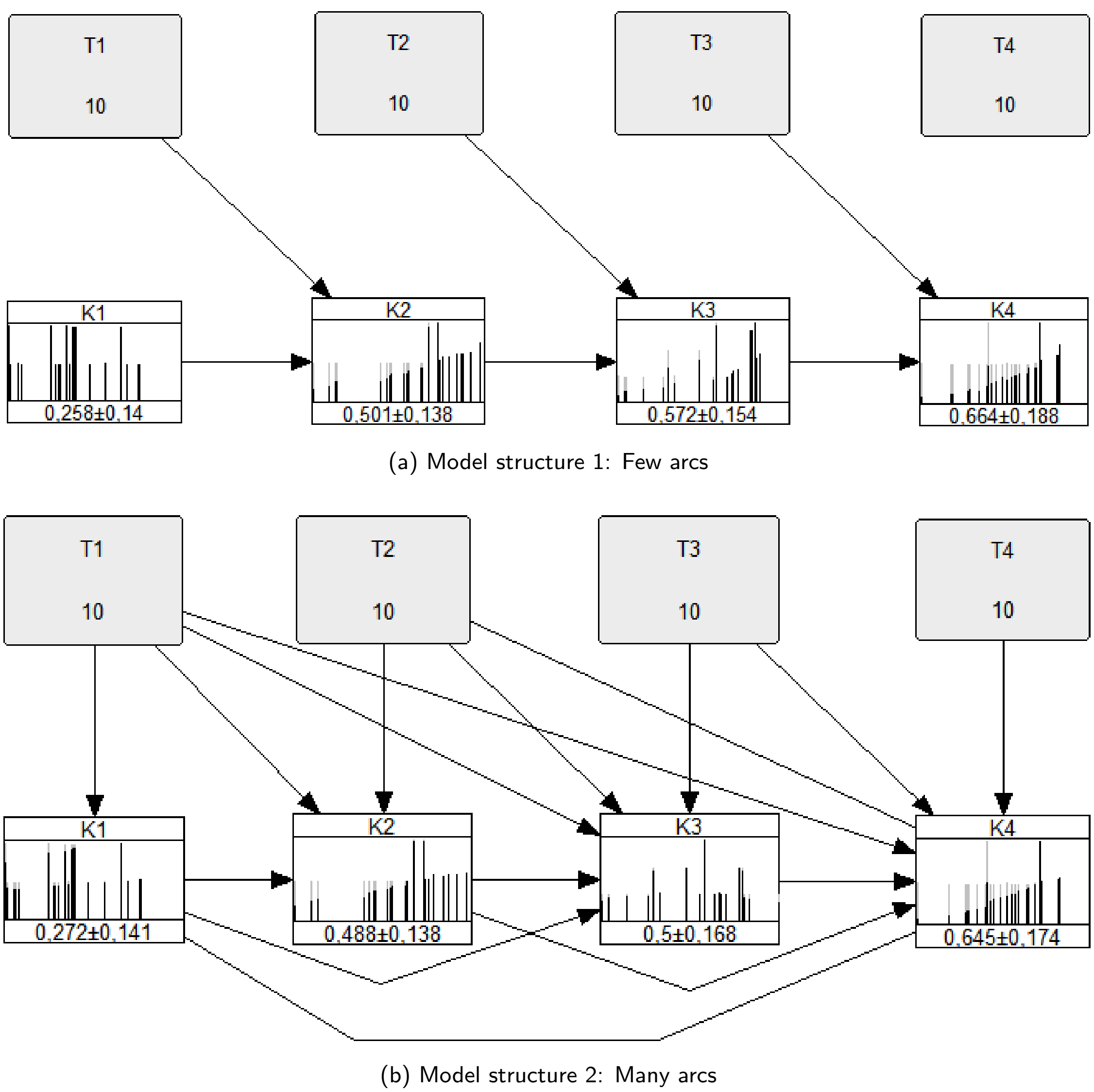


Figure 6.7: Updated distributions when conditioning on $T1 = T2 = T3 = T4 = 10$

The conditional distributions are shown in black and the original ones in gray. There is no difference in node $K1_{model1}$ and a small tendency towards higher values in $K1_{model2}$. The mean of $K2_{model1}$ is about 12%, the one of $K3_{model1}$ about 15% and the one of $K4_{model1}$ about 19% higher than with uniform training. On the contrary the mean of $K2_{model2}$ is only about 8%, the one of $K3_{model2}$ about 0.4% higher and the one of $K4_{model2}$ about 18% higher than before. The value of $K3_{model2}$ is strange, both from a logical viewpoint (Why would the relative increase in knowledge fluctuate like this?) and in comparison to the other model (Why do the $K3$ nodes differ, if nodes $K2$ and $K4$ show comparable outputs?).

Moreover, for both models the increase in variances is slightly less than before, which indicates that part of the variation in knowledge can be explained by the different training opportunities the participants had. The other part can be regarded as aleatory uncertainty; it is a natural variability within the system.

Despite the fact that the trend is an increase in knowledge, the range of knowledge values does not change in any of the models: there is a probability of having very little knowledge even after constant high training, as can be seen from the black bar in node $K4$ (the leftest bar). This kind of information is crucial for risk management. Hypothetically, consider this results as an outcome of a realistic model of training staff who operate in a high risk environment. Despite more training, there is a possibility that one of the staff members has insufficient knowledge to cope with a critical situation and, hence, there is a possibility of an accident with disastrous consequences.

6.3.2 Conditioning on High Initial Knowledge

Knowing that high training does not eliminate the possibility of having one employee whose knowledge is insufficient, another hypothetical management action could be to only promote and train those employees who have already proven to be knowledgeable, e.g. having $K1 > 0.5$ in terms of the experiment. Because it is rather complicated to condition on more than one value in Uninet, we settle for conditioning on $K1 = 0.5$. The results are shown in figure 6.8a. and 6.8b.

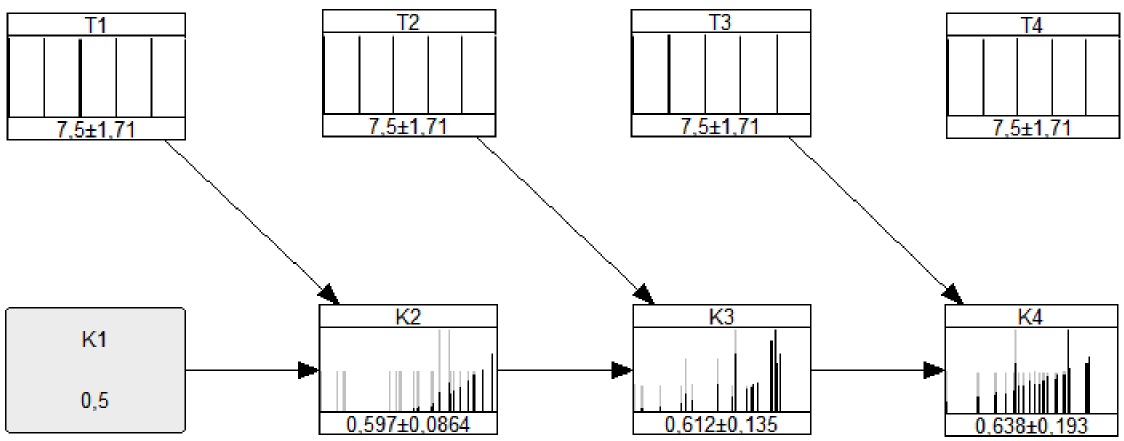
The effect on the training variables in very minor for both models. In *model 1* the effect on $K2_{model1}$ is strongest, which can be seen from the gray and black histograms. The mean of $K2_{model1}$ is $\mu = 0.679$ and the standard deviation is extremely small $\sigma_{model1} = 0.0864$. Almost the entire lower half of knowledge values receives probability zero. However, the effect diminishes towards $K4_{model1}$, which has with $\mu = 0.638$ a slightly lower mean than when conditioning on high training and the variance $\sigma = 0.192$ is comparable. Also, the lowest knowledge value receives a small probability.

In *model 2* the effect is very strong on all knowledge variables. $K2_{model2}$ seems identical to $K2_{model1}$. However, afterwards the model predicts an almost constant increase in knowledge, whereas the increase in knowledge diminishes for *model 1*, as demonstrated in table 6.2. The variances are close to zero for $K2_{model2}$ and $K3_{model2}$, but is suddenly quite high for $K4_{model2}$ ($\sigma = 0.164$). A big difference to *model 1* is that the lower values of all knowledge variable receive probability zero.

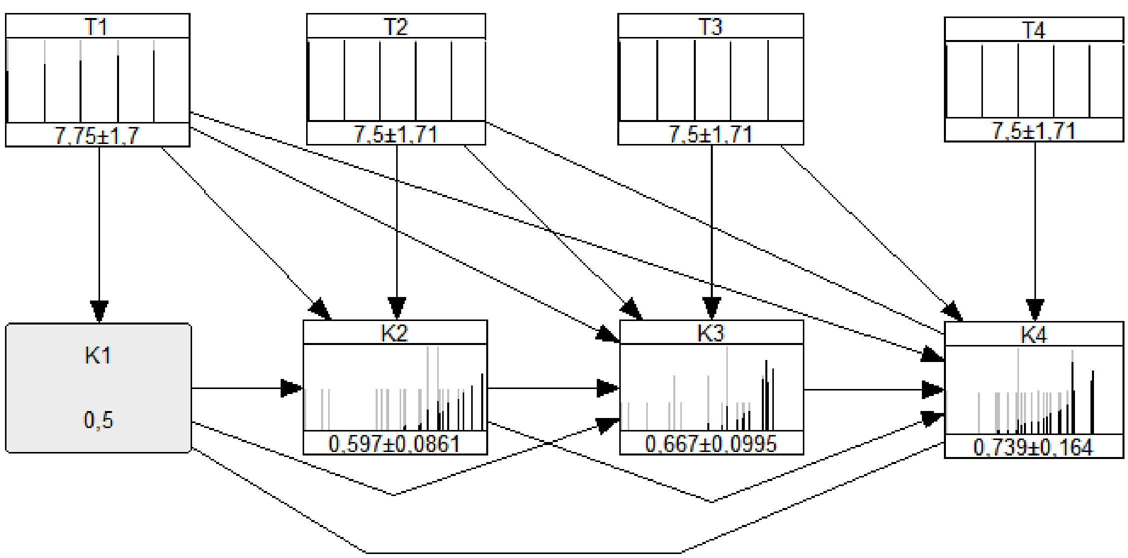
The conclusions that can be drawn differ for the two models. *Model 2* predicts that a sufficiently high initial training guarantees that high values of knowledge are reached in the fourth round. However, *model 1* does not confirm this results. A conclusion from its results of the two management actions could be to have more intense trainings and also frequent testing sessions to ensure sufficient knowledge.

Table 6.2: Comparison of the increase in knowledge predicted by the two models

	$K2 - K1$	$K3 - K2$	$K4 - K3$
model 1	0.097	0.015	0.026
model 2	0.097	0.071	0.072



(a) Model structure 1: Few arcs



(b) Model structure 2: Many arcs

Figure 6.8: Updated distributions when conditioning on $K1 = 0.5$

Nevertheless, this simple case of only 24 students playing 3Dlogic for only four rounds is not suitable to test the effect of these kind of management actions in a meaningful way.

6.3.3 Conditioning on Low Initial Knowledge

In contrast to the previous section, one can also condition on low initial knowledge. The lowest possible value is $K1 = 0.05001$. The results are shown in figures 6.9a and 6.9b.

In both models this condition hardly affects any of the training variables. Based on *model 1* the effect of this conditioning is strongest on $K2_{model1}$. The mean of $K2_{model1}$ is $\mu = 0.306$, which is ca. 30% lower than without conditioning, and the standard deviation is $\sigma = 0.151$, which is roughly the same. The highest knowledge values receives probability zero. However, the effect diminishes towards $K4_{model1}$, whose mean $\mu = 0.502$ is just 11% lower than usual and the variance $\sigma = 0.199$ is again comparable. In the case of low initial knowledge, the probability of reaching high knowledge in the fourth rounds is about half compared to the original one. Similar to the previous section the effect on $K2_{model2}$ is very similar to the one on $K2_{model1}$. In contrast, the knowledge of the subsequent rounds is predicted to be notably less than in *model 1*.

In conclusion, it can be inferred that a participant with low initial knowledge could according to *model 1* reach high values in round 4, while he/she would not be able to reach them according to *model 2*.

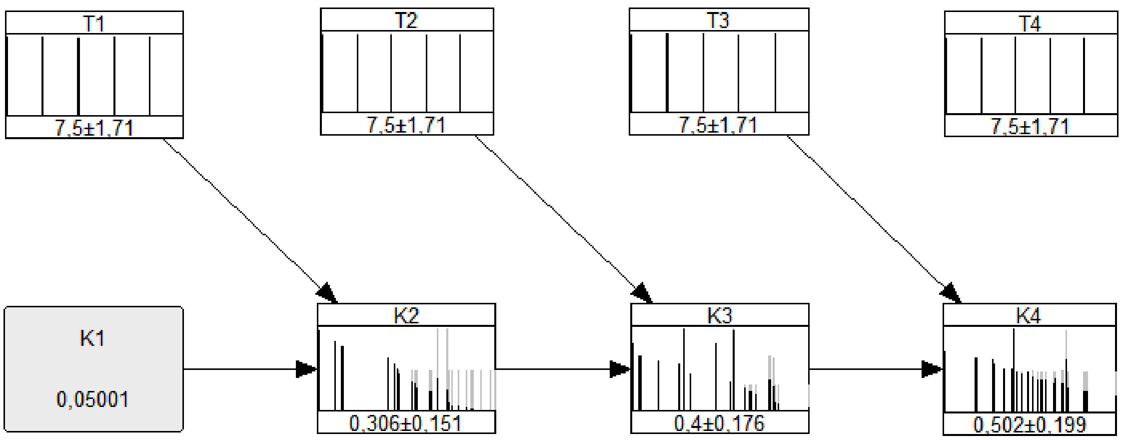
6.3.4 Conditioning on Maximal Knowledge in the Last Round

Another way to use the model is to reason diagnostically, instead of predictively as done in sections 6.3.1, 6.3.2 and 6.3.3. Conditioning on $K4 = 1$ alone implies that a participant has knowledge 1 in the fourth round, but that it is unknown which trainings or previous knowledges he/she had. Figures 6.10a and 6.10b show the updated distributions.

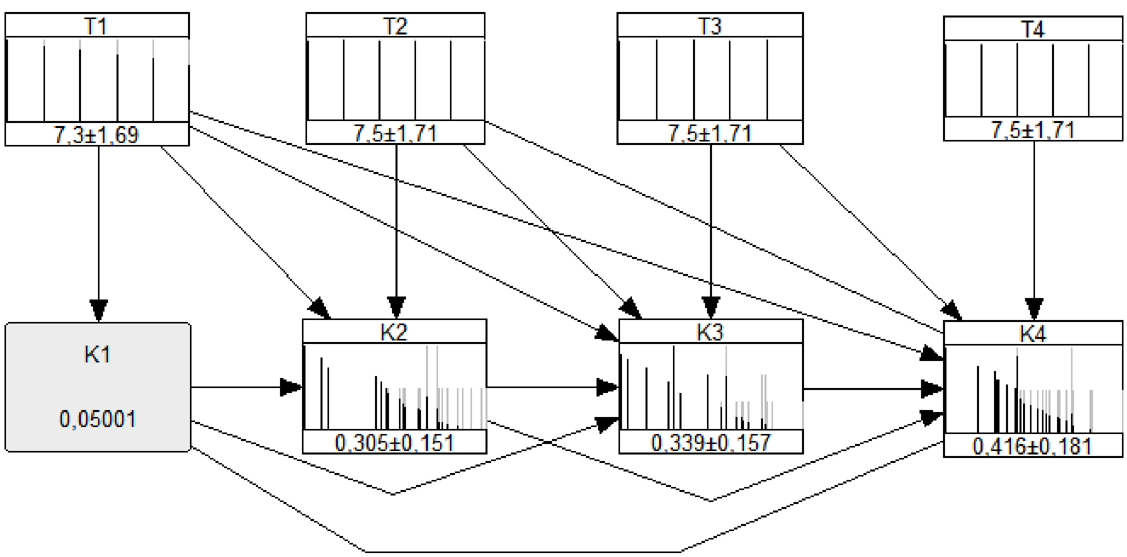
This condition has a strong impact on all other knowledge distributions. Also, it is likely that the participant has received a higher training than the average.

Model 1 indicates that, it is not impossible that he/she obtained low trainings. Most striking is that the slope of the participant's learning curve did not decrease systematically as before. From $K1$ to $K2$ his/her knowledge increased on average by 0.24, from $K2$ to $K3$ by 0.15 and from $K3$ to $K4$ by 0.24. A possibly surprising result is that he/she could have had initial knowledge of any value of the given range (approximately [0.06,0.59]) and that the mean of $K1$ is with $\mu = 0.38$ not extremely high.

On the contrary, *model 2* claims that the participant did certainly not have low trainings or low knowledge in any of the previous rounds. ($T3_{model2}$ is an exception.)

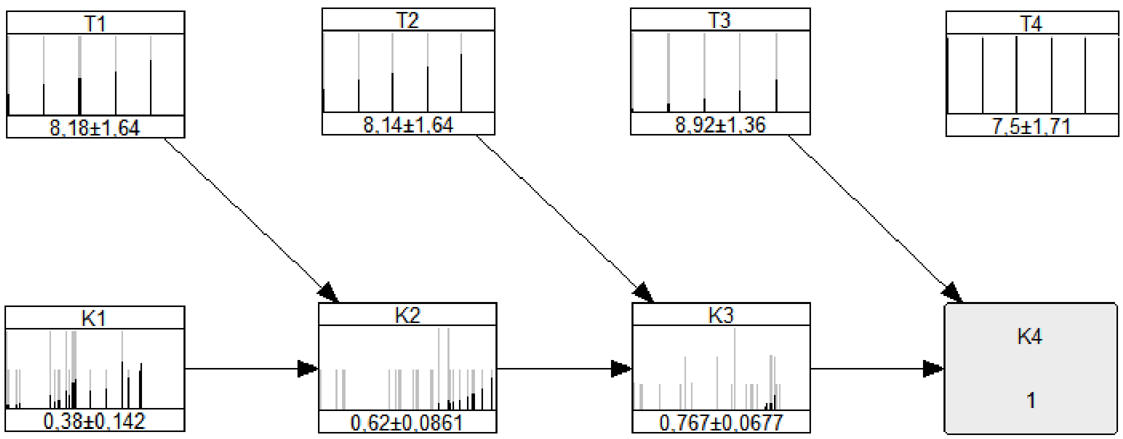


(a) Model structure 1: Few arcs

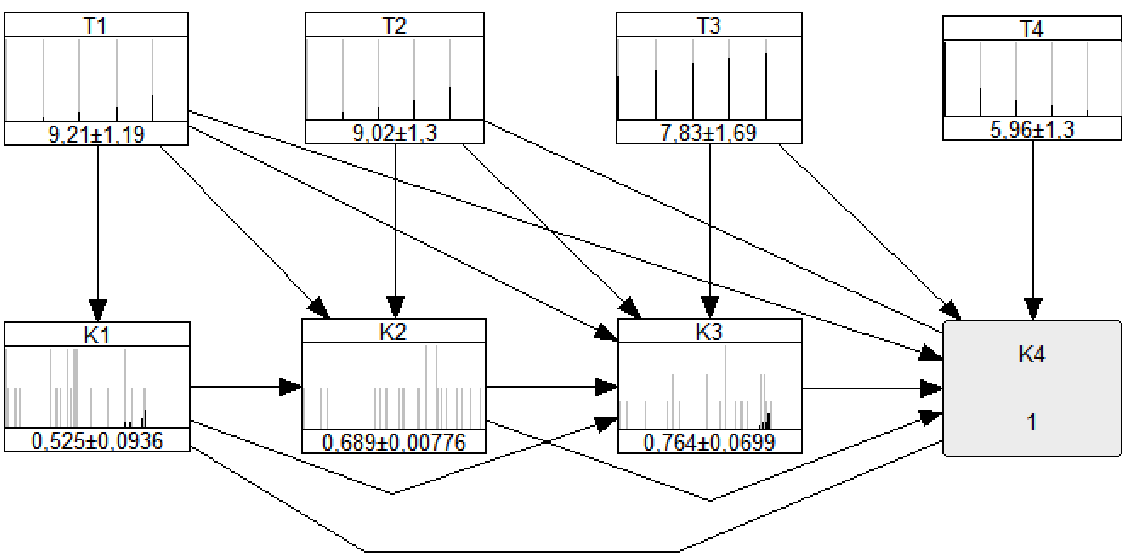


(b) Model structure 2: Many arcs

Figure 6.9: Updated distributions when conditioning on $K1 = 0.05001$, which is the minimum value possible



(a) Model structure 1: Few arcs



(b) Model structure 2: Many arcs

Figure 6.10: Updated distributions when conditioning on $K4 = 1$

6.4 Chapter Conclusions

At the beginning of this chapter, the process of building a dynamic nonparametric BBN, which adequately reflects the data from the experiment, was discussed. To this end *knowledge* and *training* of each round were regarded as individual variables of the model. It was found that several statistically valid structures exist that have a different number of arcs between the variables. In order to investigate the consequences of using one or another structure on the model outcome, a rather minimalistic structure and a structure with all plausible arcs have been compared.

Two hypothetical management actions have been translated to the context of playing *3Dlogic* and implemented to illustrate how the impact on the *knowledge* evolution over four rounds can be predicted with the dynamic BBN. First, the *training* has been increased in all rounds (BBN conditioned on $T1 = T2 = T3 = T4 = 10$) and, second, the a specific initial knowledge has been required (BBN conditioned on $K1 = 0.5$). In both models the updated distributions are skewed to the right, i.e. higher knowledge, but the range of values did not change. Even though with a low probability, there is the possibility of very low knowledge in the fourth round. In general, this kind of insight is very important in risk modeling, because low probability events may have devastating consequences and their risk needs to be mitigated. Then, the initial knowledge has been conditioned once as high ($K1 = 0.5$) and once as low ($K1 = 0.05001$). Finally, an example was given on how to reason diagnostically with a dynamic BBN. Similarly, if a participant had low initial knowledge, the updated distributions were skewed to the left.

For the last three cases the differences between the models were larger. In general, *model 1* predicted a wider range of possible values for the updated histograms of all other variables than *model 2*. In other words the output of *model 1* has more uncertainty. One could argue therefore, that *model 1* is the better choice considering the very small sample size. Moreover, it has the advantage of having a more transparent structure. Nonetheless, the purpose of this chapter was to illustrate the consequences the choice for one or the other structure on the model output. It has to be kept in mind that only demonstration case has been studied and that many of the correlations were not found to be significant. Giving a recommendation on the "best" structure is difficult, but also not useful in the current stage of research.

In section 3.2.3 the principle of time slicing and "unrolling" them up to arbitrary times T had been described. However, neither the correlations nor the distributions have been found to be time-invariant. From the present data it is not possible to infer the states of *knowledge* and *training* in future rounds.

A perspective on how this demonstration case could be extended to a more applicable model and on how to use dynamic BBNs to model human and organizational factors for risk analysis is given in the following chapter.

Chapter 7

Conclusions

This thesis started by pointing out that despite of all efforts, low probability high-consequence accidents, such as the recent blowout in the Gulf of Mexico and the nuclear disaster in Fukushima, do not appear to occur less frequently, while their consequences appear to increase. The goal of zero accidents has not been achieved with current approaches in risk management. Hence, improved safety management systems (SMS) are of great social relevance.

Since dynamic human factors and management actions are recognized as the more fundamental causes to accidents in high hazard industries than random combinations of technical failures, it is essential to integrate them in a single model with the technical components of the system. Bayesian belief nets (BBNs) seem a promising methodology. But only recent mathematical developments allow for dynamic relationships in BBNs and applications of dynamic BBNs are not well-known as yet. Accordingly, the aim of this project was to introduce dynamic BBNs as a tool to monitor the effect of management actions on safety and, thus, to support the design of improved safety management models. The main research question was:

> *How can dynamic nonparametric BBNs support the design of improved SMS by enabling a monitoring of the effect of management actions on safety?*

This concluding chapter is dedicated to illustrate to what extent the produced outcomes are congruent with the research objectives stated in the introduction. Section 7.1 therefore provides answers to the research questions and highlights the deliverables of the thesis. It also describes the limitations and reflects on how the chosen approach affects the results of the research. Then, section 7.2 discusses the relevance for the scientific community and indicates potential directions for future work in the field.

7.1 Main Findings and Limitations

The main research question has been broken down in four sub-questions, which guided the research process. The main findings for each sub-question are summarized below.

1. **Why is monitoring the effect of management actions relevant to improve safety of engineering systems?**

   A distinction of active and latent errors helps to understand how accidents arise as a concurrence of various human faults and management deficiencies. Active errors have immediate, visible consequences and are typically committed by front line operators, e.g. control room operators. Latent errors, on the contrary, are primarily associated with high-level decision makers and managers. They are dormant for a long time before an accident sequence begins and only become evident when they coincide with other faults. Often, disasters are released by active errors or external triggering events, but the latent root causes have long been present.

Unfavorable preconditions for the technical staff (also called human factors or performance shaping factors), such as insufficient knowledge of the system, time pressure or concentration, are typical latent causes for accidents. Between preconditions and active human faults there is a so-called few-to-many mapping, i.e. a bad precondition may open the pathway for many unsafe actions. In other words, bad preconditions increase the probability of human error. Naturally, it is more effective to safeguard good preconditions than trying to prevent the occurrence of unsafe acts directly. This is a first reason to monitor how management decisions influence the preconditions and, thus, the system's safety.

A second reason is that managers often seem to be more motivated to control short-term risks and direct efforts towards productivity, because their success becomes immediately visible, and ignore the risk of rare disaster. Some risk events are so rare that they are unlikely to happen during the manager's time in charge, whether he attempts to mitigate them or not. His efforts may not visible and may not be rewarded. A model that monitors the effects of all kinds of management actions on safety may make these efforts more visible and motivate managers to more intensely control low-probability risks.

2. **How can human factors and management influences be incorporated in technical risk models?**

   Event tree (ET) and fault tree (FT) analyses are the commonly used tools to model failures of technical components. However, they have severe limitations to include uncertainty as well as common cause failures. Moreover, dynamic relationships cannot be represented. Because human factors may influence the probability of human error at several points in the system, but it need not cause a fault, these tools alone are not suitable to model them.

   Two models known from the literature, a causal model for air transport safety (CATS) and a socio-technical risk analysis (SoTeRiA), attempt to incorporate human factors using a mixed logic approach. CATS combines ETs and FTs with nonparametric BBNs to a homogeneous mathematical model. It is a static, highly specific, quantified model that incorporated a few selected human factors and management influences in order to determine the accident probability of a flight. SoTeRiA, on the other hand, is rather a proof of concept emphasizing the need to include dynamics, in particular feedback, when modeling human and organizational factors. It consists individual of ET and FT modules, BBN modules and System Dynamic (SD) modules. However, no method is proposed to quantify the deterministic equations of the SD modules that deal with human factors. For this reason there are no grounds for using this methodology in this research and the decision to investigate the potential of dynamic BBNs for human factor modeling in risk analysis is substantiated.

   A problem arising when modeling human performance is the lack of metric for human factors, which makes it difficult to operationalize the factors. Proxies have to be used, although they capture only very limited part of the factor. Of course, there are countless possible proxies of a factors and for each application different choices have to be made. Therefore a realistic set of time dependencies and feedback loops between management actions and human factors can only be made on the basis of a specific example justifying the operationalizations chosen.

3. **What is the added value of dynamic nonparametric BBNs when modeling human factors and management actions?**

   Nonparametric Bayesian belief nets (BBNs) are a type of probabilistic graphical model consisting of *nodes* (variables) and *arcs* (influences). They represent the probability distributions of the variables and can realistically reflect inherent variabilities in a system. Considering that the differences in human performance among individuals are due to many factors, e.g. genetics or the situation in private life, which we have neither influence on nor can learn about thoroughly, it is essential to reflect these uncertainties in a safety management model.

   Because a specific example is needed to explore the benefits of nonparametric BBNs, it has been chosen to consider the human factor *knowledge* and the management action *training* in the limited context of students playing an online game. The data has been gathered with a simple experiment. The levels the participants completed were interpreted as *training*, whereas the *knowledge* was inferred from the required time. The model can then easily be quantified using the "new data mining model" option in the software Uninet. However, the model could also be quantified if data were not readily available, e.g. through expert judgment. A mixed approach is of course also possible. This flexibility with regard to the type of data is another major advantage of nonparametric BBNs.

   Moreover, BBNs are able to depict relationships in an intuitive manner. Cause-effect relationships are represented through the directionality of the arcs. In particular, the causal flow of time is indicated with arcs from the left to the right.

   In this research the relationship between training and knowledge has been modeled for four consecutive time steps. In principle the model could be extended to arbitrarily many time steps, but the nodes and arcs of each time step need to be quantified individually. It remains questionable, if the principle of *un-rolling*, i.e. modeling two time steps and then duplicating the structure up to the desired number of time steps, can be applied. The example case gives no indication that the nodes and arcs are time-invariant so that they can be duplicated. Hence, due to the required quantification efforts, it is at the current stage of research only feasible to model a considerably small number of time steps with nonparametric BBNs.

4. **How can management actions be developed with dynamic nonparametric BBNs?**

   The main use of BBNs is updating the distributions of the variables once new evidence on one or more variables is obtained. Because evidence can be propagated forwards (in the direction of the arcs) and backwards (in the opposite direction of the arcs), one can reason predictively as well as diagnostically. On one hand, the effect of certain management actions on a human factor in the future can be investigated. For instance, the participant's level of knowledge, if subjected to high trainings, can be predicted. Or, another hypothetical management action could be to only promote and train those participants who have already proven to be knowledgeable. By conditioning on high knowledge in the first time step, the effect on the future levels of knowledge can be predicted. On the other hand, by conditioning on high knowledge in the last time step, it is also possible ot investigate which trainings have to be given in order to achieve this desired level of knowledge.

   The output of the updated nonparametric BBN in Uninet includes mean and variance of the variables, which immediately shows the trend of the influence: does it increase or lower the value of the variable and how uncertain is that (roughly)? Additionally, the histograms of the variables are given, which enables to model user to see the whole range of possible values for it. Especially, extreme values with very low probabilities are visible. This allows to design management actions, which explicitly aim to prevent these extreme values. In the same manner it could give an incentive to managers to make efforts of mitigating low-probability-high-consequence events, since it makes these efforts visible and possible to acknowledge.

From the answers to the research questions above can be concluded that dynamic nonparametric BBNs have great potential to represent human and organizational factors in a realistic manner. They can be a useful tool to design improved SMS through monitoring the effect of management actions on safety. But of course the methodologies has some limitations as well. These limitations, especially the latter three, have to be kept in mind when building a realistic model.

- Due to constrains in time, resources and scope of the research, the demonstration case is very limited. First, the sample size was very small and correlations smaller than $|r| <= 0.344$ are not significant on the $\alpha = 5\%$ level. Second, the predictive quality of the models has not been investigated. Improvements are left for further research.

- In a realistic case the nature of the data may not be suitable for this methodology. However, this is rather unlikely, because the assumption that the data can be joint by the joint-normal copula is not very restrictive.

- It is only feasible to model a small number of time steps, because the nodes and influences have to be quantified for each time step individually.

- In this research, it has only be conditioned on point values. However, it would be more realistic to condition on a rage of values, because there is usually a minimum acceptable level or a maximum acceptable level of a factor. How to achieve this has not been investigated as part of this project.

- Because of the nature of human factors, strict assumptions have to be made in order to operationalize them. The definitions of the variables are proxies which capture a very limited part of reality. Naturally, these proxies have strong impact on the model output and have to be chosen very carefully keeping the purpose of the modeling project in mind. For this reason, the model can only make predictions and diagnoses for the system it has been designed for and can by no means be transferred to another context.

- Many different model structures may be statistically valid and the methodology gives no indication which one is the "best". The final choice has to be made by the analyst based on other factors, foremost the meaning of the variables, consistency with the causal flow of time as well as the research goal and special interests of the model user. This step has not been done in this research, because the example is too restricted to provide meaningful arguments for the choice. Instead, two model structures have been compared to demonstrate the impact different choices have on the output.

7.2 Contributions and Suggestions for Future Work

The contributions of this thesis are mainly scientific and are two-fold. On one hand it adds to the limited literature on risk models, which explicitly incorporate human performance, by proposing a new approach to model and monitor the influence of management actions on human performance in a dynamic context. The demonstration case illustrates how dynamic nonparametric BBNs can reflect inherent variabilities in the factors, which is essential in human performance modeling. The natural differences among individuals can hardly be predicted and this uncertainty needs to be represented by the model. On the other hand it is relevant to the new field of dynamic BBNs. To our knowledge, the quantification of a dynamic nonparametric BBN from data has not yet been discussed in the literature. In this thesis the steps of building and validating a model are described in detail using a demonstration case, but they could be applied to various applications, also in other fields than risk analysis and to applications in the industry.

As indicated in the discussion on the limitations in the previous section, specific recommendations for follow up research address three points. First, in order to improve the statistical significance of the results the experiment needs to be repeated with an increased sample size. According to [Ramsey, 1989] 100 samples are needed to ensure that all correlations $r \geq 0.165$ are significant. In order for lower correlation values to be significant even more samples are required, but a concrete number for the sample size cannot easily be found in the literature and may need to be computed. Since the (conditional) correlations in the demonstration models have either $r \geq 0.165$ or are very close, it seems sufficient to include around 100 samples to secure significance for all arcs. Second, it could be researched how to condition on a range of values, i.e. a critical value and higher or lower, for the factors. Third, the quality of the model predictions could be investigated so as to better assess the value of dynamic nonparametric BBNs for human factor modeling. This can be done by having a second group of participants play the game under specific conditions and compare the results with the output of the model, which has the same conditions implemented. It is expected that such an experiment helps to judge which one of the two model structures is more useful for predictions or, possibly, that a third structure not treated in this thesis may be most suitable.

More generally, for realistic applications extensive case-specific research has to be made to determine the set of human factors, which will be included in the model, and on how to operationalize them. It should further be investigated how they can be quantified: whether data is available or can be made available e.g. through experiments, or whether expert judgment has to be employed. Because this topic could be researched endlessly and a generally valid answer can probably not be found in the near future, the scope of the research has to be set for each application individually.

References

[Ale, 2009] Ale, B. (2009). *Risk: an introduction: the concepts of risk, danger and chance*. Routledge.

[Ale et al., 2009a] Ale, B., Bellamy, L., Cooke, R., Duyvis, M., Kurowicka, D., Lin, C., Morales, O., and Roelen, A. (2009a). Causal model for air transport safety. Final report. http://repository.tudelft.nl/view/ir/uuid%3Ae4c5e6b0-9e20-4e61-993e-1d9bd0e23d07/.

[Ale et al., 2006] Ale, B., Bellamy, L., Cooke, R., Goossens, L., Hale, A., Roelen, A., and Smith, E. (2006). Towards a causal model for air transport safetyan ongoing research project. *Safety Science*, 44(8):657–673.

[Ale et al., 2009b] Ale, B., Bellamy, L., Van der Boom, R., Cooper, J., Cooke, R., Goossens, L., Hale, A., Kurowicka, D., Morales, O., Roelen, A., et al. (2009b). Further development of a causal model for air transport safety (cats): Building the mathematical heart. *Reliability Engineering & System Safety*, 94(9):1433–1441.

[Ale et al., 2012] Ale, B., Hanea, D., Sillem, S., Lin, P., Gulijk, C., and P., H. (2012). Modeling risk in high hazard operations: integrating technical, organizational and cultural factors. Manuscript submitted for publication.

[Amalberti, 2001] Amalberti, R. (2001). The paradoxes of almost totally safe transportation systems. *Safety Science*, 37(2):109–126.

[Aronson et al., 1999] Aronson, J., Lustina, M., Good, C., Keough, K., Steele, C., and Brown, J. (1999). When white men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of experimental social psychology*, 35(1):29–46.

[Bedford and Cooke, 2001] Bedford, T. and Cooke, R. (2001). *Probabilistic risk analysis: foundations and methods*. Cambridge University Press.

[Bousquet et al., 2004] Bousquet, O., Boucheron, S., and Lugosi, G. (2004). Introduction to statistical learning theory. *Advanced Lectures on Machine Learning*, pages 169–207.

[de Bruijn and ten Heuvelhof, 2008] de Bruijn, H. and ten Heuvelhof, E. (2008). *Management in Networks: On multi-actor decision making*. Routledge.

[Enserink et al., 2010] Enserink, B., Hermans, L., Kwakkel, J., Thissen, W., Koppenjan, J., and Bots, P. (2010). *Policy analysis of multi-actor systems*. Lemma.

[Forrester, 1961] Forrester, J. (1961). *Industrial dynamics*. The MIT press and John Wiley & Sons.

[Frantzich, 1998a] Frantzich, H. (1998a). Risk analysis and fire safety engineering. *Fire Safety Journal*, 31(4):313–330.

[Frantzich, 1998b] Frantzich, H. (1998b). Uncertainty and risk analysis in fire safety engineering.

[Free Web Archade, nd] Free Web Archade (n.d.). 3D Logic Walkthrough. `http://www.freewebarcade.com/help/3d-logic-walkthrough.php`. Accessed: 19.01.2013.

[Gheorghe, 2010] Gheorghe, M. (2010). Non-parametric bayesian belief nets (npbbns) versus ensemble kalman filter (enkf) in reservoir simulation. Master's thesis, Technical University of Delft, Delft.

[Groth and Mosleh, 2012] Groth, K. and Mosleh, A. (2012). A data-informed PIF hierarchy for model-based human reliability analysis. *Reliability Engineering & System Safety*.

[Han and Kamber, 2006] Han, J. and Kamber, M. (2006). *Data mining: concepts and techniques*. Morgan Kaufmann.

[Hanea, 2008] Hanea, A. (2008). *Algorithms for Non-parametric Bayesian Belief Nets*. PhD thesis, Technical University of Delft, Delft.

[Hanea et al., 2010] Hanea, A., Kurowicka, D., Cooke, R., and Ababei, D. (2010). Mining and visualising ordinal data with non-parametric continuous bbns. *Computational Statistics & Data Analysis*, 54(3):668–687.

[Hanea, 2009] Hanea, D. (2009). *Human Risk of Fire. Building a Decision Support Tool Using Bayesian Networks*. PhD thesis, Technical University of Delft, Delft.

[Kurowicka, 2010] Kurowicka, D. (2010). *Dependence Modeling: Vine Copula Handbook*. World Scientific.

[Kurowicka, 2011] Kurowicka, D. (2011). Fault trees and event trees. Unpublished Lecture Notes.

[Kurowicka and Cooke, 2005] Kurowicka, D. and Cooke, R. (2005). Distribution-free continuous bayesian belief nets.

[Kurowicka and Cooke, 2006] Kurowicka, D. and Cooke, R. (2006). *Uncertainty analysis with high dimensional dependence modelling*. John Wiley & Sons Inc.

[Lauritzen and Spiegelhalter, 1988] Lauritzen, S. and Spiegelhalter, D. (1988). Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 157–224.

[Lin, 2011] Lin, P. (2011). *Safety Management and Risk Modelling in Aviation: The challenge of quantifying management influences*. PhD thesis, Technical University of Delft, Delft.

[Los Angeles Times, 2011] Los Angeles Times (2011). Report: Japan, utility at fault for response to nuclear disaster. `http://latimesblogs.latimes.com/world_now/2011/12/japans-march-11-earthquake-and-tsunami-fukushima-daiichi-nuclear-power-plant-meltdown.html`. Accessed: 11.02.2013.

[Marais et al., 2006] Marais, K., Saleh, J., and Leveson, N. (2006). Archetypes for organizational safety. *Safety science*, 44(7):565–582.

[McGlone and Aronson, 2006] McGlone, M. and Aronson, J. (2006). Stereotype threat, identity salience, and spatial reasoning. *Journal of Applied Developmental Psychology*, 27(5):486–493.

[Meadows et al., 1972] Meadows, D., Meadows, D., Randers, J., and Behrens III, W. (1972). *The Limits to Growth: A Report to The Club of Rome (1972)*. Universe Books, New York.

[Mihajlovic and Petkovic, 2001] Mihajlovic, V. and Petkovic, M. (2001). Dynamic bayesian networks: A state of the art.

[Mohaghegh et al., 2009] Mohaghegh, Z., Kazemi, R., and Mosleh, A. (2009). Incorporating organizational factors into probabilistic risk assessment (pra) of complex socio-technical systems: A hybrid technique formalization. *Reliability Engineering & System Safety*, 94(5):1000–1018.

[Mohaghegh and Mosleh, 2009] Mohaghegh, Z. and Mosleh, A. (2009). Incorporating organizational factors into probabilistic risk assessment of complex socio-technical systems: Principles and theoretical foundations. *Safety science*, 47(8):1139–1158.

[Mohaghegh-Ahmadabadi, 2007] Mohaghegh-Ahmadabadi, Z. (2007). *On the Theoretical Foundations and Principles of Organizational Safety Risk Analysis*. PhD thesis, University of Maryland.

[Morales et al., 2008] Morales, O., Kurowicka, D., and Roelen, A. (2008). Eliciting conditional and unconditional rank correlations from conditional probabilities. *Reliability Engineering & System Safety*, 93(5):699–710.

[Murphy, 2002] Murphy, K. (2002). *Dynamic Bayesian Networks: Representation, Inference and Learning*. PhD thesis, University of California, Berkley.

[Mypuzzle.org, nd] Mypuzzle.org (n.d.). mypuzzle.org. `http://mypuzzle.org/3d-logic`. Accessed: 5.12.12.

[National Commission on the BP Deepwater Horizon Oil Spill and Offshore Drilling, 2011] National Commission on the BP Deepwater Horizon Oil Spill and Offshore Drilling (2011). Deepwater: The gulf disaster and the future of offshore drilling. `http://www.oilspillcommission.gov/sites/default/files/documents/DEEPWATER_ReporttothePresident_FINAL.pdf`. Accessed: 12.07.2012.

[Norsys Software Corp., nd] Norsys Software Corp. (n.d.). A. introduction to bayes nets. `http://www.norsys.com/tutorials/netica/secA/tut_A1.htm#WhatIsABayesNet`. Accessed: 14.06.2012.

[Oxford University Press, 2005] Oxford University Press (2005). Paperback Oxford English Dictionary. Oxford University Press.

[Ramsey, 1989] Ramsey, P. (1989). Critical values for spearmans rank order correlation. *Journal of Educational and Behavioral Statistics*, 14(3):245–253.

[Rasmussen, 1997] Rasmussen, J. (1997). Risk management in a dynamic society: a modelling problem. *Safety science*, 27(2):183–213.

[Reason, 1990] Reason, J. (1990). Human error.

[Reason, 1995] Reason, J. (1995). A systems approach to organizational error. *Ergonomics*, 38(8):1708–1721.

[Reason, 1997] Reason, J. (1997). Maintenance-related errors: the biggest threat to aviation safety after gravity? In *Proceedings of the International Aviation Safety Conference: Human Factors, System Engineering, Flight Operations, Economics, and Strategies Management, VSP, Leiden*, pages 465–70.

[Senge, 1990] Senge, P. (1990). *The fifth discipline; the art and practice of the learning organization*. Doudbleday.

[Stewart and Melchers, 1997] Stewart, M. and Melchers, R. (1997). *Probabilistic risk assessment of engineering systems*, volume 2. Springer.

[Swain and Guttmann, 1983] Swain, A. and Guttmann, H. (1983). Handbook of human-reliability analysis with emphasis on nuclear power plant applications. Final report. Technical report, Sandia National Labs., Albuquerque, NM (USA).

[Tetris Holding, 2011] Tetris Holding (2011). TETRIS n-blox. `http://www.freetetris.org/game.php`. Accessed: 08.01.2013.

[Turner and Pidgeon, 1978] Turner, B. and Pidgeon, N. (1978). *Man-made disasters*. Wykeham Publications (London).

[Zichermann and Cunningham, 2011] Zichermann, G. and Cunningham, C. (2011). *Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps*. O'Reilly Media.

Appendix A

Introductory Example for BBNs

This section illustrates the basic concepts of BBNs using a (fictitious) example called *Asia*, which has been introduced by [Lauritzen and Spiegelhalter, 1988].

A BBN reflects all the possible states of (some part of) the world in form of the joint probability of the variables in the model. It is a directed acyclic graph whose nodes represent random variables and whose arcs indicate a probabilistic influence. *Asia* is a BBN with eight binary variables as displayed in figure A.1. The net indicates the influences of smoking and travel to Asia on the lung conditions of a patient. Each variable has two states: yes or no. The variables correspond to the conditions of a patient, e.g. being a smoker or not.

In general, a joint probability distribution embodies all the possible combinations of the states of the variables in a system. A system with N binary random variables has 2^N possible combinations, for which one probability value has to be stored for each. For the model *Asia* that would be $2^8 = 256$ entries. Hence, the computational effort for a reasonably simple model is quite large.

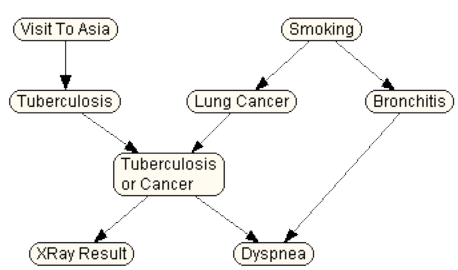


Figure A.1: Nodes and arcs of *Asia* (source: [Norsys Software Corp., nd])

BBNs are a tool to significantly lower the efforts by reducing the number of necessary parameters (probability entries) [Norsys Software Corp., nd]. The arcs point from the so-called parent node to the child node indicating a causal relationship. For instance, *visit to asia* is the parent of *tuberculosis* or, phrased differently, *tuberculosis* is the child of *visit to asia*, which means that a *visit to asia* has an influence on the probability of having *tuberculosis*. Similarly, smoking increases the risk for lung cancer and bronchitis while a trip to asia increases the risk of getting tuberculosis. Both tuberculosis and lung cancer may cause an abnormal X-ray, but bronchitis does not. "Acyclic" means that the arcs may from loops, but no cycle.

Depending on the number of arcs, the required number of parameters may be exponentially less [Murphy, 2002]. Because nodes are probabilistically related through dependence assumptions, only the possible combinations between parent and child nodes have to be stored and used.

Asia is a simplified version of a model which could help to diagnose patients arriving at a clinic for lung diseases. The following description of this model gives an intuitive description of how BBNs are used. It is based on [Norsys Software Corp., nd]. The following facts about the patients reporting to the chest clinic are assumed to be known. (They could be derived easily from experience when running a chest clinic):

- 1% has recently been to Asia and 50% smokes
- 1% has tuberculosis, 5% have lung cancer and 45% have bronchitis
- 11% have an abnormal X-ray result and 44% experience dyspnea

These probabilities describe a new patient[1] arriving at the clinic. Hence, no specific knowledge for him is available yet. For this reason they are called *prior distributions*. Figure A.2a displays the BBN that is specified according to these numbers.

When acquiring new knowledge about the patient, e.g. by questioning the patient or from medical examination, the probabilities in the net are adapted and a more accurate estimate of the likelihood for an illness is obtained. Updating is based on Bayes rule

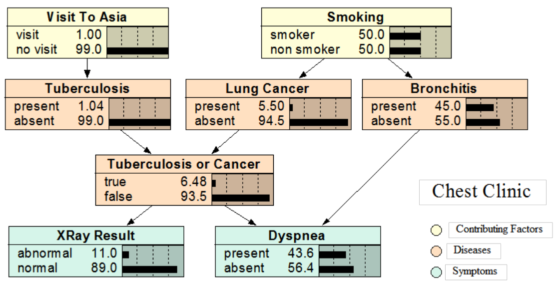
$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \tag{A.1}$$

which can be extended to more dimensions. The probabilities after updating are called *posterior distributions*. An example for acquiring and using new information of one particular patient is given in the following paragraphs.

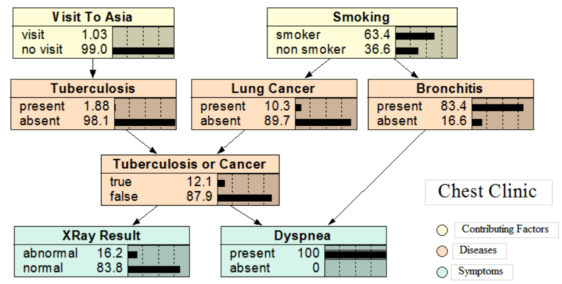
- A patient who reports dyspnea has an increased probability for all three diseases (figure A.2b). This is intuitive, because all three diseases have dyspena as a symptom and observing the symptom increases the belief in the presence of the illness.
- A recent visit to asia increases the probability for tuberculosis, but decreases the risk for lung cancer and more significantly for bronchitis (figure A.2c). This is due to the fact that the diseases are competing causes for the symptom dyspnea: If the probability for one cause increases, the probability for the other causes must decrease. This phenomenon is called "explaining away".
- Smoking increases the probability for bronchitis and lunc cancer further (figure A.2d)
- Finally a normal X-ray indicates a bronchitis with 92% (figure A.2e), whereas an abnormal X-ray indicates relatively high chances for tuberculosis and lung cancer (figure A.2f). In practice this would probably require more specific medical tests, whereas a patient could simply be diagnosed with bronchitis when having a normal X-ray result.

Asia is a typical diagnostic net. It shows that BBNs are very adaptable to the amount of knowledge that is available. In other words: complete knowledge is not required when working with BBNs. If the knowledge is limited, a preliminary model can be build and expanded as

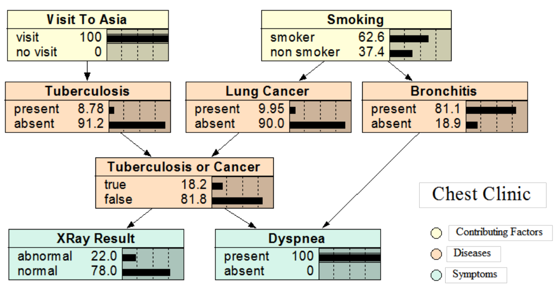
[1] Note that they are not valid for a general population, but only for the patients of the chest clinic. Keeping this assumption regarding the sample population in mind is important in order to draw valid conclusions from the network.



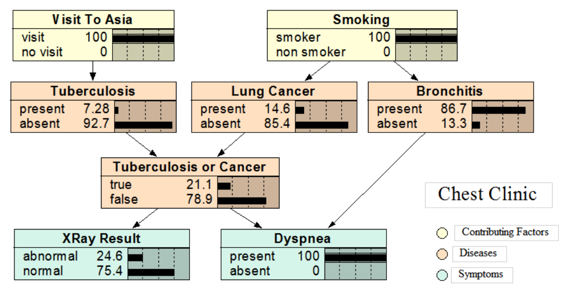
(a) Prior distributions



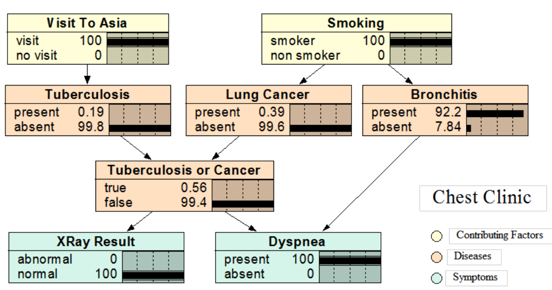
(b) Patient suffers dyspnea



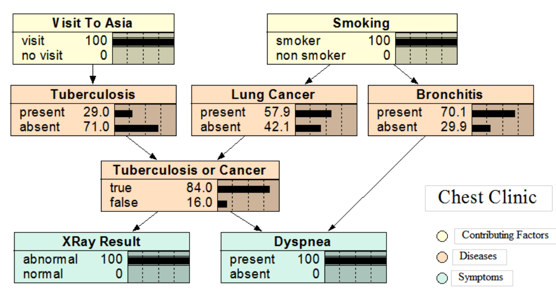
(c) Patient has recently been to asia



(d) Patient is smoker



(e) Patient has a normal X-ray



(f) Patient has an abnormal X-ray

Figure A.2: Prior and posterior distributions for *Asia* (source: [Norsys Software Corp., nd])

new information is acquired. With new information on one or more variables, the probability distributions of the other variables are updated.

A simple example of a static BBN with discrete random variables has been discussed in this section. Nevertheless, BBNs can be dynamic and have continuous random variables.

Appendix B

Selected SD modules from SoTeRiA

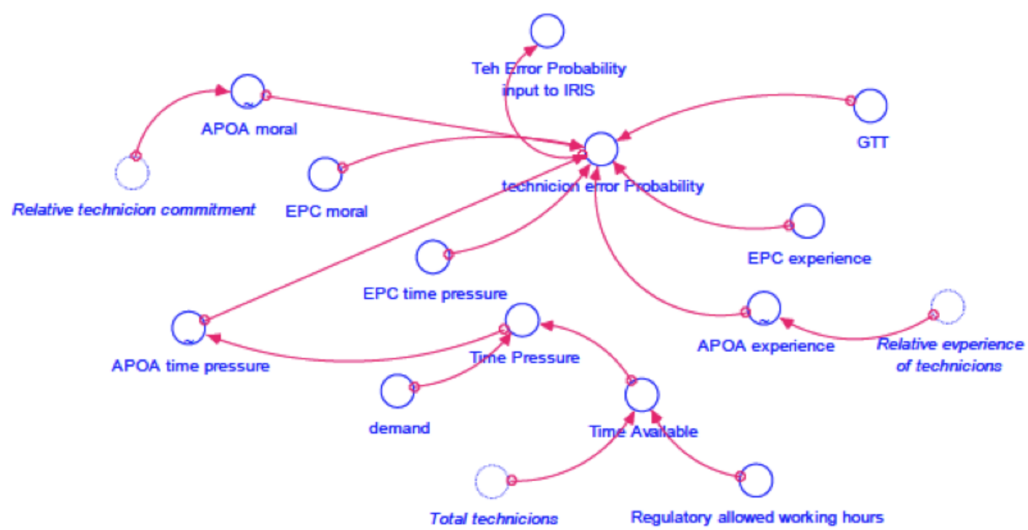


Figure B.1: Human reliability module in a System Dynamics environment (source: [Mohaghegh et al., 2009])

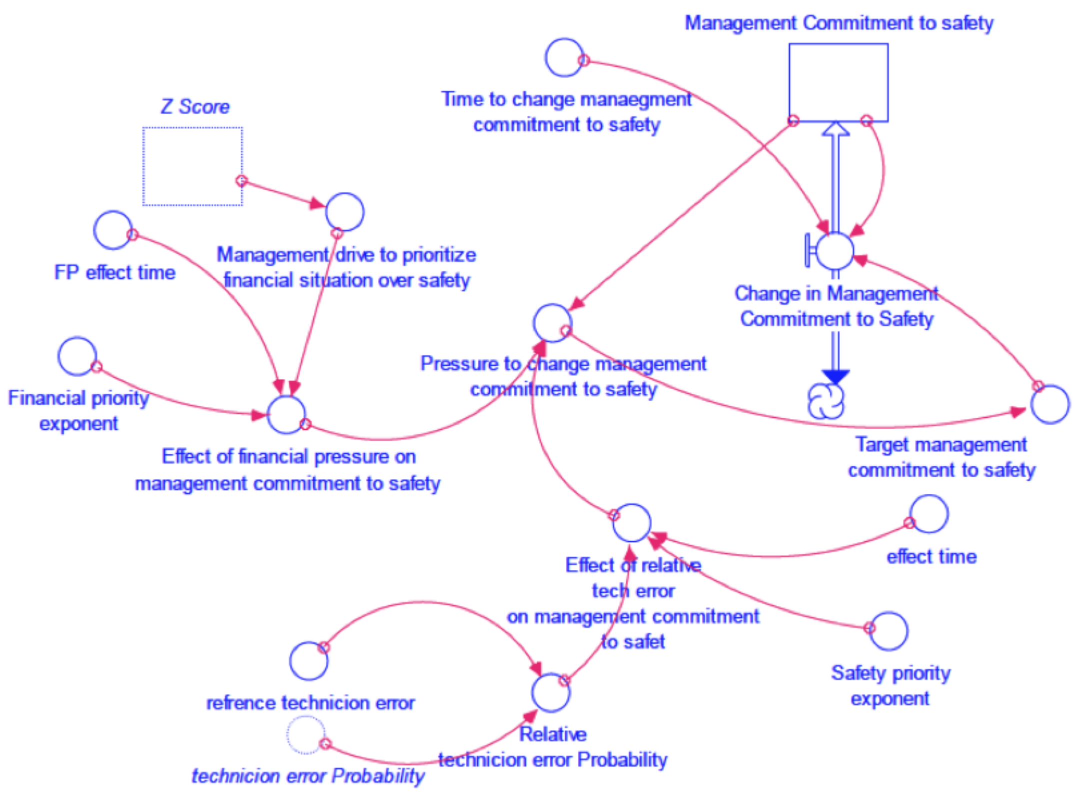


Figure B.2: Management commitment module in a System Dynamics environment (source: [Mohaghegh et al., 2009])

Appendix C

Participants and Questionnaires

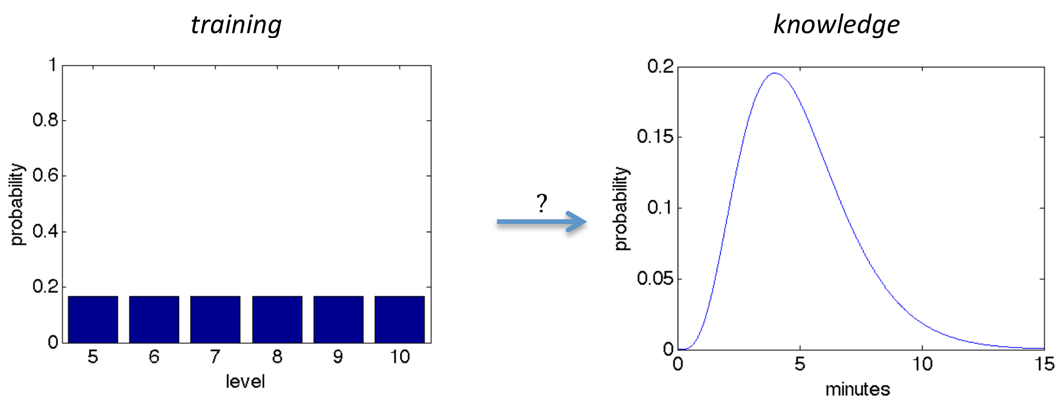
The following pages include a table with all participants and their affiliations as well as the hand out and questionnaire that have been handed out before and after the experiment.

Table C.1: Participants and their answers to the questionnaire.
**These participants indicated that they know a similar game. Due to the small sample size and because their results do not differ particularly from the others, they have been kept in the sample.*

Participant	Q1	Q2	Q3	Q4	Q5	Q6	Q7
1	m	M.Sc. Transport, Infrastructure and Logistics	no	4	3	3	4
2	m	M.Sc. Transport, Infrastructure and Logistics	no	3	1	3	4
3	m	M.Sc. Transport, Infrastructure and Logistics	no	3	1	2	4
4	f	M.Sc. Transport, Infrastructure and Logistics	no	3	3	4	2
5	f	M.Sc. Transport, Infrastructure and Logistics	no	4	4	3	3
6	m	M.Sc. Systems Engineering, Policy Analysis and Management	yes*	4	1	2	6
7	m	M.Sc. Systems Engineering, Policy Analysis and Management	no	4	1	3	4
8	m	M.Sc. Systems Engineering, Policy Analysis and Management	yes*	4	2	1	4
9	m	M.Sc. Systems Engineering, Policy Analysis and Management	no	2	1	2	3
10	m	M.Sc. Transport, Infrastructure and Logistics	no	3	2	2	3
11	m	M.Sc. Systems Engineering, Policy Analysis and Management	no	3	1	2	3
12	m	M.Sc. Systems Engineering, Policy Analysis and Management	no	1	3	2	3
13	m	M.Sc. Engineering and Policy Analysis	no	1	2	2	2
14	f	M.Sc. Civil Engineering	no	3	1	3	3
15	m	M.Sc. Engineering and Policy Analysis	no	1	1	2	3
16	f	M.Sc. Transport, Infrastructure and Logistics	no	4	1	1	3
17	f	M.Sc. Transport, Infrastructure and Logistics	no	4	1	2	2
18	f	M.Sc. Architectural Engineering	no	3	2	3	4
19	f	M.Sc. Architectural Engineering	no	2	2	3	3
20	f	M.Sc. Architecture	no	3	2	1	5
21	m	B.Sc. Commerciële economie	no	5	4	4	1
22	f	M.Sc. Mechanical Engineering	no	5	3	4	3
23	f	M.Sc. Engineering and Policy Analysis	no	3	3	4	3
24	f	M.Sc. Mechanical Engineering	no	4	2	4	4

The Experiment:

- Determining the distribution of *knowledge* transferred through *training*



3Dlogic Instructions:

Rotate the magic cube with your mouse and connect fields with the same color. You have completed a cube, if all color pairs are linked. You cannot connect cells diagonally and black cells are blocked. You can clear the complete cube and undo the last step with the buttons on the bottom. Have fun!

General Remarks:

- Please do not communicate with the other participants during the experiment.
- The levels are not equally difficult. Especially in the first round the times you need for each level may vary significantly.
- If the students around you seem to be finishing much faster than it is probably because you have to play up to a higher level.

Questionnaire (part 1):

1. **What is your gender?**

   Female ☐ Male ☐

2. **In which study program are you (or have you been) enrolled?**

   ____

3. **Do you know the game 3Dlogic?**

   Yes ☐ No ☐

4. **How important is performing "well" at this game to you?**

   Extremely important ☐

   Very important ☐

   Moderately important ☐

   Slightly important ☐

   Not at all important ☐

Procedure of the Experiment

Round 1:

1. In two separate windows open www.online-stopwatch.com and http://mypuzzle.org/3d-logic.
2. Arrange the two windows next to each other.
3. Click on "stopwatch". As a trial run start the stopwatch and pause it after 10 seconds.
4. Draw a number between 5 and 10 from the bowl and note it down as level in the table below.
5. Start the stopwatch, directly start playing 3Dlogic until you have completed your assigned level and pause the stopwatch.
6. Please **do not continue playing** and wait until the experimenter has written down your results.
7. Clear stopwatch

Rounds 2-4:

Repeat steps 4 to 7 of the first round, that is:

4. Draw a number between 5 and 10 from the bowl and note it down as level in the table below.
5. Start the stopwatch, directly start playing 3Dlogic until you have completed your assigned level and pause the stopwatch.
6. Please do not continue playing and wait until the experimenter has written down your results.
7. Clear stopwatch and press exit of 3Dlogic.

Results

Round	Level	Time
1		
2		
3		
4		

Questionnaire (part 2)

4. **How difficult did you find 3Dlogic?**

   Extremely difficult ☐

   Very difficult ☐

   Moderately difficult ☐

   Slightly difficult ☐

   Not at all difficult ☐

5. **Did you feel stress/pressure during the experiment?**

   Extremely ☐

   Very much ☐

   Moderately ☐

   Slightly ☐

   Not at all ☐

6. **How do you think you performed compared to other participants?**

   Much better ☐

   Somewhat better ☐

   About the same ☐

   Somewhat worse ☐

   Much worse ☐

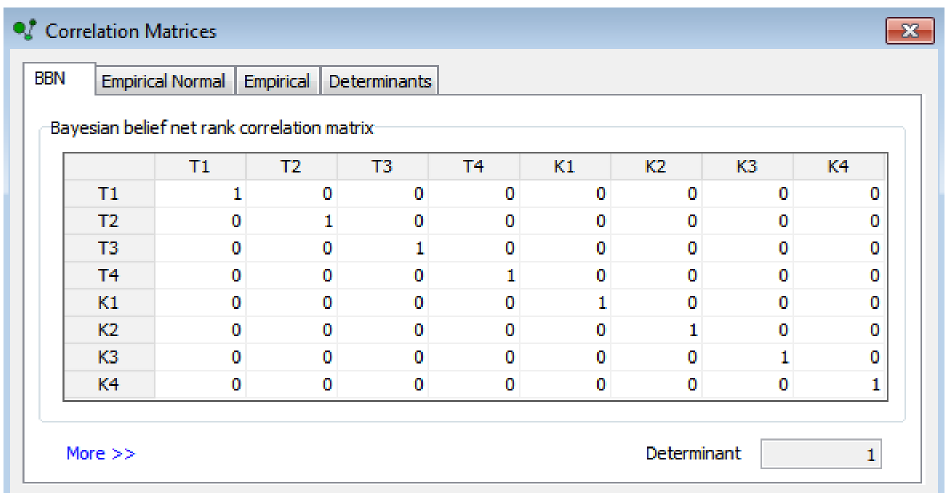
Appendix D

Validating the BBN Structure

This appendix relates to section 6.2. It contains the results of several statistical tests to validate the model structure.

D.1 Results for the Skeletal BBN

This appendix refers the skeletal BBN in figure 6.1b. It provides the correlation matrix and the output of the statistical test.


Correlation Matrices

BBN | Empirical Normal | Empirical | Determinants

Bayesian belief net rank correlation matrix

	T1	T2	T3	T4	K1	K2	K3	K4
T1	1	0	0	0	0	0	0	0
T2	0	1	0	0	0	0	0	0
T3	0	0	1	0	0	0	0	0
T4	0	0	0	1	0	0	0	0
K1	0	0	0	0	1	0	0	0
K2	0	0	0	0	0	1	0	0
K3	0	0	0	0	0	0	1	0
K4	0	0	0	0	0	0	0	1

More >>

Determinant 1

Figure D.1: Correlation matrix for the skeletal BBN

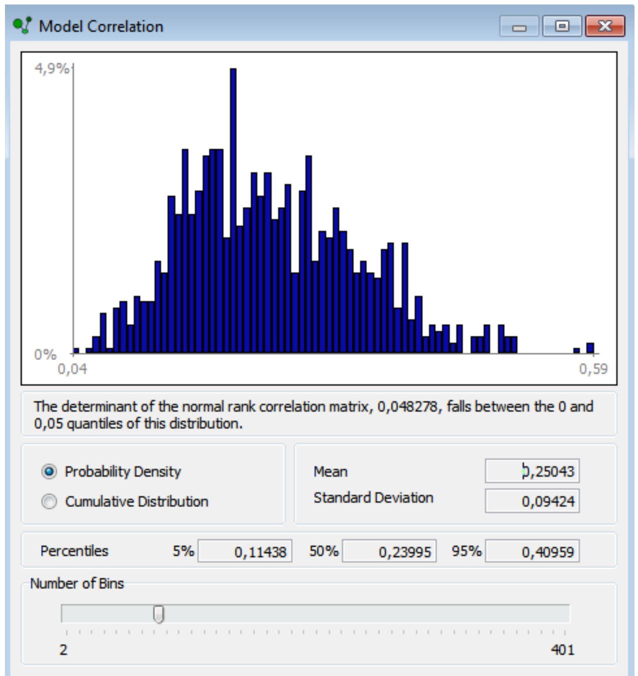
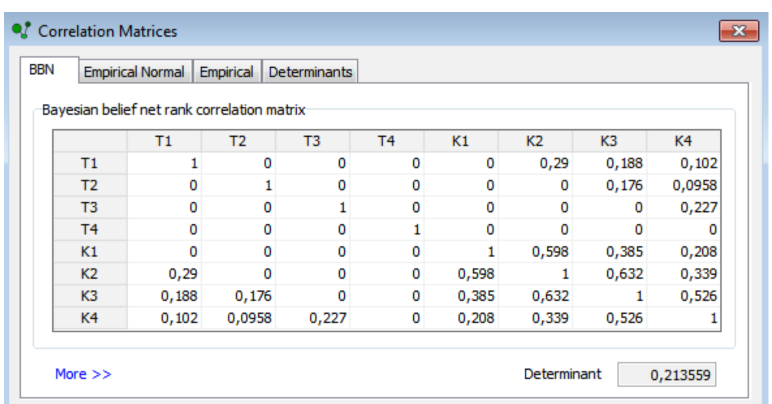


Figure D.2: Results of the "validate model" function of Uninet for the skeletal BBN

D.2 Results for the BBN with few arcs

This appendix refers the BBN structure with few arcs in figure 6.4. It provides the correlation matrix and the output of the statistical test.


Correlation Matrices

BBN | Empirical Normal | Empirical | Determinants

Bayesian belief net rank correlation matrix

	T1	T2	T3	T4	K1	K2	K3	K4
T1	1	0	0	0	0	0,29	0,188	0,102
T2	0	1	0	0	0	0	0,176	0,0958
T3	0	0	1	0	0	0	0	0,227
T4	0	0	0	1	0	0	0	0
K1	0	0	0	0	1	0,598	0,385	0,208
K2	0,29	0	0	0	0,598	1	0,632	0,339
K3	0,188	0,176	0	0	0,385	0,632	1	0,526
K4	0,102	0,0958	0,227	0	0,208	0,339	0,526	1

More >> Determinant 0,213559

Figure D.3: Correlation matrix of the BBN with few arcs

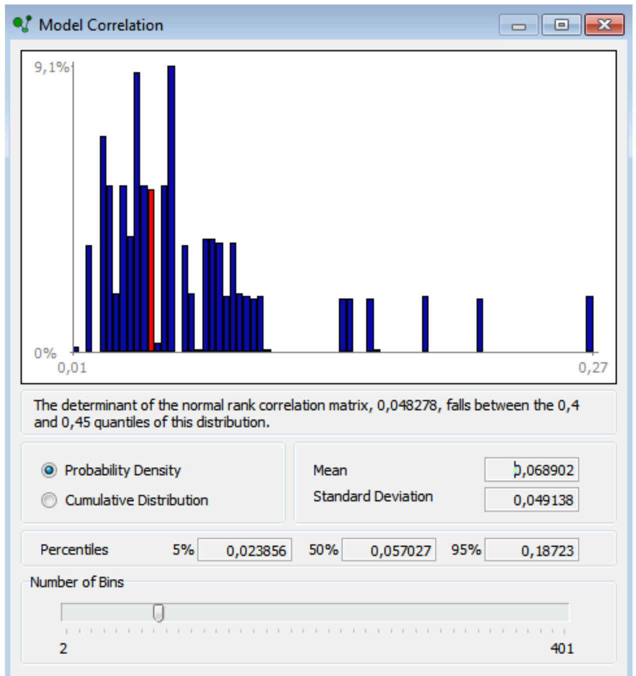
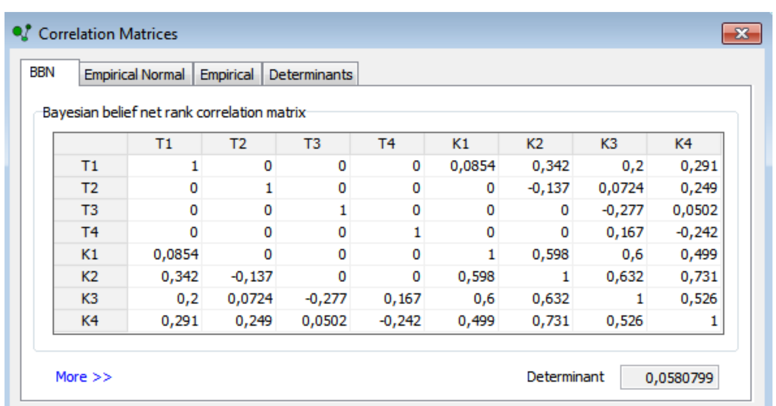


Figure D.4: Results of the "Validate Model" function of Uninet for the BBN with few arcs

D.3 Results for the BBN with many arcs

This appendix refers the BBN structure with many arcs in figure 6.5a. It provides the correlation matrix and the output of the statistical test.



Correlation Matrices

BBN | Empirical Normal | Empirical | Determinants

Bayesian belief net rank correlation matrix

	T1	T2	T3	T4	K1	K2	K3	K4
T1	1	0	0	0	0,0854	0,342	0,2	0,291
T2	0	1	0	0	0	-0,137	0,0724	0,249
T3	0	0	1	0	0	0	-0,277	0,0502
T4	0	0	0	1	0	0	0,167	-0,242
K1	0,0854	0	0	0	1	0,598	0,6	0,499
K2	0,342	-0,137	0	0	0,598	1	0,632	0,731
K3	0,2	0,0724	-0,277	0,167	0,6	0,632	1	0,526
K4	0,291	0,249	0,0502	-0,242	0,499	0,731	0,526	1

More >>

Determinant 0,0580799

Figure D.5: Correlation matrix of the BBN with many arcs

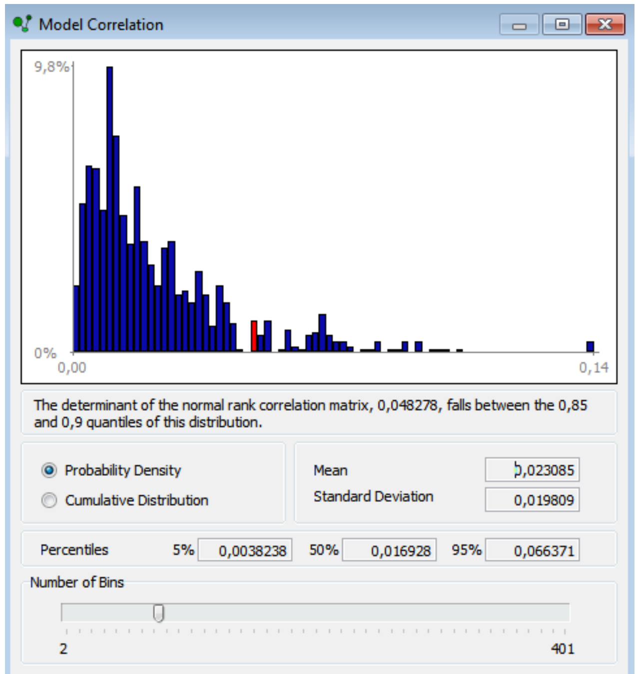


Figure D.6: Results of the "Validate Model" function of Uninet for the BBN with many arcs