Gaming Simulation and Human Factors in Complex Socio-Technical Systems

A Multi-Level Approach to Mental Models and Situation Awareness in Railway Traffic Control

Lo, Julia

DOI
10.4233/uuid:93bba631-f675-4b28-8780-3bd9d686680a

Publication date
2020

Document Version
Final published version

Citation (APA)
Lo, J. (2020). *Gaming Simulation and Human Factors in Complex Socio-Technical Systems: A Multi-Level Approach to Mental Models and Situation Awareness in Railway Traffic Control*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:93bba631-f675-4b28-8780-3bd9d686680a

Gaming Simulation and Human Factors in Complex Socio-Technical Systems:

A Multi-Level Approach to Mental Models and Situation Awareness in Railway Traffic Control

Julia Lo

Gaming Simulation and Human Factors in Complex Socio-Technical Systems:

A Multi-Level Approach to Mental Models
and Situation Awareness in Railway Traffic Control

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof.dr.ir. T.H.J.J. van der Hagen,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op
26 juni 2020 om 12:30 uur

door

Julia Chantal LO

Master of Science in Psychologie, Universiteit van Amsterdam
Master of Science in Communicatiewetenschap, Universiteit van Amsterdam
Geboren te Leiden, Nederland.

Dit proefschrift is goedgekeurd door de promotoren.

Samenstelling promotiecommissie

Rector Magnificus	voorzitter
Prof. mr. dr. J.A. de Bruijn	Technische Universiteit Delft, promotor
Prof. dr. ir. S.A. Meijer	KTH Royal Institute of Technology, promotor
Onafhankelijke leden	
Prof. dr. K.A. Brookhuis	Rijksuniversiteit Groningen
Prof. dr. R.M.P. Goverde	Technische Universiteit Delft
Prof. dr. M.A. Neerincx	Technische Universiteit Delft
Prof. mr. dr. E.F. ten Heuvelhof	Technische Universiteit Delft
Overige leden	
Drs. H. Thomassen	Airport Coordination Netherlands

Dit onderzoek is gefinancieerd door ProRail B.V. onder het Railway Gaming Suite programma.

Drukkerij: Ipskamp printing
Ontwerp omslag: Rosalie Blom & Remco Mouthaan
ISBN: 978-94-028-2078-2

Acknowledgements

It appears that things turn out like they are supposed to in the end. At least in my case. When I was in elementary school, I always answered that I wanted to become an aerospace engineer. However, when the time was there to choose my study, I did a 360° move: I decided to study Psychology. I finished my master studies in Psychology and Communication Science, went to work in industry at Philips on sustainability communication, but then realized that that was not my passion. I continued my path in research to finally become an engineering psychologist, enriched by railway professionals and researchers on gaming & simulation, governance and artificial intelligence amongst others.

This work could not have been realized if it would not have been for my supervisors Hans de Bruijn and Sebastiaan Meijer. Thank you Hans for your sharp eye on the practical relevance of this research and your ways of actually explicitizing my knowledge. Thank you Sebas for your involvement and guidance these years, and for the environment you created to independently grow as researchers and when we needed you - whether it was meeting in Delft or Stockholm or online - you were always available. I look back with warm memories to your home island Nynäshamn with your (research) family (and your lawn mowing tractor).

I have also been inspired and supported by colleagues both at TU Delft and ProRail. While now being employed at ProRail I have been given time to finish the research done during my years at TU Delft. I want to thank Chris Verstegen and Karel van Gils for giving me space to finish 'the last steps'. Here it is finally.

Thank you, Emdzad Sehic, my partner in crime and fellow researcher on the gaming simulation sessions. All our gaming simulations were very intensive in their preparation and execution. I am impressed by your dedication to the (technical) preparations of the gaming simulations and I am proud of our work for many control centers (and a bit less about the number of McDonalds and BK visits). Thank you also Gert-Jan Stolk, Berend Wouda and many other programmers from the Game Lab at TU Delft for your work on the development of the simulator. And thank you, Harrie de Haas and fellow colleagues at the control centers for supporting our research and making the simulation sessions happen.

Special thanks to Jelle van Luipen who recognized the importance of gaming simulations for ProRail and coordinated the Railway Gaming Suite program. Due to your efforts we achieved the current state of the art knowledge and experience with gaming simulations.

I would also thank my fellow colleagues Edith, Esther, Boris, Jurjen, Roderic and Wieke who I continued to work with in the meantime on the professionalization of gaming simulations and human factors at ProRail.

I want to thank my colleagues at the section of Policy, Organization, Law and Gaming (now Organization and Governance) at TU Delft for the inspiration, collaboration and/or just for the 'gezelligheid', particularly Bauke, Bill, Casper, Fabio, Femke, Geertje, Haiko, Heide, Harald, Qasim and special thanks to Igor Mayer for your involvement and guidance in my first year at TU Delft.

Jop and Iman – my PhD buddies all the way from the start. Jop, we shared the same office, were in the same projects and process, talked about new wild theories and collaborated on publications. Through our discussions you've challenged and inspired me. Iman, I remember the very first day we met for our assignment to create a syllabus of gaming simulations. Since then we were inseperatable and that big smile always returns on my face when remembering our conference trips. I'm grateful for the moments (nice and tough ones) we shared during our PhD time, and our ongoing friendship.

The Joint Research on Situation Awareness (JRSA) reading group was also an example of where work could be mixed with fun and friendship. How hours could easily fly by. Thank you Reyhan, Soufiane and Alexei for your intellectual, fun and sportive company. Up to the next adventure on our bucket list of travels.

At conferences I met peer human factors researchers, of which I particularly would like to thank Simon Tschirner, Ellemieke van Doorn, Lenore Page and Ida Löscher for our long and detailed discussions on human factors topics. I'm happy and grateful for our collaboration Lenore. Even though we live that far apart I hope we can stay in touch. Ida, from Crete to Groningen to Los Angeles – it did not matter where or at which conference we were to have a blast of a time together. I'm looking forward to our next conferences and visits.

I am indebted to Arie for supporting me in the last steps of the dissertation efforts by (proof) reading the final chapters, thank you so much.

For distractions and pokes on the current status of my thesis I could also turn to Caroline, Jolanda, Grace and Rosalie. Special thanks also to Rosalie and Remco for diving in the world of cover design and providing the finishing touch to this work.

I would also like to thank James and Eva for the fun and comforting family dinners in the weekend and my parents who always believed and supported me in my choices.

Last but not least Thomas, I think that words fall short to thank you for your care, dedication and patience during this part of our life. I am grateful to have you by my side on this entire journey.

Table of Content

1 Introduction

1.1 Problem statement

Societies are built from networks of sectors that fulfill their basic needs. Healthcare, energy and transportation are examples of such sectors and each one of these has their own historical development and challenges. However, they are all bound by their characteristic as being socio-technical systems nowadays: they are strongly engineering-oriented through their reliance on technological equipment, but simultaneously also heavily rely on the involvement of human operators. The railways are one of these so-called socio-technical systems, comprising of engineering activities for its infrastructural elements, machinery and automation while its maintentance, driving and control operations rely on human activities.

After many decades of straightforward focus on maintenance and utilization of the railways, the privatization of the railway sector triggered a transformation in many countries around the 1990s (Knieps, 2013). With changed institutional governance in terms of a split between infrastructure and transportation management, the railways are worldwide facing a higher demand in passenger and freight demands which coincide with capacity constraints (e.g. Lai & Barken, 2009).

Similarly, ProRail, the Dutch railway infrastructure manager in the Netherlands, has set goals in collaboration with the Dutch government to increase the capacity of the infrastructure in order to provide for the increasing number of passengers nationwide and cargo from the port of Rotterdam. The capacity increase on the railways comes with a number of challenges.

Firstly, ProRail faces the inability of solving capacity constraints by solely physical infrastructural expansion. Due to a highly densely built environment in many parts of the Netherlands, space-related solutions are not a straightforward option. Other types of solutions are needed to facilitate the increase in capacity: allowing more trains on the infrastructural network, which can also be realized instead by process optimalizations or digitalization. These solutions are sought in short-term and long-term (e.g. over eight years) implementation timelines.

Another challenge lies within the organizational culture and management. After decades of no radical changes, the rigorous nature of the redesign of the railway system requires a new approach where the conservative and institutionalized way of working is challenged and can no longer be substantiated. Given the invasiveness of the changes, projects are more intertwined and collaboration between projects is important for their success (Van den Hoogen, 2019).

A third challenge relates to the impact of the large-scale changes and complexity of these process optimalizations and digitalization solutions on the task-space of railway traffic operators. The involvement of human operators in the railway system introduces an element of uncertainty about the predicted performance of the railway system, as human operators may exhibit individual differences, such as in their experience and skills. As such, it is necessary to understand what the implications of the planned changes are on the cognitive processes of the train traffic and network operators and their level of diversity, in order to deal with the increased pressure on the resilience of the railway system. Investigating the cognition of railway traffic operators could provide more insight into their reasoning and the quality of their decisions, which in turn could support improved compatible redesigns in processes or optimalizations. As such a basic understanding of the cognition of railway traffic operators would be beneficial and a valuable starting point in the (re)design of this complex socio-technical system. The studies in this dissertation will mainly focus on investigations relating to this particular challenge with the use of state-of-the-art tools such as gaming simulation.

As will be elaborated upon further on in this chapter, gaming simulation can be seen as a simulation in which human participants take part. Throughout this work, gaming simulation will be used in a broad sense. They may exist in many forms (e.g. analog or digital) and purposes (e.g. for research, training etc.). The current studies in Chapters 4 to 8 focus on analog (i.e. tabletop) and digital (i.e. simulator) gaming simulation that can be employed to simulate alternative modes of the system. Theoretical implications of the use of gaming simulation as a research environment and research tool are another focal point in conducting psychological research. The cognitive concepts of mental models and situation awareness (SA) are particularly investigated as, in short, mental models serve as a fundament for SA and in turn, SA is a predictor of the quality of decision-making of operators in complex socio-technical environments. These characteristics make these concepts especially interesting to investigate. Thus, two fields will be elaborately covered in this dissertation: (1) gaming simulation and (2) human factors research on mental situation awareness on multiple units of analysis, i.e. individual, team and network level.

A theoretical background of several fields are discussed in the following sections: section 1.2 provides a brief introduction of the railway sector and its characteristics for the Dutch railways, followed by section 1.3 on gaming simulation including their use for participatory design, section 1.4 focuses on the cognitive components of mental models and situation awareness, section 1.5 focuses on the system theory and specifically the perspective of complex adaptive systems, and section 1.6 elaborates on the role of agent-based modeling in relation to the complex adaptive systems perspective and its role in an operational context. Finally, section 1.7 introduces the research questions and chapter outlines in this dissertation.

1.2 The Dutch railways

In describing the characteristics of the Dutch railway sector, different approaches can be applied. To support a uniform approach, the following framework is applied (see Table 1.1) as proposed by Golightly, Sandblad, Dadashi, Andersson, Tschirner, and Sharples (2013), which was developed to compare designs or deployment decisions in train traffic control across countries. In the third column the different characteristics of the framework are described for the Dutch railway system.

A description of railway traffic control and the different roles can also be found in Chapter 7.

Table 1.1: Characteristics of the Dutch railway system[1] based on typologies by Golightly et al., (2013).

Characteristic	Description	Dutch railway system
National characteristics	Density, complexity, service and performance context; organisation of the railways.	7097 kilometers of track, 27 passenger and freight train operating companies, 6661 switches, 24 tunnels, 2585 level crossings, 11622 signals, 400 stations, ProRail as Dutch railway infrastructure manager, Nederlandse Spoorwegen (NS) as principle train operating company, trains every 15 minutes between large cities.
Organization	Centralization vs. decentralization, work organisation, division of control tasks between different roles.	13 regional traffic control centers, with one main national control center; three operator roles: train traffic controller (TTC), regional network controller (RNC) and a national network controller (NNC). Control centers with large and complex geographical areas often have a TTC planner, who is responsible for keeping the timetable up-to-date until 15 min prior to execution.
Roles	Structure and relations between different roles. Work processes and control tasks for each role.	TTC and RNC co-located, depending on control center: a TTC planner and one or two RNCs.

[1]The terminology in the current work may vary and deviate across chapters from common English railway terminology. The following terminologies are used for the railway infrastructure manager (IM): railway infrastructure manager, agency or organization. For railway undertakings (RU) or train operating companies (TOC): passenger transport manager or train service providers. To distinguish the traffic control centers of the IM and RU/TOC: control centers of the IM are called railway traffic control, control centers of the RU/TOC are called: passenger traffic control or passenger traffic service organization. The term train traffic controller is used for the combined role of signaler and dispatcher.

Communication	Communication patterns and channels between different roles in the control process, e.g. other control roles, train drivers, railway undertakers etc.	TTC communicates with train driver, TTC planner and RNC, RNC communicates with TTC, TTC planner and NNC. NNC communicates with RNC.
Technology	Type of signalling and safety system, traffic control system, switch box technology, interlocking system, train protection system etc.	Traffic management systems (TMS) for all TTCs and RNCs between all control centers. Predominant signalling system is NS54/ATB (>90%) in 2019.
Automation	Structure and complexity of automation. Single automatic systems or a complex structure. Interaction between different automatic systems. Different modes of automation. Control-by-awareness or Control-by-exception	TMS system: automatic route setting for TTCs.
Interfaces	Observability - are the automatic functions and their actions transparent and easy to understand? Controllability - possibilities for turning on/off, changing modes, re-programming etc. Representation - schematic versus train graph.	Tabular timetable for TTCs and TTC planners, platform occupancy graphs for TTCs and TTC planners, dynamic distance-time graphs for RNCs and NNCs.

1.3 Gaming simulations

1.3.1 Definition

Throughout the past decades games have also been widely known as 'serious games', 'simulation games', 'gaming simulations' and many more synonyms. It is stated that the varying terms are caused by a different emphasis focusing on the functionality or the artifact of certain forms (Klabbers, 2009; Narayanasamy, Wong, Fung, & Rai, 2006). An example of such a debate is whether simulators are considered to be a type of gaming simulation.

In line with the recognition of these connotations, a broad perspective of gaming simulations (in short 'games') is taken upon as simulations with human participants, in which game design principles are applied. As such, the function of the simulation of a reference system is predominantly emphasized. To exemplify this with the railway domain as an application, the gaming simulation as a simulated part of the system reflects a part of the subsystem, which in turn is only a segment of the real world system, i.e. the railway system (see Figure 1.1). For instance, the railway system as a whole can be captured with non-human and human artefacts, such as its tracks, switches, train traffic control centers, operators, freight trains, passenger trains, train drivers etc. However, when investigating the railway system only certain parts of the railway system can be investigated to ensure a clear and focused research question. As such only a part of the real-world system can be investigated; one or more operators in a control center, the interaction between train driver and train traffic controller; a single individual train traffic controller etc. When designing a gaming simulation as a simulated environment of the identified subsystem, different approaches can be taken in the translation of physical and procedural characteristics. For instance, in the case of simulating an entire control center, the gaming simulation design can be operationalized in different manners: the entire simulated environment can be designed to be identical in its appearance, tools and people. An alternative design could be to design the operators' system identical to that in the reference system, i.e. the functionalities and interfaces of operators are fully similar to those in the real world system. Another alternative design could be that only parts of an operator's system are included in the simulation or are represented in a different but isomorphic manner.

The physical translation as part of the gaming simulation design from the subsystem as reference system touches upon the representation level, also known as the fidelity level of the game environment. As such, low-tech but also high-tech environments can be distinguished (Meijer, 2015). In Figure 1.1, an illustration is provided of a simulated system. At the top, an individual train traffic operator is depicted who uses a simulator as a high-tech simulated environment. To the bottom, a tabletop environment is depicted with multiple operators used to simulate a larger part of the railway system using a low-tech simulated environment.

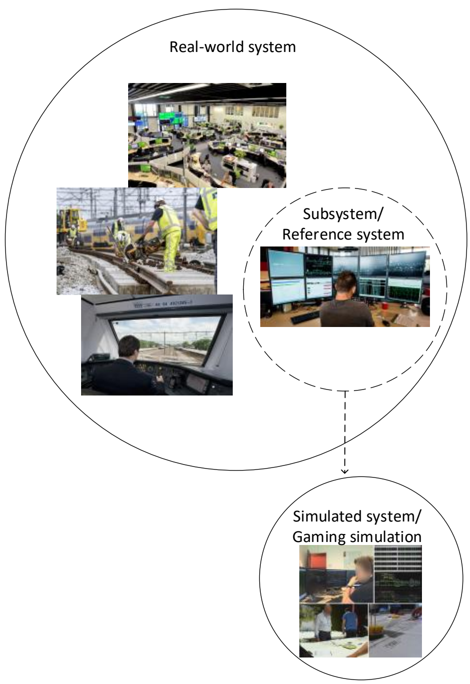


Figure 1.1: Creating simulated environments through an extraction of a subsystem of the real world.

Given this context of a large diversity of types of simulation environments, a broader definition of gaming simulation is applied throughout this book, covering human-in-the-loop simulators and tabletop simulation environments in a socio-technical system such as the railway sector.

1.3.2 Gaming simulation design principles

In addition to to the presence of human participants in the simulated environment as a characteristic of a gaming simulation, another element that is reflected in the current definition is that gaming simulation design principles can be identified. As such, the gaming simulation design looks into the explicit translation of the roles, rules, objectives, constraints, load and situation (see also Figure 1.2 by Meijer, 2008). Different configurations can be realized based on the design choices, with different outcomes of the gaming simulation. Figure 1.2 also illustrates two dimensions by ascribing each to the analytical science and design science. Analytical science refers to the research purpose of games (Klabbers, 2006), while games for training and intervention purposes reside under the umbrella of design science, which focuses on a change of participant(s) or an organization through experiences in the game session.

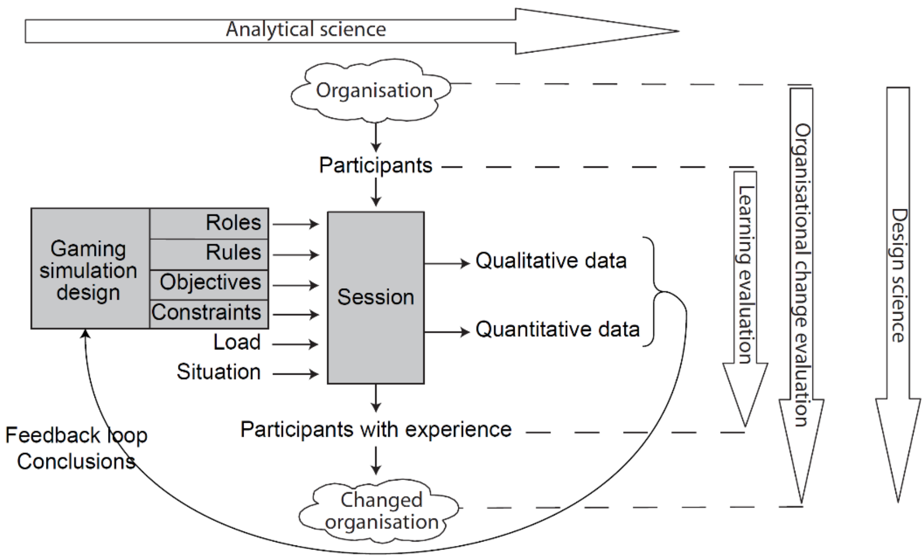


Figure 1.2: Meta-framework for design and analytical science (Meijer, 2008).

In creating a simulated environment that is based on its reference system, the notions of design-in-the-large (DIL) and design-in-the-small (DIS) acknowledge the relevance of the context-in-use (Klabbers, 2006; Van den Hoogen, 2019). In the context of the Dutch railways, gaming simulation has been used in the light of process innovations, which can be acknowledged as the DIL. Herein, the design of a gaming simulation i.e. the simulated environment can be seen as the DIS. Depending on the goal and approach of the gaming simulation session, which can be to test hypotheses or to improve change processes, different perspectives are held, influencing the relations between DIS (i.e. the gaming simulation) and DIL (i.e. the innovation process) (see Figure 1.3).

	Analytical Science	Design Science
Function	Test hypotheses	Improve change processes
Assessment on	Validity and Reliability	Usability and Credibility
DIL vs DIS	DIS linearly follows DIL and vice versa -> they are part of the same cyclical model	DIS and DIL are separated (and temporal) cycles that interact in more complex ways
Underlying model of innovation	Linear model of innovation, cognitive problem-solving, Quality of innovation can be determined objectively, from the outset No relevant context, innovation is an isolated process Timing is irrelevant	Non-linear model of innovation, Socio-institutional dynamics, Parallel processes Quality of innovation is socially constructed, during the process Context is relevant Timing is relevant because of multiple parallel and interdependent processes

Figure 1.3: Relation between a gaming simulation as design-in-the-small (DIS) and process innovation as design-in-the-large (DIL) from two perspectives (Van den Hoogen, 2019).

Different gaming simulation types can be connected based on this framework, which will be further introduced and discussed in Chapters 2, 3 and 11.

1.3.3 Gaming simulation as a tool in a participatory system design

The involvement of various stakeholders, e.g. users and/or designers, in the design and research of a product, artifact, or process is a core principle in the participatory design field (Muller, 2003; Spinuzzi, 2005). In the design process different tools and techniques can be applied, in which gaming simulation is one of the context-oriented tools (Kensing & Blomberg, 1998; Muller & Kuhn, 1993).

In the use of gaming simulation as part of a change implementation, two approaches can be distinguished: the programmatic view and the participatory view (Russ, 2010). The programmatic view follows a top-down manner of communication on the changes that are designed and planned by management. Gaming simulation is used as a tool to convince employees of these changes. The participatory view on gaming simulation focuses on dialogical communications. In line with this view, gaming simulation is used as a tool to create a common ground, in which employees can develop a shared understanding, are able to co-design and support consensus in the decision-making. The current work takes upon a participatory view on gaming simulation. Chapters 8 and 10 will address this topic.

1.4 Human factors/cognitive engineering

The field of human factors and ergonomics itself is interdisciplinary in nature, covering disciplines such as psychology and engineering. It can be characterized by interactions between the humans and technology in socio-technical systems (Karwowski, 2012; International Ergonomics Association, 2003). Human factors research recognizes different methods and techniques to study the human nature in engineering systems, which can draw on experimental, descriptive and evaluation research (e.g. Jacko, Yi, Sainfort, & McClellan, 2012).

The cognition of human operators has many facets that can be investigated. In line with the scope of the problem statement, i.e. to understand the cognition of operators in gaming simulations, the focus is particularly on two cognitive constructs that are profound in complex and dynamic command and control operation: mental models and situation awareness.

1.4.1 Mental models

Mental models are deemed important as they serve as knowledge structures, in which an individual's representation of a physical system can be described (Endsley, 2000; Klimoski & Mohammed, 1994; Mathieu, Heffner, Goodwin, Salas, & Cannon-Bowers, 2000). They have been frequently used by researchers to explain individuals' cognitive functioning and performance (Salas, Stout and Cannon-Bowers, 1994). Moreover, mental models are assumed to be a fundamental mechanism for the establishment of situation awareness (SA), as without a well formulated mental model, attention is not directed to certain cues and thereby operators might oversee certain elements in the environment (level 1 SA), might not be able to establish a good comprehension of the situation (level 2 SA), or might be unable to make a good prediction of future states (level 3 SA) (e.g. Endsley, 1988). Subsequently, a high situation awareness is expected to be a predictor of good decision-making in operational settings.

However, operators in complex socio-technical systems hardly execute tasks in a solitary environment. On the contrary, the high interdependence between actors often defines and shapes the complexities of these systems, emphasizing the importance of team knowledge for operators' cognition. Team mental models are also known as common cause maps, teamwork schemas, shared frames, socio-cognition, transactive memory (Klimoski & Mohammed, 1994). Mental models in teams have been predominantly measured through the accuracy and similarity between team members (Mohammed, Ferzandi, & Hamilton, 2010). As a result of the need to compare individual mental models, the mental model construct has been operationalized to task-team types of mental models and knowledge structure models.

One type of conceptualization of mental models can be distinguished in terms of Equipment/technology, Task/job, Team-interaction and Team (ETTT) models (Cannon-Bowers, Salas, & Converse, 1993; Lim & Klein, 2006; Matthieu et al., 2000). The equipment/technology model is related to the technology and equipment that is used to execute tasks in a team. This also involves indirect interaction, such as changing the direction of railway switches through computerized systems. The equipment/technology model is the most stable of the four types as the components in this model do seldom change. The task/job model is related to the perception and understanding of procedures, strategies and so forth, in which operators need to understand the ways of how to accomplish their task, e.g. necessary information and procedures. The influence of environmental conditions on the task and task demands, such as changed weather conditions or sudden peaks in passenger flow, are also part of the task/job model. Thirdly, the understanding of the responsibilities, norms and interaction patterns of other team members is part of the team-interaction model. Procedures, such as which team members need to interact with each other, what kind of particular information is needed, but also knowledge when to help team members are also knowledge contents related to the team-interaction model. Finally, the team model is related to the understanding of knowledge, preferences, skills, attitudes, strengths and weaknesses of other team members. The team model has a rather low model content stability due to frequent changes in teams, e.g. as railway traffic operators work in shifts, they often need to collaborate with different colleagues. Thus, as they might not work together in the same team configuration for a long period of time, team members develop their knowledge about the abilities, preferences etc. of their colleagues more slowly. Table 1.2 illustrates the different knowledge types and knowledge components from the railway domain.

Table 1.2: Conceptualization of mental models in terms of Equipment/technology, Task/job, Team-interaction and Team (ETTT) mental model conceptualization (Cannon-Bowers et al., 1993; Matthieu et al., 2000).

Type	**Knowledge contents**	**Railway knowledge components**	**Stability of the model content**
Equipment/ technology model	Equipment functioning, operating procedures, equipment limitations, likely failures	Network layout, such as railway tracks, switches, signals, computerized systems, such as the PRL (train traffic management) system, dynamic timetable interface	High
Task/job model	Task procedures, task strategies, environmental conditions, likely contingencies, likely scenarios	Task procedures, such as the role dependent operating procedure, TAD (train order protocol); environmental conditions, such as the weather	Moderate
Team-interaction model	Roles/responsibilities, role interdependencies, information patterns, information sources, communication patterns	Roles as defined in the operating procedure	Moderate
Team model	Knowledge over teammates' knowledge, skills, abilities, preferences, tendencies	Team configurations as in planned working shifts	Low

A second type of operationalization of mental models can be realized through knowledge structures, i.e. declarative, procedural and strategic (DPS) mental models (see Table 1.3) (Mohammed et al., 2010; Salas et al., 1994). Declarative models refer to knowledge of facts, rules and relationships (knowledge of what). Procedural models refer to the timing and sequential type of knowledge (knowledge of how). Strategic models refer to knowledge that forms the basis for problem solving (knowledge of the concept and contingency plans). In relation to the ETTT mental model conceptualization, these three types of knowledge can be applied to one single knowledge content. For example, declarative knowledge can be related to facts and rules of a railway switch (e.g. single slip, double slip, outside slip). Procedural knowledge of a railway switch is related to how a railway switch works and how it can be operated. An illustration of strategic knowledge is using a specific switch to reroute a train to a different railway track if the original/planned railway track for that train is blocked, knowing that no other train is currently using the alternative railway track.

Table 1.3: Conceptualization of mental models in terms of knowledge structures (DPS mental model conceptualization) (Mohammed et al., 2010; Salas et al., 1994).

Type	Definition	Knowledge contents	Example
Declarative model	Information about concepts and elements, and their relationship Knowledge of what/knowing that	Facts, rules, relationship, knowledge about the overall system task goals, the relation among system components, equipment/hardware, position/roles, and the team members themselves	Umbrella: size, shape, function, knowledge that an umbrella is used to keep yourself dry
Procedural model	Sequential and timing type of knowledge Knowledge of how/knowing how	Task action/goal relationship, and external influences on this relationship	Use of an umbrella
Strategic model	Information that is the basis of problem solving Knowledge of what and how and applied to the context	Action plans to meet specific goals, knowledge of the context in which procedures should be implemented, actions to be taken if a proposed solution fails, and how to respond if required information is absent	Applied use of an umbrella not only for rain, but also sun, sandstorms etc.

1.4.2 Situation awareness

The concept of situation awareness (SA) emerged from aviation psychology, initially intended to describe the component of the pilot's comprehension at tactical flight operations (Durso & Gronlund, 1999). It is also positioned to be a predictor of good decision making and performance in complex environments. In terms of theoretical implications, SA is often viewed as a buzzword for a variety of cognitive processes (Prince, Salas, & Brannick, 1999; Sarter & Woods, 1995). Croft, Banbury, Butler and Berry (2004) argued situation awareness to be an epiphenomenon of cognition. As a construct of cognition, it can be placed in line with other cognition constructs, such as Mental Models, and in terms of team cognition together with e.g. group learning, strategic consensus and transactive memory, in which the constructs are conceptually different due to their scopes (Mohammed, Ferzandi & Hamilton, 2010).

Despite the discussion on definition, the cognitive construct of situation awareness is identified as a unique one. One of the best-known definitions of situation awareness from a psychological approach was proposed by Endsley (1988): "The perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future". Three separate levels of SA can be derived from this definition: level 1 - perception (of the elements in the environment, e.g. identification of an aircraft, mountains, warning light etc. by pilots), level 2 - comprehension (of the current environment or situation, e.g. determining the status of a power plant system through disparate bits of data by a power plant operator), level 3 - projection (of the future status, e.g. predicting which airplane runways will be free in order to prevent collisions by traffic controllers) (Endsley, 1995). Mastering one level increases the likelihood of apprehending the next level more successfully. Endsley differentiates between situation awareness as product and SA as process respectively, by using the term 'situation awareness' for "a state of knowledge from the processes used to achieve that state", whereas the term 'situation assessment' refers to "the process of achieving, acquiring or maintaining SA" (Endsley, 1995, p.36). Other frameworks also exist, such as the perceptual cycle by Neisser (1976).

It should be noted that although situation awareness is a common term that can be applied in practice, this cognitive construct is specifically linked to dynamic situations in which operators are faced with timely decisions.

1.4.3 Situation awareness on different levels

The three-level model by Endsley (1988) has been widely accepted as a definition for situation awareness on the individual level. However, this is not the case for the theoretical acceptance of team situation awareness. Situation awareness on a team or network/system level has been approached from multiple theoretical perspectives: the classic psychological information-processing perspective (Endsley, 1988), the team cognition perspective (Cooke, Gorman, Myers, & Duran, 2013), and distributed situation awareness which builds on the distributed cognition perspective (Hutchins, 1995).

According to Endsley's three-level model, team situation awareness can be operationalized by aggregating individual SA to the team level. Team situation awareness can be defined as "the degree to which every team member possesses the SA required for his or her responsibilities", (Endsley, 1995, p.39). Additionally, as a component of team SA, the shared SA within a team should be high as well. Shared SA is defined as "the degree to which team members have the same SA on shared SA requirements", (Endsley & Jones, 1997, p. 47). Factors that contribute to the shared SA development in teams are shared SA requirements (overlap of necessary information to take decisions), shared SA devices (communication, shared displays and shared environment), shared mechanisms (shared mental models, i.e. shared knowledge structures of team members) and shared processes (team process behaviors) (Endsley, Bolté, & Jones, 2003). The level of similarity and accuracy of knowledge are often properties analyzed by the information-processing and related shared cognition paradigm. This approach posits that knowledge is held on the individual level.

The Interactive Team Cognition (ITC) theory holds its roots from perspectives such as ecological psychology, situated cognition and activity theory (Cooke et al., 2013). This theory states that it might be too simplified to only look at the aggregation of individual SA and team process behaviors, although doing so might be better suited for smaller and homogeneous teams. ITC theory puts its emphasis on the interactions as cognitive processes at the team level instead of knowledge as a cognitive structure. Therefore, the cognition of a team is holistically defined by the interactions of team members, rather than by the static knowledge structures. Team situation awareness is constructed from: (1) an alteration of the environment that is perceived by two or more team members (through their system), (2) a coordinated perception and interpretation of the change and (3) a coordinated action by one or more team members to overcome future negative impacts of the change. Following this theory, knowledge is held on the team level.

Thirdly, the Distributed Situation Awareness (DSA) theory is inspired from the distributed cognition theory and cognitive systems engineering discipline, which posits that SA resides within human and non-human agents (Salmon, Stanton, Walker, & Jenkins, 2009; Stanton, Stewart, Harris, Houghton, Baber, McMaster, et al., 2006). This theory takes upon a system's perspective in the analysis of human-computer interaction and states that knowledge is distributed across the system. An illustration of this is the representation of certain information displays, which carry the translation of mental efforts to support the cognition of operators. Additionally, the DSA theory assumes knowledge as being compatible and not shared, as stated by the three-level model of SA. According to DSA, the knowledge of each team member is unique, but compatible or fitting to the knowledge of another team member. Although it might seem that knowledge is held at the individual level, information from objects in the system is accounted for as well. Thus, team SA can be interpreted as a unit of the system contrary to a unit of the individual or as a unit of the team. Although this perspective has a high relevance, this theory will not be investigated in this work.

Table 1.4 depicts similarities and differences between the three theories. The criteria distinguish the theories in terms of:

- Unit of analysis: on which level can the theory be used to investigate situation awareness, i.e. on the individual, team and/or network/system level
- Paradigm and perspective in this paradigm: the classical information-processing paradigm in psychology, next to the paradigm of macrocognition
- Theoretical model: perspectives on causal interaction, which can follow (1) the classic simple cause and effect perspective, in line with the Input-Process-Output (IPO) model or (2) the dynamic causal interaction perspective, in which the individual operators as 'team members' and the 'team' co-exist in a circular relation (Gorman, Cooke, & Winner, 2006; Illgen, Hollenbeck, Johnson, & Jundt, 2005)
- Situation awareness as structure or process: SA as structure, in terms of knowledge, or SA as a process in terms of a continuous perception-action process (Endsley, 1988; Gorman, Cooke, & Winner, 2006). Additionally, on team level knowledge of team members can be shared in multiple ways: similarity (identical information held in common by team members), overlap (knowledge that overlaps among team

members), compatibility (unique and complementary knowledge of team members) and transactive (identical knowledge which may be used differently by different team members) (Mohammed, & Dumville, 2001; Salmon et al., 2009)

- Required simulated physical elements: requirements of the simulated environment in terms of its physical features
- Required simulated process elements: requirements of the simulated environment in terms of its processes

Table 1.4: Three theoretical perspectives on team situation awareness.

	Three-level model (Endsley, 1988)	**Interactive Team Cognition (Cooke, Gorman, Myers, & Duran, 2013)**	**Distributed Situation Awareness (Stanton, Stewart, Harris, Houghton, Baber, McMaster, et al., 2006)**
Unit of analysis	Individual, team, & network/system	Team & network/system	Network/system
Paradigm	Information-processing	Macrocognition, group cognition	Macrocognition, distributed cognition
Theoretical model	Simple cause and effect: Input-Process-Output (IPO)	Dynamic causal interaction: emergence through interactions	Simple cause and effect: Input-Process-Output (IPO)
Situation awareness as a structure or process	Cognitive structure; focus team knowledge in terms of similarity and overlap	Cognitive processes: emphasis on cognitive team processes compared to knowledge	Cognitive structure; focus team knowledge (including artifacts) in terms of compatible and transactive knowledge
Required simulated physical elements	Representation of the information system is irrelevant as long as all necessary (shared) information is provided, realistic (co-)location of participants	Representation of the information system is irrelevant as long as all necessary information is provided, realistic (co-) location of participants	Highly realistic and similar information systems and work environment
Required simulated process elements	Highly similar shared mental models and shared processes	Highly similar conditions to support task and team processes	Highly similar process elements

Chapters 4 and 8 focus on (shared) mental models studies, while Chapters 5 to 7 focus on studies revolving around situation awareness. Each chapter differs in the extent to which a unit of analysis is addressed.

1.5 System theory

In order to understand the role of human operators as part of a system such as the railways, a high, system level approach can be applied (Sheridan, 2010; Zarboutis & Marmaras, 2002). The railway domain is exemplary of a socio-technical system as it consists of multiple subsystems, each with an interdependence of physical, technical and human components and each constantly changing, reorganizing and evolving. These characteristics are applicable to that of a complex adaptive system (Holland, 1992; McCarthy, 2003). According to Holland (1992), all complex adaptive systems share three similar characteristics: (1) evolution, (2) aggregate behavior and (3) anticipation. Firstly, evolution occurs through the ability of components in the system to adapt and learn due to interactions with the surroundings. Secondly, aggregated behavior in a system cannot be simply derived from the actions of the components, rather it emerges from the interaction of the components. Thirdly, adapting to changing circumstances through anticipation of the consquences of certain responses is another feature that adds to the complexity of the system.

In studying systems with the complex adaptive systems perspective, three levels can be distinguished: the system level, network level and agent level (Bekebrede & Meijer, 2009). The agent level focuses on agents (i.e. professionals or operators) and their behavior and adaptations on individual and team levels, in which key properties are for instance agent diversity and adaptiveness. The network level focuses on the network dynamics in terms of the interaction between agents amongst themselves and with the formalized systems and on network evolution in terms of the physical, technical and human components. The system level focuses on properties, such as self organization of (teams of) agents, path dependency of processes and robustness and instability in terms of processes and strategies.

Parallels can also be drawn with the levels of decision-making, in which the agent, network and system level can be matched to respectively operational, tactical and strategic forms of decision-making (Van den Hoogen, 2019). Chapters 4 to 11 focus on the operational decision-making level, while Chapter 10 and 11 also include the tactical and strategic decision-making level.

1.6 Agent-based modeling and simulation

Following the field of complex adaptive systems, agents are characterized by their dynamic, interactive, autonomous and non-lineair behavior (Berry, Kiel, & Elliott, 2002; Macal & North, 2009). On the contrary to traditional methodologies that investigate social dynamics, agent-based modeling and simulation assume that social structure and social facts are created bottom-up through the interactions of agents. Through these microlevel interactions, emergent behavior can be observed. As such, agent-based modeling can also be seen as another, new way of doing science (Axelrod, 1997; Smith & Conrey, 2007).

The field of agent-based modeling is a multi-disciplinary one, as it spans across researchers from computational and social sciences to life sciences and ecology (Bandini, Manzoni, & Vizzari, 2009; Niazi, 2013). Social sciences and particularly that of applied psychology/human factors, primarily focuses on human behavior modeling (Gluck & Pew, 2001). To develop higher fidelity human behavior models and predict individual performance, a cognitive model of human operators is needed (Reyling, Lovett, Lebiere, Reder, & Demiral, 2004; Wellbrink, 2003). Next to their use for simulation, software agents can be applied to support human operators by (1) taking decisions autonomously, (2) reactive or (3) pro-active support operators by providing alerts or information to support operator decisions, and (4) collaborate with operators or other agents (Wooldridge & Jennings, 2009).

The current work investigates the development of a railway traffic operator's cognitive model based on their situation awareness. Using software agents, a human operator can be simulated in a gaming simulation when operators are unable to participate in a gaming simulation session. This topic is investigated in Chapter 9.

1.7 Research focus

The previous sections touched upon the concepts of mental models and situation awareness, which are investigated using gaming simulations as a participatory design tool within the railway sector. Finding a basic understanding of the cognition of operators is deemend relevant to identify the resilience of the system, in an organizational context where robustness of the railway system takes a central role.

In introducing the aforementioned concepts, this dissertation focuses on two main thematic areas: gaming simulations and human factors. The aim is to bridge the gap on theoretical and methodological constraints that can be found when using different gaming simulation designs for research purposes. Research questions that are addressed are:

1. How are cognitive concepts such as mental models and situation awareness of train traffic and network operators relevant for gaming simulations and vice versa?
2. Which situation awareness theories can be used in the Dutch railway traffic control domain?
3. What are the requirements on research gaming simulations in order to measure the mental models and situation awareness of operators?
4. To what extent can gaming simulation be used as a formal research environment for complex operational environments?
5. How can mental models and situation awareness contribute to system design processes using gaming simulations?

In answering these questions this dissertation is grouped and ordered as follows.

1.7.1 Section I

Validity is of essence in the design and execution of a gaming simulation. The topic of validity in relation to gaming simulation design and types are addressed in the following chapters, in which is presented:

- A framework of different validity types in gaming simulation (Chapter 2)
- A framework on cognitive structures in relation to different gaming simulation types (Chapter 3)

1.7.2 Section II

Each of the chapters in Chapters 4 to 8 address a combination of different dimensions (actor/level, gaming simulation representation, cognition) (see Figure 1.4):

- Actor/level: single or multi actor, i.e. is the unit of analysis in the study an individual operator (single) or a team or network (multi-actor)
- Gaming simulation: low or high-tech, i.e. is an analog, tabletop environment (low-tech) used or a digital, human-in-the-loop simulation environment (high-tech)
- Cognition: mental models or situation awareness, i.e. is the cognitive construct of mental models or situation awareness investigated

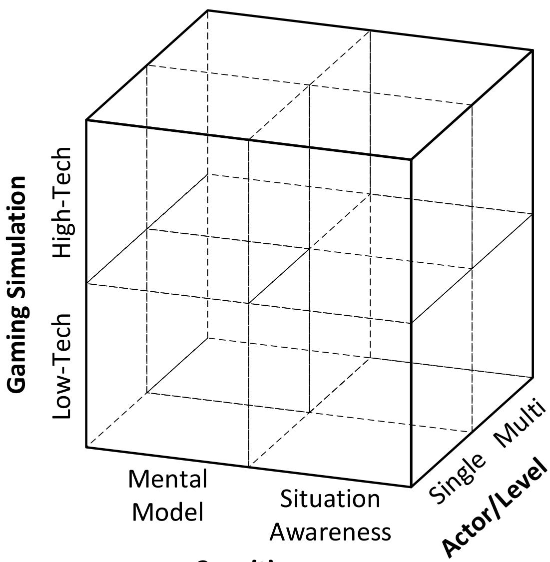


Figure 1.4: The three-dimensional research space that is spanned by Chapters 4 to 8.

In mapping these configurations to each chapter, the chapters can be distinguished as follows (see Table 1.5). For instance, the research study in Chapter 4 focuses on the investigation of the mental models of train traffic controllers in a high-tech gaming simulation (i.e. human-in-the-loop simulator), while the study in Chapter 8 focuses on the investigations of shared mental models (multiple operators) in a low-tech gaming simulation.

Table 1.5: Human factors research studies in this dissertation (Chapter 4 to 8).

Chapter	**Cognition**	**Actor/level**	**Gaming simulation**
Not investigated	Mental model	Single	Low-tech
Not investigated	Situation awareness	Single	Low-tech
4	Mental model	Single	High-tech
5	Situation awareness	Single	High-tech
8	Mental model	Multi	Low-tech
7	Situation awareness	Multi	Low-tech
Not investigated	Mental model	Multi	High-tech
6	Situation awareness	Multi	High-tech

1.7.3 Section III

Chapters 9 to 11 focus on various applications and syntheses of human factors studies

- For intelligent agents (Chapter 9)
- For participatory system design and strategic decision-makers (Chapter 10)
- For hybrid gaming simulations and the implications for validity (Chapter 11)

This dissertation concludes with the discussion and conclusion in Chapter 12. A summary of the research outline of this dissertation is depicted in Figure 1.5.

It should be noted that Chapters 4 to 8 have been submitted and/or published in scientific journals with the work (literature review, data collection, analysis, manuscript) being done independently by the author and reviewed by co-authors. Chapter 9 was published as a conference paper which resulted from a collaboration. Reyhan Aydoğan and Julia Lo together developed the conceptual model, Reyhan Aydoğan focused on the formalization of the model and Julia Lo focused on the human factors literature and the test study. Sebastiaan Meijer and Catholijn Jonker reviewed the work.

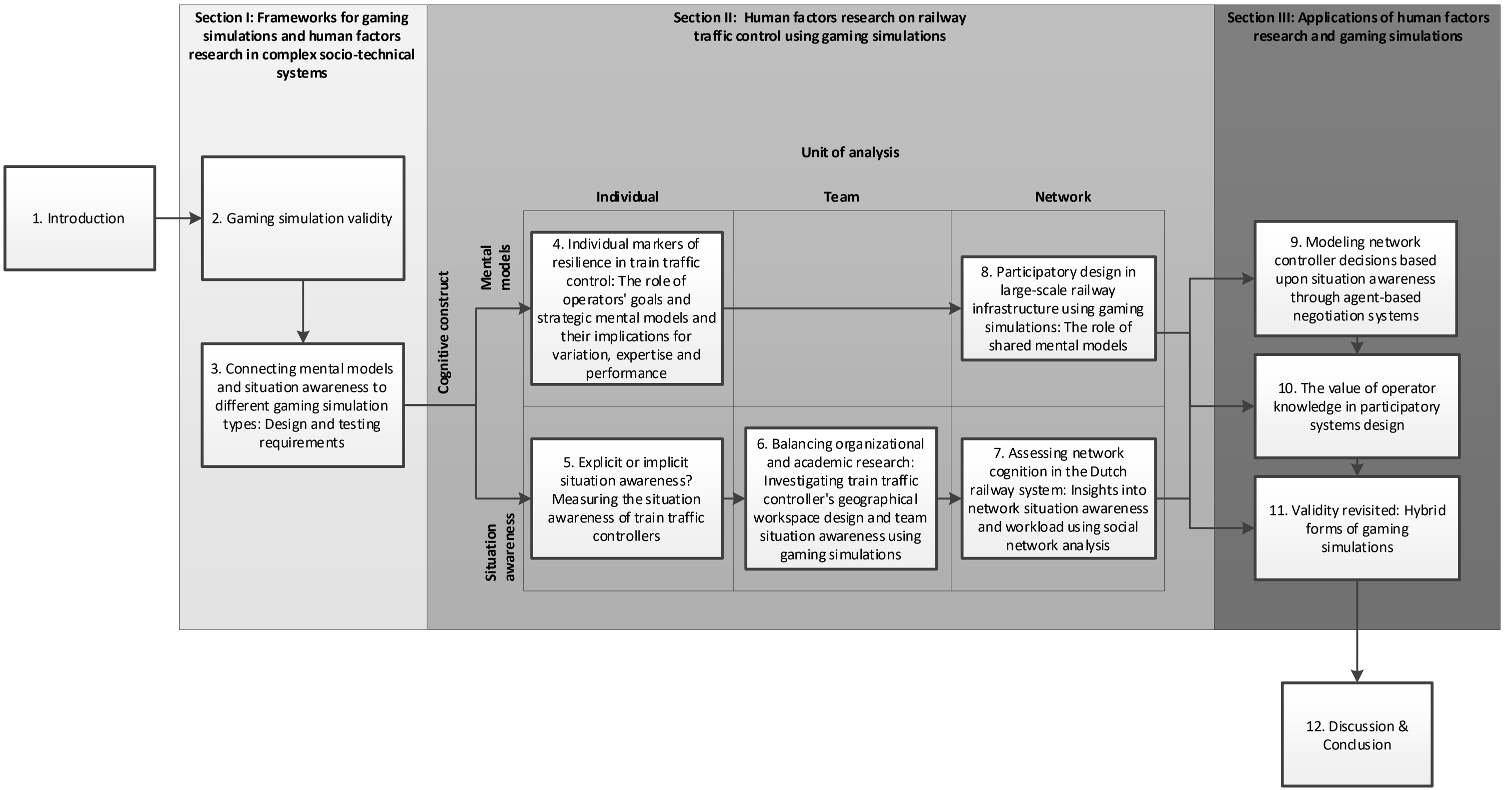


Figure 1.5: Research outline of this dissertation.

References

Axelrod R. (1997) Advancing the Art of Simulation in the Social Sciences. In R. Conte, R. Hegselmann, & P. Terna (Eds.) Simulating Social Phenomena. *Lecture Notes in Economics and Mathematical Systems*, 456. Springer, Berlin, Heidelberg

Bandini, S., Manzoni, S. & Vizzari, G. (2009). Agent Based Modeling and Simulation: An Informatics Perspective*. Journal of Artificial Societies and Social Simulation 12*(4): 4

Bekebrede, G., & Meijer, S. A. (2009). Understanding complex infrastructure systems: the case of SimPort-MV2. *Second International Conference on Infrastructure Systems and Services: Developing 21st Century Infrastructure Networks (INFRA).*

Berry, B. J. L., Kiel, L. D., & Elliott, E. (2002). Adaptive agents, intelligence, and emergent human organization: Capturing complexity through agent-based modeling. *Proceedings of the National Academy of Sciences, 99*(3), 7187-7188.

Cannon-Bowers, J. A., Salas, E., & Converse, S. (1993). Shared mental models in expert team decision making. In J. John Castellan, N. (Ed.), *Individual and Group Decision Making: Current Issues*. Hillsdale, New Jersey: Lawrence Erlbaum

Cooke, N. J., Gorman, J. C., Myers, C. W., & Duran, J. L. (2013). Interactive Team Cognition. *Cognitive Science, 37*(2), 255-285.

Croft, D. G., Banbury, S. P., Butler, L. T., Berry, D. C. (2004). The role of awareness in situation awareness. In S. Banbury & S. Tremblay (Eds.), *A Cognitive Approach to Situation Awareness: Theory and Application* (pp. 82-103). Bodmin, Cornwall: MPG Books Ltd.

Durso, F. T., & Gronlund, S. D. (1999). Situation awareness. In F. T. Durso, R. Nickerson, R. Schvaneveldt, S. Dumais, M. Chi, & S. Lindsay (Eds.), *Handbook of Applied Cognition*. New York: Wiley.

Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement. *Proceedings of the 32nd Human Factors Society Annual Meeting*, 32, 97-101.

Endsley, M. R. (1995). Towards a Theory of Situation Awareness in Dynamic Systems. *Human Factors, 37*(1), 32-64.

Endsley, M. R. (2000). Direct measurement of situation awareness: Validity and use of SAGAT. In M. R. Endsley & D. J. Garland (Eds.), *Situation Awareness Analysis and Measurement* (pp. 147-174). Mahwah, NJ: Lawrence Erlbaum Associates.

Endsley, M. R., Bolté, B. & Jones, D. G. (2003). *Designing for Situation Awareness*. New York, NY: Taylor & Francis Group.

Endsley, M. R. & Jones, W. M. (1997). *Situation Awareness, Information Dominance, and Information Warfare.* United States Air Force Armstrong Laboratory Technical Report 97-01.

Flach, J. M. (1995). Situation Awareness: Proceed with Caution. *Human Factors, 37*(1), 149-157.

Gluck, K. A., & Pew, R. W. (2001b). Overview of the agent-based modeling and behavior representation (AMBR) model comparison project. *Proceedings of the 10th Computer Generated Forces and Behavioral Representation Conference*. Orlando, FL: Division of Continuing Education, University of Central Florida.

Golightly, D., Sandblad, B., Dadashi, N., Andersson, A. W., Tschirner, S., & Sharples, S. (2013). A socio-technical comparison of rail traffic control between GB and Sweden. In N. Dadashi, A. Scott, J. R. Wilson, & A. Mills (Eds.), *Rail Human Factors: Supporting Reliability, Safety and Cost Reduction* (pp. 367–376). London: Taylor & Francis.

Gorman, J. C., Cooke, N. J., & Winner, J. L. (2006). Measuring team situation awareness in decentralized command and control environments. *Ergonomics, 49*(12-13), 1312-1325.

Holland, J. H. (1992). Complex Adaptive Systems. *Daedalus, 121*(1), 17-30.

Hutchins, E. (1995). *Cognition in the Wild*. Cambridge, MA: MIT Press.

Ilgen, D. R., Hollenbeck, J. R., Johnson, M., & Jundt, D. (2004). Teams in Organizations: From Input-Process-Output Models to IMOI Models. *Annual Review of Psychology, 56*(1), 517-543.

International Ergonomics Association (IEA) (2003). *IEA triennial report: 2000–2003*. Santa Monica, CA: IEA Press.

Jacko, J. A., Yi, J. S., Sainfort, F., & McClellan, M. (2012). Human factors and ergonomic methods. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (pp. 289-329). New York: Wiley.

Karwowski, W. (2012). The discipline of human factors and ergonomics. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (pp. 3-37). New York: Wiley.

Kensing, F., & Blomberg, J. (1998). Participatory Design: Issues and Concerns. *Computer Supported Cooperative Work (CSCW), 7*(3), 167-185.

Klabbers, J. H. G. (2006). A framework for artifact assessment and theory testing. *Simulation & Gaming, 37*(2), 155-173.

Klabbers, J. H. G. (2009). *The magic circle: Principles of gaming & simulation*. Brill Sense.

Klimoski, R., & Mohammed, S. (1994). Team mental model: construct or metaphor? *Journal of Management, 20*(2), 403-437.

Knieps, G. (2013). Competition and the railroads: A European perspective. *Journal of Competition Law & Economics, 9*(1), 153-169.

Lai, Y.-C., & Barkan, C. P. L. (2009). Enhanced Parametric Railway Capacity Evaluation Tool. *Transportation Research Record, 2117*(1), 33-40.

Lim, B. C., & Klein, K. J. (2006). Team mental models and team performance: A field study of the effects of team mental model similarity and accuracy. *Journal of Organizational Behavior, 27*(4), 403-418.

Macal, C. M., & North, M. J. (2009, 13-16 Dec. 2009). Agent-based modeling and simulation. *Proceedings of the 2009 Winter Simulation Conference (WSC).*

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology, 85*(2), 273-283.

McCarthy, I. P. (2003). Technology management–a complex adaptive systems approach. *International Journal of Technology Management, 25*(8), 728-745

Meijer, S. A. (2008). *The Organisation of Transactions: Studying Supply Networks using Gaming Simulation*. Wageningen: Academic Publishers.

Meijer, S. (2015). The power of sponges: Comparing high-tech and low-tech gaming for innovation. *Simulation & Gaming, 46*(5), 512-535.

Mohammed, S., & Dumville, B. C. (2001). Team mental models in a team knowledge framework: expanding theory and measurement across disciplinary boundaries. *Journal of Organizational Behavior, 22*(2), 89-106.

Mohammed, S., Ferzandi, L., & Hamilton, K. (2010). Metaphor No More: A 15-Year Review of the Team Mental Model Construct. *Journal of Management, 36*(4), 876-910.

Muller, M. J. (2003). Participatory Design: The Third Space in HCI. In A. Sears, & J.A. Jacko (Ed.) *Human-Computer Interaction: Development Process*, 165-186.

Muller, M. J., & Kuhn, S. (1993). Participatory Design. *Commun. ACM, 36*, 24-28.

Narayanasamy, V., Wong, K. W., Fung, C. C., & Rai, S. (2006). Distinguishing games and simulation games from simulators. *Computers in Entertainment, 4*(2), 9.

Neisser, U. (1976). *Cognition and Reality: Principles and Implications of Cognitive Psychology.* New York, NY, US: W H Freeman & Co.

Niazi, M. A. (2013). Complex Adaptive Systems Modeling: A multidisciplinary Roadmap. *Complex Adaptive Systems Modeling, 1*(1), 1.

Prince, C., Salas E., and Emery L. (1999). Situation awareness: What do we know now that the 'Buz' has gone?" In D. Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics, 3* (pp. 215-222).

Rehling, J., Lovett, M., Lebiere, C., Reder, L., & Demiral, B. (2004). Modeling Complex Tasks: An Individual Difference Approach. *Proceedings of the Annual Meeting of the Cognitive Science Society, 26*.

Russ, T. L. (2010). Programmatic and Participatory: Two Frameworks for Classifying Experiential Change Implementation Methods. *Simulation & Gaming, 41*(5), 767-786.

Salas, E., Stout, R. J., & Cannon-Bowers, J. A. (1994). The role of shared mental models in developing shared situational awareness. In R. D. Gilson, D. J. Garland & J. M. Koonce (Eds.), *Situational Awareness in Complex Systems* (pp. 297-304). Daytona Beach, Florida: Embry-Riddle Aeronautical University Press.

Salmon, P. M., Stanton, N. A., Walker, G. H., Jenkins, D., Ladva, D., Rafferty, L., & Young, M. (2009). Measuring Situation Awareness in complex systems: Comparison of measures study. *International Journal of Industrial Ergonomics, 39*(3), 490-500.

Sarter, N. B., & Woods, D. D. (1995). How in the World Did We Ever Get into That Mode? Mode Error and Awareness in Supervisory Control. *Human Factors, 37*(1), 5-19.

Sheridan, T. B. (2010). The System Perspective on Human Factors in Aviation. In E. Salas & D. Maurino (Eds.), *Human Factors in Aviation* (Second Edition) (pp. 23-63). San Diego: Academic Press.

Smith, E. R., & Conrey, F. R. (2007). Agent-Based Modeling: A New Approach for Theory Building in Social Psychology. *Personality and Social Psychology Review, 11*(1), 87-104.

Spinuzzi, C. (2005). The Methodology of Participatory Design. *Technical Communication 52*(2), 163-174.

Stanton, N. A., Stewart, R., Harris, D., Houghton, R. J., Baber, C., McMaster, R., Salmon, P. Hoyle, G, Walker, G. Young, M. S., Linsell, M, Dymott, R., & Green, D. (2006). Distributed situation awareness in dynamic systems: theoretical development and application of an ergonomics methodology. *Ergonomics, 49*(12-13), 1288-1311.

Tenney, Y. J., & Pew, R. W. (2006). Situation Awareness Catches On: What? So What? Now What? *Reviews of Human Factors and Ergonomics, 2*(1), 1-34.

Van den Hoogen, J. (2019). *The Gaming of Systemic Innovations: Innovating in the Railway Sector using Gaming Simulations*. Ede: Print Service.

Wellbrink, J. (2003). Modeling Reduced Human Performance as a Complex Adaptive System. PhD thesis.

Wickens, C. D. (2008). Situation Awareness: Review of Mica Endsley's 1995 Articles on Situation Awareness Theory and Measurement. *Human Factors, 50*(3), 397-403.

Wooldridge, M., & Jennings, N. R. (2009). Intelligent agents: theory and practice. *The Knowledge Engineering Review, 10*(2), 115-152.

Zarboutis, N., & Marmaras, N. (2002). Emergency Management Respecting System's Com-plexity: The Case of Fire in a Metro Railway System. In S. Bagniara, S. Pozzi, A. Rizzo & P. Wright (Eds.), *Cognition, Culture & Design* (pp. 275-281). Sienna: Instututo di Sienze et Tecnologie dela Cognizione.

Section I

Frameworks for gaming simulation and human factors research in complex socio-technical systems

The chapters in section 1 focus on the theoretical foundation of this dissertation by connecting different research fields: gaming simulation, human factors and methodology in social and computational science.

Chapter 2 reviews different types of validity in computer sciences, social sciences and human factors, and provides a framework that captures the different stages and types of validity that are wished for in gaming simulations.

Chapter 3 connects the relevance and application of the cognitive concepts of situation awareness and mental models from the human factors field to the different types of gaming simulation, such as testing and training.

2 Gaming simulation validity

This chapter is based on the following work:
Lo, J. C., Van den Hoogen, J. & Meijer, S. A. (2013). Using Gaming Simulation Experiments to Test Railways Innovations: Implications for Validity. In R. Pasupathy, S. H. Kim, A. Tolk, R. Hill & M. E. Kuhl (Eds.), *Proceedings of the 2013 Winter Simulation Conference (WSC)* (pp. 1766-1777).

Page, L., Lo, J., Velazquez, M. & Claudio, D. (in review). Introducing a validation framework to standardize and improve human factors and ergonomics research.

Abstract
The field of gaming simulations has grown substantially in the last decade with a strong emphasis on the business environment. In this context an increasing need for the assessment of gaming simulation validity is observed. This chapter focuses on a literature review of gaming simulations along with developments in validity testing. Given the interdisciplinary nature of gaming simulations, this review includes validity types in closely related fields, such as computer science and human factors. As a result of unclear and overlapping definitions and constructs, the path towards validity testing of a gaming simulation is non-trivial. This review is concluded by a synthesis on gaming simulation validity types and methodological directions.

Keywords: Ecological validity; measurement/test validity; external validity; internal validity; computer simulation; human-in-the-loop simulation

2.1 Introduction

Under a plethora of synonyms for gaming simulations, the gaming field with in particular business environments, has grown substantially in the last decade where computer and internet developments played a major role (Faria, Hutchinson, Wellington, & Gold, 2008). The interest of organizations is reflected in applications of games as part of their learning curriculum, to facilitate the policy development process on a strategic level and to understand operational processes (e.g. Harteveld, 2012; Mayer et al., 2013). Given the business context where impact, efficiency and effectiveness are important driving factors, the validity of a game used in business is a returning question. As there is an increasing need to research the validity of games, this chapter aims to provide a literature review on validity developments within the gaming field. As such, the aim of this review is to identify the current state of the art and gaps in literature around the validity of gaming, especially for its non-educational use. Given the interdisciplinary nature of gaming simulations, validity developments in fields that are closely related to the fields of computer simulations and human factors are also considered. Herein different validity types relevant for each field are identified, which are intersecting in gaming simulation and provide us with an environment for validity research analysis.

The following section firstly presents a review on the development of gaming simulation validation, followed by an overview of validation studies in closely related fields such as computer simulations and human factors.

2.2 Validity development in gaming simulation

Insights in the popularity of validity in games can be obtained through a search in different search engines. On the terms 'game OR gaming' and 'validity OR validation' in the title and abstract, 2602 hits can be found on Scopus between 1975 and mid November 2015 (see Figure 2.1). Here an increase in the number of published articles over time can be observed.

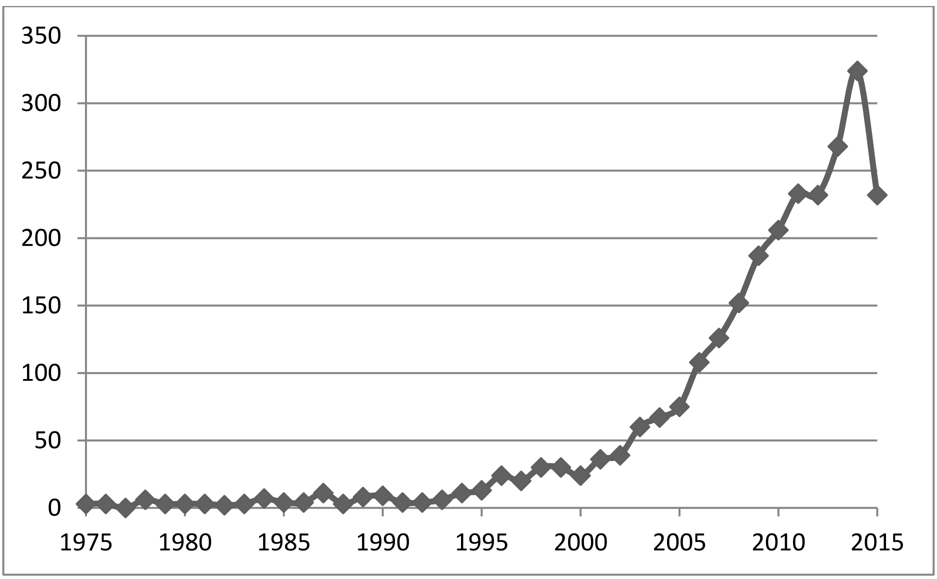


Figure 2.1: 2602 hits on a search on the Scopus engine search on the terms 'game OR gaming' and 'validity OR validation' in the title and abstract.

An additional search was performed on Google Scholar on 'game' and 'validity OR validation' in the title. A slightly deviating search entry was entered compared to the previous search due to limited search configurations. Based on this entry, 224 hits with 18 missing publication years were found (see Figure 2.2). In both search engine results, a positive trend can be observed for the increase over time in the number of articles on validity in games.

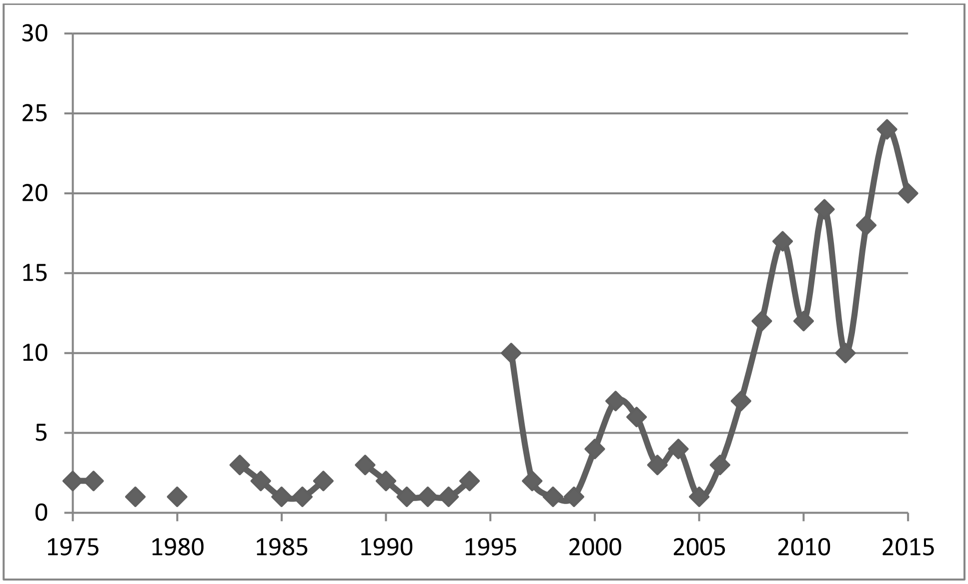


Figure 2.2: Google Scholar engine search on the terms 'game' and 'validity OR validation' in the title based on 224 hits.

2.3 A brief literature overview

Up until the 1970s gaming simulation was a budding new field of science where especially computerized business games were introduced (e.g. Newgren, Stair, & Kuehn, 1981). A novel methodological approach in developing validity testing methodology was introduced by Raser (1969). First views were presented on the validation of gaming simulations and three relevant validity components were defined: structural validity, process validity, and predictive validity. Psychological reality refers to the similarity of the psychic environment, on which Raser posits that participants might perceive the simulated environment "not as dissimilar as one might think" (Raser, 1969, p.147). Structural validity relates to "the degree that its structure (the theory and assumptions on which it is built) can be shown to be isomorphic to that of the reference system" (Raser, 1969, p.144). Process validity relates to "the degree that the processes observed in the game are isomorphic to that of the reference system" (Raser, 1969, p.144). Finally, predictive validity refers to "the degree that it can reproduce historical outcomes or predict the future" (Raser, 1969, p.144).

In the 80s, validation studies strongly focused on the internal and external validity in relation to business gaming simulations, which were mostly used for training purposes (e.g. Dickinson, Whiteley, & Faria, 1990; Norris & Snyder, 1981; Whiteley & Faria, 1989; Wolfe & Roberts, 1986). However, these definitions varied amongst researchers. For some, external validity was determined using evaluative questions about the worth of a simulation game

(Norris & Snyder, 1981), while other scholars measured to which extent the performance in the gaming simulation compared to the real-world business performance or measured the impact of the gaming simulation effectiveness through a longitudinal study (Dickinson, Whiteley, & Faria, 1990). It should be noted that more common definitions of external and internal validity in psychological research are used in terms of respectively generalization of research findings and causal relations (e.g. the American Educational Research Association, American Psychological Association, National Council on Measurement in Education and Joint Committee on Standards for Educational and Psychological Testing, 1999).

Another marking development was the distinction of different gaming simulation types into three categories: policy, training and research games (Peters, Vissers, & Heijne, 1998). Here, policy games focus on the exploration of options that provide support in solving an organizational problem or improving its solution, whereas training games are used as a tool for teaching or education. Thirdly, research games are used to investigate certain questions that a researcher has about the reference system. Games are also seen as an experimental setting (e.g. Vissers, Heyne, Peters, & Geurts, 2001). Each of these gaming simulation types would have different implications for validity. Both Peters, Vissers and Heijne (1998) and Feinstein and Cannon (2001) show that validity in gaming simulation is still developing at the start of the 21st century. They appoint this to the lack of a widely accepted methodology of validation and the lack of a developed construct of validity. To further develop this area in gaming simulation, Feinstein and Cannon provide an extensive taxonomy of the terms of validity used in the field of gaming simulation, in which gaming simulations should be based on verification, internal validation, and external validation.

More recently many games are once again focused on training (Graafland, Schraagen, & Schrijven, 2012). Stainton, Johnson and Borodzicz (2010) introduced a methodological framework for the assessment of educational validity, especially applicable for the validity of business gaming simulations. However, it can be questioned whether this framework rather reflects effectiveness in terms of transfer of learning.

Zomer, Moustaid and Meijer (2015) introduced a framework in which they identified elements in the characteristics of a gaming simulation that altogether lead to a multitude of varying gaming simulation configurations. These characteristics can be the output (data, simulation or human interference), input (data, simulation or human interference), type of combination (stand-alone, unidirection, bidirectional, human interference, participatory experiments or simulations) and type of communication (data, information and knowledge). These elements can be used to specify the type of gaming simulation in more detail and identify the extent of computer simulation and human participation in a specific configuration.

In sum, although there have been evaluations into the effectiveness of gaming simulations, there has also been much discussion on how evaluation should be approached. We posit that effectiveness is not validation, as often applied in training gaming simulations. Feinstein and Cannon (2001) remark that one of the problems plaguing gaming simulation research is confusion on the definition and the testing of validity. Another element of confusion is caused by the multidisciplinary nature of the gaming simulation field: gaming simulations increasingly involve computer simulation models as well as cognitive and behavioral aspects of human participants combined. As such, validation types and methods in both fields of computer science and human factors should be considered in the design and use of a gaming simulation. In the following sections, the role of validation in the area of human factors is explored with an emphasis on relevant validity types in this field. Subsequently, validation studies and their specific validity types in the area of computer simulations are reviewed.

2.4 Validation in the human factors and ergonomics field

Chapter 1 described the human factors and ergonomics as an interdisciplinary field, drawing on psychological, physiological and engineering disciplines. The foundation for research designs of studies is drawn from psychological research, in which applied settings are often simulated. The simulated environments that are used in this field moreover consist of simulators, such as aircraft cockpit human-in-the-loop simulators. As mentioned earlier, human-in-the-loop simulators are marked as a type of gaming simulation, i.e. one which uses gaming design principles but with a lower game play element. Human-in-the-loop simulation has a focus on a detailed simulation of a system, including the human inside it. Next to the connotation of human-in-the-loop simulator, flight, driving and sailing simulators are also often denoted together as vehicle simulators, where the environment is represented by a partial imitation of the cabin and the outside world representation is matched with a virtual and dynamic image (Padmos & Milders, 1992). The design aim for these vehicle simulators can be summarized: to create a realistic image and environment that is as close as possible to reality (Caro, 1973). Human-in-the-loop simulations slightly differ in definition from vehicle simulators in the wider range of the definition and the requirements that are set for the validity and degrees of freedom of the simulation design.

The research principles in the psychology domain distinguish a number of validity types: internal, measurement and external validity (Zechmeister, Zechmeister, & Shaughnessy, 2001). Figure 2.3 shows part of a research process, in which a specific research configuration eventually impacts the validity of the study (Page, Lo, Velazquez, & Claudio, in review). The research configuration refers to the experimental design of the study including sample size. The measurement instrument focuses on the testing instrument, such as questionnaires. Finally, the research environment refers to the simulated environment, respectively in a lab, gaming simulation or in a natural environment.

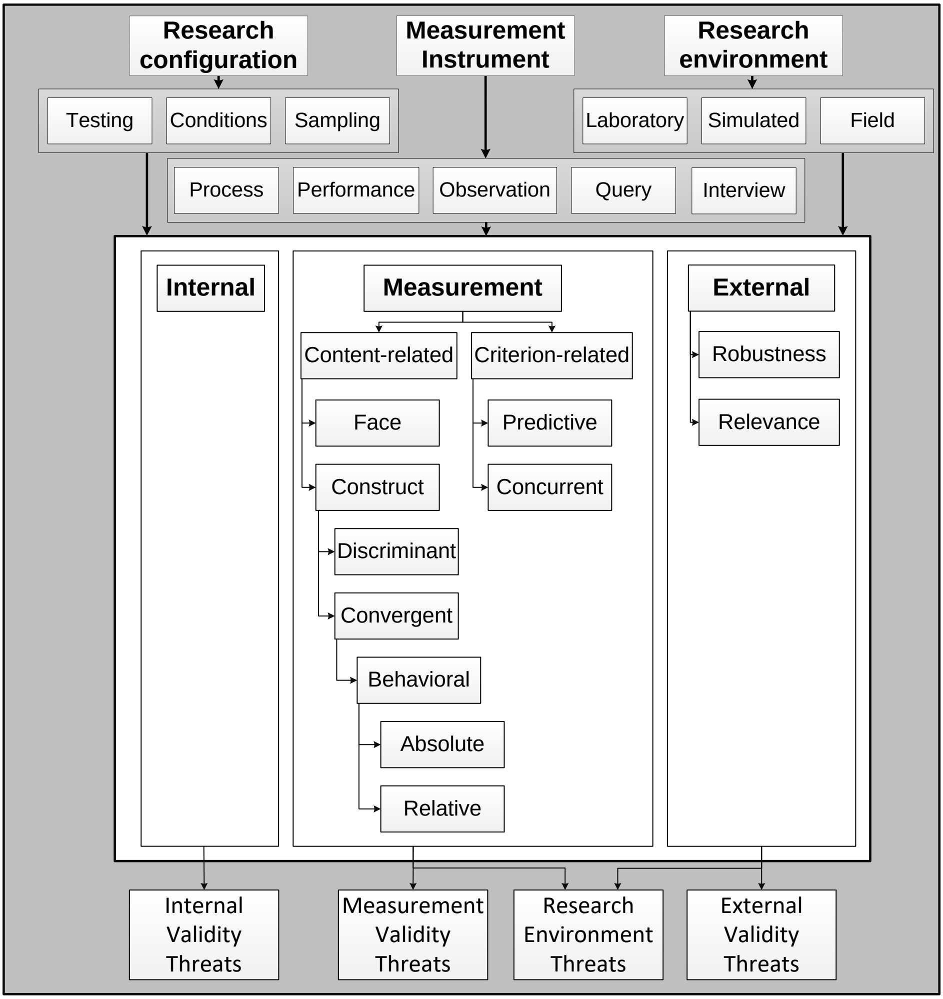


Figure 2.3: Validation testing, research design & execution (Page, Lo, Velazquez, & Claudio, in review).

2.4.1 Internal validity

Internal validity is defined as the ability to confidentially "state that the independent variable caused differences between groups on the dependent variable" (Zechmeister et al 2001, p.149). In order to make a causal inference, the experiment needs to establish a relationship between the independent and dependent variable, the cause must precede the effect, and finally, plausible alternative explanations should be outruled. To ensure the latter, confounding factors or internal validity threats need to be controlled (Campbell & Stanley, 1966, Zechmeister et al., 2001).

2.4.2 Measurement / Test validity

Measurement instruments are necessary to extract the information about causality from the research environment. Contrary to observable variables as in physics studies that are directly measurable, psychological variables as being part of cognitive processes in the human mind are characterized in that they can only be measured indirectly, for instance as with personality traits. This adds to the importance of measuring exactly what was intended to be measured. Measurement or test validity refers to the validity of measurement instruments, in which multiple typologies can be identified. Where the American Psychological Association started with three components that comprise measurement or test validity, the most recent version only identifies two (American Educational Research Association, American Psychological Association, National Council on Measurement in Education and Joint Committee on Standards for Educational and Psychological Testing, 1999; Trochim & Donnelly, 2008).

It should be noted that although traditional psychometric research focuses on measurement/test validity of questionnaires, human factors and ergonomics research also use these categories for the validation of a simulated environment.

2.4.3 Content-related validity

Content validity is considered a logical validity (McCoy, 1963) and is assessed by the degree of how well either an instrument or measure explains a sample of situations, content, or domain of interest (Constantine & Ponterotto, 2006; McCoy, 1963). With this in mind, content validity is composed of two subcategories of validity: face and construct.

Face validity is defined as an environment that looks like what it is supposed to measure; it is a subjective validity as it is difficult to quantify (George, 2003). In addition to an environment, this also applies to psychological tests. This validity is based on the judgement of users of the test, in terms of the obviousness of content as well as the situation in which the test is administered (Holden, 2010). The use of questionnaires to assess to what extent a simulator is representative to a reference system is an example of face validity.

The other subdivision of content-related validity is *construct validity*. A construct is "some postulated attribute of people, assumed to be reflected in test performance" (Cronbach & Meehl, 1955, p.284). Construct validity relates to all of the variables in the study (dependent and independent) and how they demonstrate the theoretical constructs being studied (Constantine & Ponterotto, 2006; McCoy, 1963; Wampold, 2006; Wiggins, 1968). Researchers attempt to quantify this logical validity by measuring the degrees of presence or absence of the corresponding measure. The independent variable indicates a construct and the dependent variables operationalize the construct (Blalock, 1968; Wampold, 2006). This type of logical validity (McCoy, 1963) is determined from expert opinion and is further divided into two other components.

Construct validity is composed of convergent and discriminant validity in instrument measurements (Constantine & Ponterotto, 2006). *Convergent validity* determines if the scales correlate positively and significantly with other measuring instruments that measure the same construct. Behavioral validity is a sub-category of convergent validity and is heavily used in transportation safety research and specifically focuses on the validity of a simulated environment (Reimer, D'Ambrosio, Coughlin, Kafrissen, & Biederman, 2006). *Behavioral validity* measures the difference, for example, between a person operating a driving simulator compared to a person operating a car (Blaauw, 1982). *Absolute validity* shows the same directions and magnitudes (Blaauw, 1982) while *relative validity* shows a similar, but not the exact same, direction and magnitude (Godley, Triggs, & Fildes, 2002). At the same time, while convergent validity measures similarity, *discriminant validity* demonstrates the lack of correlation between a theory and construct(s) that it should not correlate to. Assuming the scales being compared are good constructs, these two aspects of construct validity measure and assess how well a construct is measured and that the construct is clear of other theories that could confound the measure.

2.4.3.1 Criterion-related validity

Criterion-related validity demonstrates a theory that predicts an experimental outcome (Constantine & Ponterotto, 2006). it is comprised of predictive and concurrent validity. *Predictive validity* describes how well predictions can be made from tests or measures (McCoy, 1963). Those tests and measures are then confirmed by subsequent observation using statistical correlation analysis between test scores or criterion scores or measures. *Concurrent validity* is similar to predictive validity, but this validity involves a temporal component which shows how well scores or performance statistically correlates with concurrent status or performance. Concurrent validity usually occurs when one measurement is used instead of another (McCoy, 1963).

2.4.3.2 External validity

External validity is defined as "the extent to which findings from an experiment can be generalized to individuals, settings, and conditions beyond the scope of the specific experiment" (Zechmeister et al., 2001, p.161). This validity type shows a high similarity to Raser's predictive validity type.

2.5 Validation in computer simulations

In line with developments in the computer industry, more gaming simulations became digitalized and opportunities to use simulation models to reflect complex dynamics could be incorporated. For games that make use of simulation models, the field of computer simulation models and their approaches for validity were a factor to consider for validation research. Validation and verification are considered the two main types for the validation of computer simulation models and are captured by the IEEE community in standards, such as the Distributed Simulation Engineering and Execution Process (IEEE Computer Society, 2011). Validation refers to "substantiating that the model, within its domain of applicability, behaves with satisfactory accuracy consistent with the study objects" (Balci, 1994, p.121). Verification is about "substantiating that the model is transformed from one form into another, as intended, with sufficient accuracy" (Balci, 1994, p.123). In other words, validation relates to comparable outcomes of the model in comparison to the reference system, while verification focuses on the computer program working as intended (e.g. Kleijnen, 1995).

In comparing computer simulation with psychological experiments, especially internal validity and test validity issues become significant. As a closed system, and thus lacking the problems of confounding factors, computer simulation does not have internal validity issues. Even in non-deterministic simulations, Monte Carlo methods help in averaging out the influence of an independent variable on a dependent variable and showing if this influence is statistically significant. However, internal validity-like issues do appear during the computer programming of a conceptual model into a computerized model (Sargent, 2004). In computer simulation literature the mitigation of this validity threat is done using verification activities. In gaming simulation sessions, the introduction of game players makes the experiment inherently open, allowing all sorts of confounding variables to distort the causal picture of one independent variable and one dependent variable. Furthermore, as more soft variables are used to assess system behavior, e.g. workload and resilience, which do not need to be fully operationalized, gaming simulation, more than computer simulation, runs the risk of not measuring exactly that what was intended to be measured.

2.6 Towards a synthesis for gaming simulation validity types

As gaming simulation is an artefact where different disciplines overlap one another, investigating validity reaches a different level. As such, the topic of validation in gaming simulation can be very confusing due to the interwoven nature of multiple disciplines. In the case of a technical, digital game environment, simulation models can be used as a basis where a simulated environment is built upon. For instance, a system dynamics model may be applied to simulate a certain domain, i.e. the supply chain dynamics in the Beer Game using a system dynamics model. On the other hand, through the involvement of human participants in games, social scientific models and predominantly psychological approaches on validity should be considered as well.

Computer simulation as well as gaming simulation are both methods of simulating a reference system, each with their own properties and related strengths and weaknesses. Different purposes guide the development of both types of simulations. For computer simulations that are used for research, it is necessary to look into the process of simulation and conducting the research, which include the development of the model, the data analysis and the feedback of the results to others (Axelrod, 2003). However, this is also the case for gaming simulations. In essence, gaming simulations experiments (or direct experiments) follow more or less the same research process as computer simulations (or thought experiments) (Axelrod, 2003; Sterman 1987). In Figure 2.4, the research process of both types of simulations is presented where the use of computer simulation models may be an integral part of a gaming simulation and which both focus on two levels: (1) to model or create a simulated system that represents the reference system, and (2) to select valid simulation strategies or facilitate natural behavior in participants whilst controlling the research environments for confounding factors. This is in line with the process where a problem entity is translated into a computerized model through a conceptual model (Sargent, 2004).

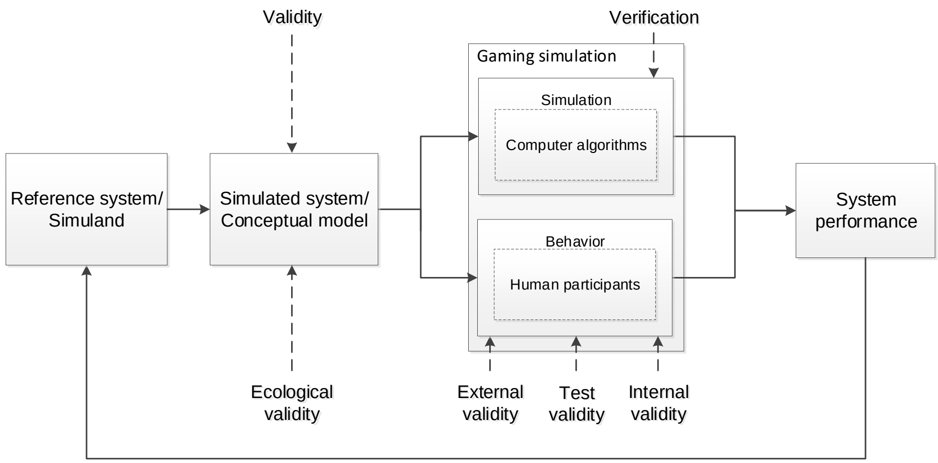


Figure 2.4: Types and stages of validating a gaming simulation (Lo, Van den Hoogen, & Meijer, 2013).

Reviewing how different validity types are used in the computer science and human factors field, validation can be identified in multiple stages within the process of designing and conducting a gaming simulation. The first stage involves the translation from the reference system to the simulation system, which often can be viewed as the conceptual model. When computer simulations are included in the gaming simulation design, the created model can be tested on its validity of the reference system. In case that the gaming simulation design only uses analogue methods, such as tabletop or board designs, the validity that can be assessed could be inferred using Raser's proposed validity types, i.e. structural validity, process validity and psychological reality. Additionally, the validity of the design of the gaming simulation as an environment can also be called ecological validity[2]. Ecological validity can be assessed through the degree to which the actors, information, physical elements etc. of the gaming simulation are comparable to the reference system (i.e. structural validity) and through the degree that participants perceive the gaming environment as realistic (psychological reality).

At the second stage, the gaming simulation session with its designed model is run. During test runs the computer model may be assessed on its verification. In terms of the overall gaming simulation session, the choice of the participants and measurement instrument facilitate respectively the external and test (or measurement) validity. Often the test or measurement validity of the instruments are assessed after the gaming simulation session. To make causal claims from the collected data (i.e. internal validity) the session needs to control for internal validity threats.

2.7 Discussion & conclusion

This chapter provided a structured and integrated interdisciplinary approach in the assessment of gaming simulation validation by reviewing different crucial validity types for gaming simulation. Notably, validation of the gaming simulation can be pinpointed to two stages: the design of the simulated environment and secondly the session itself. As such the design of the gaming simulation environment as a measurement tool can be separated from measurement tools to collect data within the gaming simulation session.

This chapter also focused on capturing the ecological validity of a gaming simulation to be able to draw implications on the validity of the design of a gaming simulation. So far validation often occurs by face validity, for instance subject matter experts evaluating the design in a qualitative manner. A quantitative manner in the form of questionnaires by subject matter experts has not been established yet as far as known. Also making use of behavioral validity tests could be a method to establish the ecological validity of a gaming simulation design.

References

American Educational Research Association, American Psychological Association, National Council on Measurement in Education and Joint Committee on Standards for Educational and Psychological Testing. (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.

Axelrod, R. (2003). Advancing the art of simulation in the social sciences. *Japanese Journal for Management Information Systems, 12*(3): 1-19.

---

[2]The authors recognize that ecological validity is inconsistently used throughout the social science literature, but it is chosen as a more neutral term in this context, in comparison to terms such as gaming simulation design validity or research environment validity.

Balci, O. (1994). Validation, verification, and testing techniques throughout the life cycle of a simulation study. *Annals of Operations Research, 53*(1), 121-173.
Blalock, H. M. (1968). The measurement problem: A gap between the languages of theory and research. In H. M. Blalock, & A. B. Blalock (Eds.), *Methodology in Social Research* (pp. 5-27). United States of America: McGraw-Hill, Inc.
Blaauw, G. J. (1982). Driving experience and task demands in simulator and instrumented car: A validation study. *Human Factors, 24*, 473-486.
Campbell, D. T., & Stanley, J. C. (1966). *Experimental and Quasi-Experimental Designs for Research*. Skokie, Illinois: Rand McNally & Company.
Caro, P.W. (1973). Aircraft simulation and pilot training. *Human Factors, 15*, 502-509.
Constantine, M. G., & Ponterotto, J. G. (2006). Evaluating and selecting psychological measures for research purposes. In F. T. Leong, & J. T. Austin (Eds.), *The Psychology Research Handbook: A Guide for Graduate Students and Research Assistants* (2nd ed., pp. 104-113). United State of America: Sage Publications, Inc.
Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281-302.
Dickenson, J. R., Whiteley, T. R., & Faria, A. J. (1990). An empirical investigation of the internal validity of a marketing simulation game. *Developments in Business Simulation & Experiential Exercises, 17*, 47-52.
Faria, A. J., Hutchinson, D., Wellington, W. J., & Gold, S. (2008). Developments in business gaming: A Review of the Past 40 Years. *Simulation & Gaming, 40*(4), 464-487.
Feinstein, A. H., & Cannon, H. M. (2001). Fidelity, Verifiability, and Validity of Simulation: Constructs for Evaluation. *Developments in Business Simulation and Experiential Learning*, 28.
Feinstein, A. H., & Cannon, H. M. (2003). A hermeneutical approach to external validation of simulation models. *Simulation & Gaming, 34*(2), 186-197.
George, C. F. P. (2003). Driving simulators in clinical practice. *Sleep Medicine Reviews, 7*(4), 311-320.
Godley, S. T., Triggs, T. J., & Fildes, B. N. (2002). Driving simulator validation for speed research. *Accident Analysis & Prevention, 34*(5), 589-600.
Graafland, M., Schraagen, J. M., & Schrijven, M. P. (2012). Systematic review of serious games for medical education and surgical skills training. *British Journal of Surgery, 33*(10), 1322-1330.
Harteveld, C. (2012). *Making Sense of Virtual Risks.* IOS Press.
Holden, R. R. (2010). Face validity. In I. B. Weiner, & W. E. Craighead (Eds.), *The Corsini Encyclopedia of Psychology: Volume 2 D-L* (4th ed., pp. 637-638). Hoboken, NJ: John Wiley & Sons, Inc.
IEEE Computer Society (2011). *IEEE Recommended Practice for Distributed Simulation Engineering and Execution Process (DSEEP).* IEEE Std 1730TM-2010.
Kleijnen, J. P. C. (1995). Verification and validation of simulation models. *European Journal of Operational Research, 82*(1), 145-162.
Lo, J. C., Van den Hoogen, J. & Meijer, S. A. (2013). Using Gaming Simulation Experiments to Test Railways Innovations: Implications for Validity. In R. Pasupathy, S. H. Kim, A. Tolk, R. Hill & M. E. Kuhl (Eds.), *Proceedings of the 2013 Winter Simulation Conference (WSC)* (pp. 1766-1777).
Mayer, I., Zhou, Q., Lo, J., Abspoel, L., Keijser, X., Olsen, E., Nixon, E., & Kannen, A. (2013). Integrated, ecosystem-based Marine Spatial Planning: Design and results of a game-based, quasi-experiment. *Ocean & Coastal Management, 82*(0), 7-26.
McCoy, W. K. (1963). Problems of validity of measures used in investigating man-machine systems. *Human Factors, 5*, 373-377.
Newgren, K. E., R. M. Stair & Kuehn, R. R. (1981). Decision efficiency and effectiveness in a business simulation. *Developments in Business Simulation and Experiential Learning, 8*(1), 263-265.
Norris, D. R. & Snyder, C. A. (1981). External validation: An experimental approach to determining the worth of simulation games. *Proceedings, Eight ABSEL Conference*, pp. 247-250.
Padmos, P. & Milders, M.V. (1992). Quality criteria for simulator images: A literature review. *Human Factors, 34*(6): 727-748.
Page, L., Lo, J. C., Velazquez, M., & Claudio, D. (in review). Introducing a validation framework to standardize and improve human factors and ergonomics research.
Peters, V., Vissers, G., & Heijne, G. (1998). The Validity of Games. *Simulation & Gaming, 29*(1), 20-30.

Raser, J. C. (1969). *Simulations and Society: An Exploration of Scientific Gaming*. Boston, MA: Allyn & Bacon.
Reimer, B., D'Ambrosio, A., Coughlin, J. F., Kafrissen, M. E., & Biederman, J. (2006). Using self-reported data to assess the validity of driving simulation data. *Behavior Research Methods, 38*(2), 314-324.
Sargent, R. G. (2004). Validation and Verification of Simulation Models. In R. G. Ingalls, M. D. Rossetti, J. S. Smith, & B. A. Peters (Eds.), *Proceedings of the 2004 Winter Simulation Conference*, pp. 17-28. Piscataway, New Jersey: Institute of Electrical and Electronics Engineering, Inc.
Stainton, A. J., Johnson, J. E., & Borodzicz, E. P. (2010). Educational validity of business gaming simulation: A research methodology framework. *Simulation & Gaming, 41*(5), 705-723.
Sterman, J. D. (1987). Testing behavioral simulation models by direct experiment. *Management Science 33*(12), 1572-1592.
Trochim, W. M. K., & Donnelly, J. P. (2008). *Research Methods Knowledge Base* (3rd ed.). USA: Atomic Dog.
Vissers, G., G. Heyne, V. Peters, and J. Geurts. 2001. The validity of laboratory research in social and behavioral science. *Quality & Quantity, 35*, 129-145.
Wampold, B. E. (2006). Designing a research study. In F. T. Leong, & J. T. Austin (Eds.), *The Psychology Research Handbook: A Guide for Graduate Students and Research Assistants* (2nd ed., pp. 93-103). United State of America: Sage Publications, Inc.
Whiteley, T. R., & Faria, A. J. (1989). A Study of the Relationship Between Student Final Exam Performance and Simulation Game Participation. *Simulation & Games, 20*(1), 44-64.
Wiggins, J. A. (1968). Hypothesis validity and experimental laboratory methods. In H. M. Blalock, & A. B. Blalock (Eds.), *Methodology in Social Research* (pp. 390-427). United States of America: McGraw-Hill, Inc.
Wolfe, J., & Roberts, C. R. (1986). The External Validity of a Business Management Game: A Five-Year Longitudinal Study. *Simulation & Games, 17*(1), 45-59.
Zomer, L.-B., Moustaid, E., & Meijer, S. (2015). A meta-model for including social behavior and data into smart city management simulations. *Proceedings of the 2015 Winter Simulation Conference*, Huntington Beach, California.
Zechmeister, J. S., Zechmeister, E. B., & Shaughnessy, J. J. (2001). *Essentials of Research Methods in Psychology*. New York, NY: McGraw-Hill.

3 Connecting mental models and situation awareness to different gaming simulation types: Design and testing requirements

This chapter is based on the following work:
Lo, J. C. & Meijer, S. A. (2014). A Framework to Assess Cognition in Different Types of Gaming Simulations. *Poster Presentation at the Human Factors and Ergonomic Society Europe Chapter*.

Abstract
Reviewing the validity of gaming simulations as discussed in the previous chapter, it can be observed that the validity topic surrounding gaming simulations resides on an interdisciplinary crossing. Depending on the inclusion of simulation models in the gaming simulation, validation approaches from the computer science discipline may need to be considered next to the validation approaches for human research during the gaming simulation session. The current chapter first presents an updated framework that constitutes four main types of purposes for gaming simulations. Secondly, the cognitive implications that are required for each type of purpose are identified and linked with cognitive constructs such as mental models and situation awareness. The aim of the chapter is to provide a structured approach in understanding and recognizing the requirements for conducting human research in games with different purposes.

Keywords: Gaming simulation types; mental models; situation awareness

3.1 Introduction

The purpose of a gaming simulation takes a crucial role in relation to validation. Building on the previous chapter, the validity requirements of a gaming simulation varies with the different types of gaming simulations, i.e. training, policy and research (Peters, Vissers, & Heijne, 1998). In addition to different typologies of games, different terms such as business games are also found, which predominantly have the purpose of training participants (Wolfe, 1994) and additional ways in which games can be framed, e.g. as an intervention method or as a design artifact (Mayer et al., 2014). The latter is a fairly new terminology, which refers to the game as a socio-technical design. Although differently than these authors positioned, design games are also more commonly seen as a tool in the participatory design of products and services (e.g. Kensing & Blomberg, 1998) and more recently as part of the design of complex systems (Meijer, Reich, Subrahmanian, 2014).

With the recognition of design games as a new typology, this chapter first aims to provide an updated framework on different types of gaming simulation. This new framework specifically focuses on typologies of gaming simulations that can be applied in a complex socio-technical system, such as the railways. Following our line of research to investigate human cognition in gaming simulations, an approach is presented that identifies which cognitive components are addressed. As such the main research question in this chapter will be: which cognitive construct can be measured in a gaming simulation with a specific purpose and what are the requirements for the design of this particular type of gaming simulation? The implications of different gaming simulation types on a number of cognitive components, such as mental models and situation awareness, are investigated.

3.2 Four gaming simulation types

An earlier framework by Peters, Vissers, & Heijne (1998) distinguished three purposes for which gaming simulations exist: (1) policy games (e.g. Duke & Geurts, 2004; Geurts, Duke, & Vermeulen, 2007), (2) training games, (e.g. De Freitas & Griffiths, 2007; Garris, Ahlers, & Driskell, 2002) and (3) research games (e.g. Klabbers, 2003; 2006; Vissers, Heyne, Peters, & Geurts, 2001). An additional type of gaming simulation, namely that of design games, has been proposed by Grogan and Meijer (2017), which reflects the use of gaming simulations in organizations. As a slightly different categorization of the framework has a better fit with the current research, the following adaptations are made: instead of the dimension 'knowledge type' with differentiations in whether the knowledge type is generalizable beyond the scope of a particular game scenario or is contextual in the sense that deep insights of a particular game scenario are obtained, the dimension 'type of game outcome' is introduced. Games that differ on this dimension are distinguished between: the creation of knowledge (i.e. learning) versus the creation of policies and artefacts. Secondly, games can differ to the extent that the game outcomes are relevant for either the players or participants themselves, or for the observer who could be a researcher, members of a project team etc. Thus, these dimensions result in four types of gaming simulations (see Figure 3.1).

		Transfer/relevance of game outcome	
		Player	Observer (researcher, project team, ..)
Type of game outcome	Learning (creation of knowledge)	**Training**	**Research**
	Creation of policies and artefacts	**Intervention**	**Design**

Figure 3.1: Four types of gaming simulation (based on Grogan and Meijer, 2017).

3.2.1 Training games

Games for learning or education have so far been the most popular type of gaming simulation in decades, in which 'serious games' bear this connotation under a wider public. The purpose of training games is fairly obvious: players should obtain insights or develop knowledge by playing this game. Psychologists often refer to the development of declarative (e.g. knowledge of what) and procedural (e.g. knowledge of how) knowledge (Anderson, 2010). In the applied psychological (i.e. human factors and ergonomics) field, researchers often also refer to a third category: strategic knowledge (e.g. Mohammed, Ferzandi and Hamilton, 2010; Salas, Stout and Cannon-Bowers, 1994). Here, strategic knowledge refers to the knowledge of what and how applied to the context. This set of three knowledge types is also used in this field as a mental model.

3.2.2 Research games

In research or testing games, the outcomes that are found in a session are not primarily relevant for the participant, but in this case relevant to the observer. Games could be used as a controlled research environment that incorporates more physical characteristics of the environment opposed to a traditional laboratory environment. In addition to the framework, two types of research games can be distinguished: (1) research games with an experimental focus aim at formulating hypotheses, which leads to an explanatory outcome, and (2) research games that are exploratory in nature, with a stronger focus on descriptive results and on obtaining (qualitative) insights from the gaming simulation session (Lo, Van den Hoogen, & Meijer, 2013).

3.2.3 Design games

Design games focus on the creation of elements, such as policies and artefacts, which are relevant for the observer. Due to the applied nature of these types of games, observers are often members of an organization, such as policy-makers, advisors or project members. Design games provide the platform for a participatory environment, in which interactive participation can take place between the user as participants and the designer as observer (Grogan & Meijer, 2017).

3.2.4 Intervention/policy games

The final gaming simulation type is that of an intervention game. Although an intervention game resembles much of the characteristics of a design game, the main difference is that the outcomes of the game have implications for the participants themselves. While design games are more future oriented due to the lead role of observers in translating the outcomes, intervention games are rather short-term oriented as players can implement their designs or solutions directly.

Although these four games all have different purposes, they all make use of game design principles that eventually lead to a configuration that characterizes the gaming simulation. The next section introduces these game design principles.

3.3 Gaming simulation design principles

For the current approach, the meta-framework from Meijer (2008) is used to assess different game characteristics. The following game (design) components can be derived from the framework with examples from the railway traffic domain:

- Roles: the personification of an individual within the game, e.g. train traffic controllers playing their own role of a train traffic controller or a colleague's role, for instance that of a train driver
- Rules: the specific and general do's and don'ts for the different roles, e.g. the extent to which a train traffic controller needs to execute his/her own task as a train traffic controller in the gaming simulation
- Objectives: the goal(s) of an individual or multiple roles, e.g. goals related to their own tasks
- Constraints: the range of possible actions in the game, e.g. limitations to collaborate with other colleagues due to absence of certain roles

In addition to the game design elements, parameter settings need to be defined:

- Load: difficulty of the roles, rules, objectives and constraints, which affects the difficulty of the gaming simulation
- Situation: external variables that might influence the session, e.g. location, selection of participants, e.g. the amount of game facilitators or colleagues or managers who want to be present during a gaming simulation session

Cognitive processes of participants are influenced by the extent to which a participant is familiar with the task or needs to learn this task. For instance, a student in civil engineering who needs to play the role of a train traffic controller most likely first needs to learn all the basic principles of a train traffic controller's task. It is therefore highly questionable as to what extent he/she will be able to participate in a research game that is supposed to investigate what the implications are on cognitive activities, such as their situation awareness.

3.4 Theoretical implications of game design on mental models and situation awareness

In line with our application to employ gaming simulations in railway traffic control, cognitive structures such as mental models (MM) and situation awareness (SA) are particularly investigated in this context as they can be used as an indicator for participant's decision-making, behavior and performance. Table 3.1 presents an overview of the definitions of mental models and situation awareness, and the differences between the two cognitive constructs in terms of knowledge and certain requirements for game design choices.

In terms of difference in knowledge, mental models can be seen as abstract knowledge structures, while situation awareness is the application of this knowledge in a concrete situation. Consequently, this implies that a well-developed and solid knowledge base is needed in order to guarantee for a high situation awareness. In other words, in order to be able to develop a good situation awareness no strong development in the mental model should take place. Finally, as situation awareness is characterized by a dynamic operational environment, a continuous time flow is another requirement for establishing valid levels of situation awareness in a simulated environment. On the other hand, studies on strategic decision-makers involve the assessment of mental models (e.g. Klimoski & Mohammed, 1994; Schwenk, 1988).

Table 3.1: Definition and differences between two cognitive structures.

	Mental models	**Situation awareness**
Definition	"mechanisms whereby humans are able to generate descriptions of system purpose and form, explanations of system functioning and observed system states, and predictions of future states" (Rouse & Morris, 1985, p.7)	"the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" (Endsley, 1988, p.97). Situation awareness is developed from an existing mental model
Game design in terms of knowledge	Abstract (strategic) knowledge that is built from declarative and procedural knowledge	Applied knowledge in a specific situation that is built on mental models, often more relevant in an operational context. Makes use of short-term and long-term memory.
	Knowledge can be similar or different from the real environment (e.g. different roles, rules etc.)	Knowledge should be similar to the real environment (same roles, rules, objectives, constraints) (no or little change in mental model)
Game design in terms of time frames	No specific requirement for the design of time (e.g. step-wise, continuous time flow)	Continuous time flow in order to reflect the dynamic characteristics of a complex system for operational decision-making

3.5 Cognition in different gaming simulation types

The previous sections described different gaming simulations types and requirements for mental models and situation awareness in terms of gaming simulation design. The current section connects the concepts of mental models and situation awareness and their gaming simulation requirements to the different gaming simulations types. Figure 3.2 illustrates the results of the mapped concepts.

		Transfer/relevance of game outcome	
		Player	Observer (researcher, project team, ..)
Type of game outcome	Learning (creation of knowledge)	**Training** Mental model development	**Research** Exploratory: expert mental model Explanatory: situation awareness
	Creation of policies and artefacts	**Intervention** Novice or expert mental model	**Design** Novice or expert mental model

Figure 3.2: Gaming simulation conditions and cognitive structures in four types of gaming simulations.

Two propositions are introduced in relation to the mapped concepts.

Proposition 1: mental models or situation awareness are not necessarily impacted by the representation level of a gaming simulation
As an initial assumption, we posit that the representation level of a gaming simulation does not necessarily impact the mental models or situation awareness of players, depending on validity levels in terms of structural and process validity and psychological reality (Lo & Meijer, 2014; Raser, 1969).

Proposition 2: gaming simulation type and design have implications on the level of representation of operator's mental models and/or situation awareness
Secondly, we posit that the selected purpose and choices in the game design do have implications on whether mental models and/or situation awareness can be assessed. The extent to which the simulated system needs to incorporate the characteristics of a future reference system could affect the mental model development of operators. These changes could take place in the product, social or institutional changes, respectively introducing for instance a new interface, new roles or new procedures (see Lo, Van den Hoogen, & Meijer, 2014). Subsequently, these changes are implemented in the design principles such as roles, rules, objectives, which impact the extent to which participants need to learn their new roles, and to which extent mental model development takes place during the gaming simulation session. As such, we posit that it is possible to establish and assess mental models of participants in all four types of gaming simulations. However, only in research gaming simulations is it possible to establish and assess valid levels of situation awareness.

For training gaming simulation, it is apparent that the purpose of this gaming simulation type is on mental model development of the participant, i.e. learning. Without a solid mental model base, it is therefore not possible to establish valid levels of situation awareness that can be expected in comparison to the actual work setting.

Research gaming simulation can facilitate both mental models and situation awareness of participants. For explanatory research gaming simulation it is required for the simulated system to have a high ecological validity in order to investigate variables in the gaming simulation session in order to adhere to the rigid experimental conditions for hypothesis testing. This implies that cognitive activities should be comparable to those in the reference system, and that well-developed mental models are therefore a prerequisite for research games. Subsequently, valid levels of situation awareness are expected to be established. However, as the game design choices can vary, for instance in terms of step-wise rounds instead of a continuous time-flow or requirements for measuring situation awareness may not be met. In exploratory research gaming simulation hypothesis generation is desired based on rich, contextual, (qualitative) data. As such, expert mental models are *minimally* needed and valid levels of situation awareness would be wished, but not demanded for.

As design gaming simulations have the purpose for participants to come up with a new configuration of the system, they do not follow regular cognitive information-processing demanding activities following their operational decisions as they would in regular conditions. Other cognitive processes may even be encouraged, such as focusing on creativity and critical reflections related to the current system. As such, these types of gaming simulations are not suitable to facilitate situation awareness. Further on, non-subject matter experts may participate, which implies a novice mental model base to draw new ideas.

Finally, intervention gaming simulations similarly focus on identifying a new configuration of the system, in which participants directly could apply the solution into the reference system. Similar cognitive activities may be expected as with design gaming simulation, which facilitate the use of novice or expert mental models.

3.6 Discussion and conclusion

This chapter introduced an updated framework on four gaming simulation types: training, research, design and intervention gaming simulations. Different purposes and characteristics were used as a foundation to related relevant cognitive constructs such as mental models and situation awareness. We posited that a number of varying configurations can trigger different cognitive components that are relevant to invoke the conditions for mental models and situation awareness to establish. The current approach emphasizes that solid mental model bases are key determiners in assessing mental models and/or situation awareness can be assessed in a gaming simulation. As such, we posited that valid levels of situation awareness can only be established in research gaming simulations. Although a specific type of gaming simulation can be related to certain game design 'pre-sets' it is not a prerequisite. For instance, design games can have a high ecological validity.

References

Anderson, J. R. (2010). *Cognitive Psychology and Its Implications*. Worth Publishers: New York, NY.

De Freitas, S., & Griffiths, M. (2007). Online gaming as an educational tool in learning and training. *British Journal of Educational Technology, 38*(3), 535-537.

Duke, R. D., & Geurts, J. L. A. (2004). *Policy Games for Strategic Management*. Dutch University Press: Amsterdam.

Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement. *Proceedings of the 32nd Human Factors Society Annual Meeting, 32*, 97-101.

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming, 33*(4), 441-467.

Geurts, J. L. A., Duke, R. D., & Vermeulen, P. A. M. (2007). Policy gaming for strategy and change. *Long Range Planning, 40*(6), 535-558.

Grogan, P. T., & Meijer, S. A. (2017). Gaming Methods in Engineering Systems Research. *Systems Engineering, 20*(6), 542-552.

Kensing, F., & Blomberg, J. (1998). Participatory Design: Issues and Concerns. *Computer Supported Cooperative Work (CSCW), 7*(3), 167-185.

Klabbers, J. H. G. (2003). Simulation and gaming: Introduction to the art and science of design. *Simulation & Gaming, 34*(4), 488-494.

Klabbers, J. H. G. (2006). A framework for artifact assessment and theory testing. *Simulation & Gaming, 37*(2), 155-173.

Klimoski, R., & Mohammed, S. (1994). Team mental model: Construct or metaphor? *Journal of Management, 20*, 403–437.

Lo, J. C. & Meijer, S. A. (2014). Gaming Simulation Design for Individual and Team Situation Awareness. In S. A. Meijer & R. Smeds (Eds.), *Frontiers in Gaming Simulation* (pp. 121-128). Berlin: Springer.

Lo, J. C., Van den Hoogen, J. & Meijer, S. A. (2013). Using Gaming Simulation Experiments to Test Railways Innovations: Implications for Validity. In R. Pasupathy, S. H. Kim, A. Tolk, R. Hill & M. E. Kuhl (Eds.), *Proceedings of the 2013 Winter Simulation Conference (WSC)* (pp. 1766-1777).

Lo, J. C., Van den Hoogen, J. & Meijer, S. A. (2014). Testing Changes in the Railway System through Gaming Simulation: How Different Types of Innovations Affect Operators' Mental Models. In T. Ahram & T. Marek (Eds.), *Proceedings of the 5th International Conference on Applied Human Factors and Ergonomics (AHFE)* (pp. 8054-8065).

Mayer, I., Bekebrede, G., Harteveld, C., Warmelink, H., Zhou, Q., van Ruijven, T., Lo, J., Kortmann, R. & Wenzler, I. (2014). The research and evaluation of serious games: Toward a comprehensive methodology. *British Journal of Educational Technology, 45*(3), 502-527.

Meijer, S. A. (2008). *The Organisation of Transactions: Studying Supply Networks using Gaming Simulation*. Wageningen: Academic Publishers.

Mohammed, S., Ferzandi, L., & Hamilton, K. (2010). Metaphor No More: A 15-Year Review of the Team Mental Model Construct. *Journal of Management, 36*(4), 876-910.

Meijer, S., Reich, Y. & Subrahmanian, E. (2014). The future of gaming for design of complex systems. In R.D. Duke & W. Kriz (Eds.) *Back to the Future of Gaming*, pp.154-167.

Nefs, M., Gerretsen, P., Dooghe, D., Mayer, I. S., & Meijer, S. (2010, 11-13 Nov. 2010). Gaming the interrelation between rail infrastructure and station area development: Part 1 - modeling the serious game 'SprintCity'. Paper presented at the Third International Conference on Infrastructure Systems and Services: Next Generation Infrastructure Systems for Eco-Cities (INFRA).

Peters, V., Vissers, G., & Heijne, G. (1998). The Validity of Games. *Simulation & Gaming, 29*(1), 20-30.

Raser, J. C. (1969). *Simulations and Society: An Exploration of Scientific Gaming.* Boston, MA: Allyn & Bacon.

Rouse, W. B., & Morris, N. M. (1985). On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin, 100*(3), 349-363.

Salas, E., Stout, R. J., & Cannon-Bowers, J. A. (1994). The role of shared mental models in developing shared situational awareness. In R. D. Gilson, D. J. Garland & J. M. Koonce (Eds.), *Situational Awareness in Complex Systems* (pp. 297-304). Daytona Beach, Florida: Embry-Riddle Aeronautical University Press.

Schwenk, C. R., (1988). The Cognitive Perspective on Strategic Decision Making. *Journal of Management Studies 25*(1), 41-55.

Vissers, G., Heyne, G., Peters, V., & Geurts, J. (2001). The validity of laboratory research in social and behavioral science. *Quality and Quantity, 35*(2), 129-145.

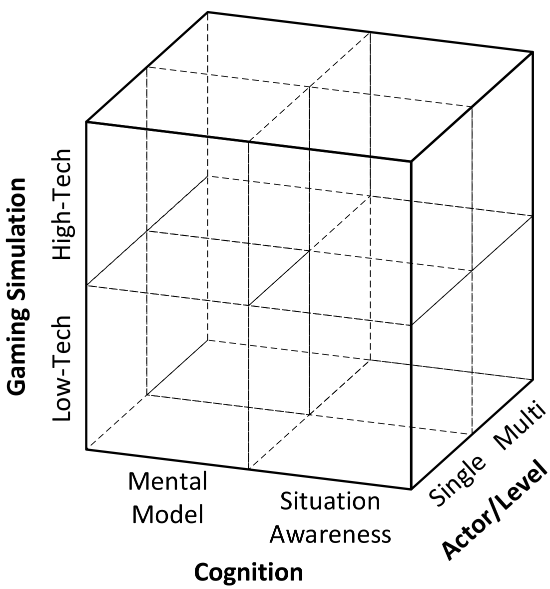
Wolfe, J. (1994). Recollections on 25 years of simulation/gaming. *Simulation & Gaming, 25*(2), 274-278.

Section II

Human factors research on railway traffic control using gaming simulations

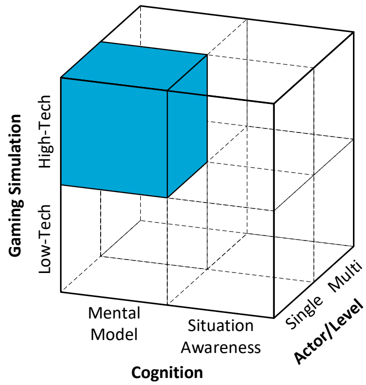
This section is comprised of several human factors studies of the railway traffic control domain using low- and high-tech gaming simulation. The studies focus on the investigation of cognitive components, specifically on the mental models and situation awareness of single and multiple operators. The main nature of these studies is descriptive, in which multiple cognitive theories are addressed.

Chapter 4 - Single / high-tech / mental model
Chapter 5 - Single / high-tech / situation awareness
Chapter 6 - Multi / high-tech / situation awareness
Chapter 7 - Multi / low-tech / situation awareness
Chapter 8 - Multi / low-tech / mental model



4 Individual markers of resilience in train traffic control: The role of operators' goals and strategic mental models and implications for variation, expertise, and performance

This chapter has been published as:
Lo, J. C., Pluyter, K. R., & Meijer, S. A. (2015). Individual Markers of Resilience in Train Traffic Control: The Role of Operators' Goals and Strategic Mental Models and Implications for Variation, Expertise, and Performance. *Human Factors, 58*(1), 80-91.



Abstract
The chapter focuses on the topic of resilience in train traffic control, in which resilience is investigated through train traffic operator's goals and strategic mental models as individual markers. Train traffic controllers (N=22) enacted two scenarios in a human-in-the-loop simulator. Their experience, goals, strategic mental models, and performance were assessed through questionnaires and simulator logs. Goals were operationalized through performance indicators and strategic mental models through train completion strategies. A variation was found between operators for both self-reported primary performance indicators and completion strategies. The primary goal of only 14% of the operators reflected the primary organizational goal (i.e., arrival punctuality). An incongruence was also found between train traffic controllers' self-reported performance indicators and objective performance in a more disrupted condition. The level of experience tends to affect performance differently. All in all, the results show a gap between primary organizational goals and preferred individual goals. Further, the relative strong diversity in primary operator goals and strategic mental models indicates weak resilience at the individual level.

Keywords: Resilience; organizational goals; operator goals; mental model

4.1 Introduction

Resilience engineering studies are relevant in multiple domains, especially in those that are highly complex and known for their hazards (Nemeth, Wears, Patel, Rosen, & Cook, 2011). Domains that are most heavily investigated are aviation (22%), health care (19%), the chemical and petrochemical industry (16%), nuclear power plants (10%), and railway (8%) (Righi, Saurin, & Wachs, 2015).

For the Dutch railway infrastructure managing organization ProRail, the notion of resilience and robustness strongly resonates in the organization to improve the system along these concepts (Meijer, 2012). The idea is that when the system cannot maintain the regular way of working, resilience is required to respond through the adaptation of strategies (Burnard & Bhamra, 2011; Hollnagel, 2014; Hollnagel & Woods, 2005).

However, resilience is linked across different levels with influencing mechanisms on the industry at the highest level, followed by plant and operations (organization), and teams and individuals at the lowest level (Back, Furniss, Hildebrandt, & Blandford, 2008; Sheridan, 2002). Research often focuses on a specific unit of analysis, as it is not yet well understood how resilience is linked across these different levels (Righi et al., 2015). The study of cross-level interactions inside the system is, however, crucial to prevent brittleness in the overall system, which can be facilitated through proactive safety management (Gomes, Woods, Carvalho, Huber, & Borges, 2009).

An analysis of railway safety operations in the Netherlands revealed poor to mixed resilience levels (Hale & Heijer, 2006). The debundling and privatization of the railway system that was widely introduced across Europe in the 1990s, causing extensive institutional fragmentation of the system, is a possible reason for the low resilience levels (Hale & Heijer, 2006; Knieps, 2013). The debundling of the railway system inextricably led to more brittle operational processes for railway traffic operators, resulting in, for example, unclear and conflicting goals and the development of multiple coping strategies (Steenhuisen & De Bruijne, 2009; Veeneman, 2006). This phenomenon can be also be labeled as a gap between the system as designed or imagined and the system as it is actually operated, which results in a distance between the various levels (Dekker, 2006).

At an individual level, resilience engineering can help operators to develop robust yet flexible responses to disturbances inside or outside the organization (Chialastri & Pozzi, 2008; Lengnick-Hall & Beck, 2005). As such, performance variability is normal, though it needs to be controlled. Performance variability that leads to positive outcomes should be promoted (Hollnagel, 2008, 2014). Having shared goals and experiences, robust responses to simple problems, and flexible responses to complex problems is essential to the development of a resilient organization (Lengnick-Hall, Beck, & Lengnick-Hall, 2011).

Departing from resilience studies in the Dutch railways at a system and organizational level, this study focused on the individual level of railway traffic operators in order to provide recent and quantitative insights to further the understanding of variations in their cognition and behavior and the implications thereof. The central research questions were as follows: To what extent do organizational and individual goals correspond? What is the level of diversity in the goals and strategic mental models of train traffic operators given operators' work experience, and how does it relate to their performance?

The following section briefly introduces the Dutch railway system from a number of perspectives. This overview is followed by a brief presentation of the theoretical background to goals and strategic mental models. The subsequent sections present the method, results, and discussion and conclusion.

4.2 A multilevel overview: railway transport in the Netherlands

The Dutch railways transport more than 1 million passengers and operate about 350 freight trains per day in a relatively small country, making it one of the busiest railway systems in Europe and even the world (Meijer, 2012; ProRail, 2013; Ramaekers, De Wit, & Pouwels, 2009). The debundling and privatization of the Dutch railway sector in 1995 initially led to a decrease in the performance of the system: Between 2000 and 2001, arrival punctuality dropped from about 87% to 80% (Algemene Rekenkamer, 2012; Steenhuisen, 2009). Although the punctuality of trains has recovered over the years, reaching 92% in 2014, both the principle passenger transport manager (Nederlandse Spoorwegen [NS]) and the infrastructure manager (ProRail) were penalized for performing insufficiently on the agreed performance indicators: NS for passenger dissatisfaction in terms of punctuality and quality of service (e.g., number of available seats, crowdedness during peak hours) and ProRail for the insufficient availability of the infrastructure due to malfunctions (Rijksoverheid, 2015).

4.2.1 Organizational performance indicators

Safety, reliability, service, and capacity use can be seen as key public values in the railway domain (Wilson, Farrington-Darby, Cox, Bye, & Hockey, 2007). The general public values that are held in the governance of railway transport are rather stable over time, unlike the operationalization and quantification of these values into goals or performance indicators (Veeneman & Van de Velde, 2006). For instance, reliability can be conceptualized in a number of ways, such as punctuality, which can be further operationalized in terms of, for example, arrival, departure, or overall (arrival and departure) punctuality. Departure punctuality was a performance indicator until 2006, when arrival punctuality became the indicator (Veeneman, 2006). However, both railway infrastructure and passenger transport managers set different thresholds in arrival punctuality, namely, <3 min and <5 min, respectively (NS, 2015; ProRail, 2015a). The formalization of performance indicators is an annual iterative process with occasionally ad hoc organizational reactions throughout the year in the case of unexpected large-scale disruptions that are subject to media scrutiny.

4.2.2 Train traffic control

Railway traffic operations differ between European countries in a number of ways, such as organization, roles and responsibilities, and level of automation (Golightly et al., 2013). In the Netherlands, a train traffic management system is used to execute the timetables, which are operated by train traffic controllers. The primary responsibility of these controllers is to execute train timetables in an accurate and punctual manner (Sulmann, 2000). Maintaining the operational safety of the rail system and recovering after disruptions and accidents is an essential part of their job (Crawford, Toft, & Kift, 2014). Train traffic controllers do not perceive their primary task as challenging as long as routes are already scheduled (Roth & Patterson, 2005). However, a more active role is needed in unsafe situations that cannot be controlled by the automated safety system or when there is a system malfunction (Sulmann, 2000).

4.2.3 Future developments

In terms of future developments, ProRail and the government stated their intention to double the railway track capacity between 2008 and 2020 (now extended to 2028), which should lead to a timetable that supports both an intercity and a local train service six times per hour in both directions between a number of major cities (Meijer, Van der Kracht, Van Luipen, & Schaafsma, 2009; ProRail, 2015b). Given the restriction of a capacity increase through the mere addition of tracks, a change in the organizational processes is also required. As such, process optimization programs are being implemented that focus on, for instance, increasing the centralization of decision making to the national control center (operational control center rail [OCCR]) for disruption mitigation procedures and restructuring the roles and responsibilities of operators. Switches are increasingly being removed at major stations (e.g., 110 of the 170 switches are being removed at Utrecht Central Station) in order to, for example, facilitate corridor management, shorter travel times, and more reliable traffic control, while bottleneck areas in the infrastructure are being expanded and upgraded. Finally, the replacement of the current traffic management system is being explored.

4.3 Goals and mental models

4.3.1 Goals

Goals are states or ends that someone wants to achieve (Latham & Locke, 1991; Mohammed, Klimoski, & Rentsch, 2000; Popova & Sharpanskykh, 2011). Operators' goals influence their mental model selection and therefore their decision making and performance (Endsley, 1995). In a dynamic environment, individuals focus on elements in the environment that are goal related. Deriving the meaning of the elements and the projection to the future is done in light of the goal and the active mental models (Endsley, 1995). Goals influence the valuation of multiple options during decision making (Mohammed et al., 2000). In order to achieve resilience, operators need to have a common set of goals (Lengnick-Hall et al., 2011).

4.3.2 Mental Models and expertise

Mental models are mental representations of humans, systems, artifacts, and situations formed by experience, observation, and training (Endsley, 1995; Schaffernicht & Groesser, 2011; Wilson, 2000). Mental models store knowledge that is necessary for human-environment interaction (Klimoski & Mohammed, 1994; Mathieu, Heffner, Goodwin, Salas, & Cannon-Bowers, 2000). This knowledge is crucial for effectively solving problems, such as those faced by train traffic controllers when confronted with multiple disruptions to the train schedule. Visual attention and evaluation of relevant information in complex problem situations improve when mental models are well developed.

The degree of development of mental models differs between novices and experts. Experts with extensive domain knowledge have developed the ability to perceive important patterns and features that are not seen by novices (Bogard, Liu, & Chiang, 2013; Bransford, Brown, & Cocking, 2000; Glaser & Chi, 1988). Experts also have the capacity to better recognize meaningful patterns due to their superior knowledge organization and extensive domain knowledge (Glaser & Chi, 1988). In contrast, novices' knowledge consists of facts, procedures, and formulas that are not as well organized, as they do not have integrated mental models. Novices are therefore oriented toward surface characteristics in problem solving (Bogard et al., 2013; Glaser & Chi, 1988). Furthermore, experts have developed a condition–action ability through practice. Experts have conditioned knowledge: The recognition of specific patterns triggers an appropriate response that is useful for problem solving (Bransford et al., 2000; Glaser & Chi, 1988). Different levels of expertise may influence the performance of train traffic controllers and therefore resilience at an individual level (Lengick-Hall et al., 2011).

4.4 Method

4.4.1 Experimental setting

A simulator session was used to familiarize train traffic controllers with the new infrastructure that would result from the removal of 66 switches in 3 months' time. The simulator was strongly focused on the logistical aspects of train traffic control and much less on technical safety-related aspects. The infrastructure that was simulated was the train traffic area around Utrecht Central Station. This area is operated by two train traffic workstations. One controller was responsible for the trains that belong to the "turn" (in Dutch, *keer*) area, and a second controller was responsible for the “through” (in Dutch, *door*) area. The role allocation was reversed in the second round.

Two scenarios were designed for the participants: Scenario 1 consisted of a light disruption in the train traffic flow caused by minor train delays, whereas Scenario 2 represented a moderately to severely disrupted flow. In the first round, train traffic controllers participated in Scenario 1. In the second round, 10 participants participated in Scenario 1 and 12 participated in Scenario 2. Both scenarios were designed in collaboration with two senior train traffic controllers. Train traffic controllers were asked to perform their job as they typically would at their actual workstation. No interaction between the train traffic controllers was needed to conduct their tasks.

4.4.2 Participants

All 22 train traffic controllers (18 males, four females) worked at Utrecht Central Station. They took part in a 2 (workstation area: turn or through) × 2 (severity of disruption: high vs. low level of train delays) within-subject experimental design.

4.4.3 Materials

Work experience and *job role* were assessed using questionnaires. Participants were assigned to a high- or a low-experience group based on their work experience as train traffic controllers. The cutoff point was set at 10 years, as a new traffic management system had been implemented 10 years earlier (Bary, 2015).

Operator goals were operationalized through performance indicators (Popova & Sharpanskykh, 2011). A list of performance indicators for train traffic controllers was created prior to this session by six senior train traffic controllers. Participants ranked these performance indicators on a scale of 1 to 7 (1 = most important, 7 = least important).

Speed of acquaintance was included to find out how fast the participants were able to get accustomed to the new infrastructure. This item was measured on a 5-point Likert scale, ranging from fully disagree to fully agree. Participants could also opt for I do not know as an answer.

Performance was measured using five performance indicators, namely, arrival punctuality, departure punctuality, amount of arrival delay, amount of departure delay, and platform consistency. Arrival and departure punctuality was operationalized through trains that arrive at (or depart from) Utrecht Central Station on time or with less than a 3-min delay. These trains were counted, summed up, and divided by the total number of arrived/departed trains. For the arrival and departure delay in minutes, the amount of delays in minutes was summed up and divided by the total number of arrived/departed trains. With regard to platform consistency, all trains that did not arrive at the planned track were counted and summed up, and the same was done for all trains that did not arrive at the planned platform. Second, the total number of trains that did not arrive at the planned platform and at the planned track were summed up and divided by the total number of arrived trains for each train traffic controller.

Strategic mental models. Mental models can be conceptualized as declarative (knowledge of what), procedural (knowledge of how), or strategic (knowledge of what and how, and applied to the context) (Mohammed, Ferzandi, & Hamilton, 2010; Salas, Stout, & Cannon-Bowers, 1994). Strategic mental models can also be operationalized by generating lists of actions with subject matter experts (Webber, Chen, Payne, Marsh, & Zaccaro, 2000). As such, the completion strategies of a train traffic controller could be an indicator of the controller's strategic mental model. Simulator logs were used to analyze the completion strategies when different ways of dealing with the train delays (i.e., the different order of departure of trains that were handled given their delay) were expected to be possible. Given the length of Scenario 1, three conflict points for completion strategies for the through workstation and one conflict point for the turn workstation were identified; for Scenario 2, one and two completion strategies were identified for the through workstation and the turn workstation, respectively. Different completion strategies were subsequently assessed by analyzing whether the completion strategies were followed according to the preferred completion strategy (as was scheduled) and the different strategies applied, to assess the variability per operator and per conflict point. Analyses were done based on participants who enacted Scenario 1 in both rounds and those who enacted Scenario 1 and subsequently Scenario 2, in order to obtain four conflict points per individual.

Simulator validity was measured through three components—structural validity (the degree of similarity in structure between the simulated and the reference system), processes validity (the degree of similarity in processes between the simulated and the reference system), and psychological reality (the degree to which the participants perceived the simulated system as realistic)—in line with Raser (1969), using a questionnaire designed by Lo, Sehic, and Meijer (2014). An example of a structural validity item is "I can apply the information from the information sources in the simulator in a similar way as in the real world" (α = .65 with the removal of one item). The item "The train traffic flow in the simulator is similar in their [sic] processes to the real world train traffic flow" represented process validity (α = .60). An example of psychological reality (α = .67) is "The simulation environment feels more or less like my own work environment." These items were measured on a Likert scale.

4.4.4 Procedure

The participants completed a questionnaire before the start of the session. They then enacted the two 40-min scenarios. At the end of each round, they completed another questionnaire. During the second round, knowledge probes were administered for the purpose of another study. Video recordings were made throughout both sessions.

4.5 Results

Six of the 22 participants were excluded from the simulator data analysis because they had known about the train delays. Another two participants were excluded as they enacted Scenario 2 twice. As there were a few problems with the simulator, not all train traffic controllers received the same number of trains. This issue was controlled for by using an average score of the objective performance measures and reviewing the severity of issues through video recordings for events that hindered participants in their options or decision making.

The average score of the participants' work experience in their current function was 10.3 years (*SD* = 9.24).

4.5.1 Simulator validity

The findings show that the participants tended to be slightly positive about the validity of the simulator considering the task they were given (see Table 4.1). The participants also indicated that they had quickly gotten used to the simulator.

Regarding learning effects between scenarios, the participants indicated that they had gotten used to both workstations relatively quickly.

Table 4.1: Descriptive statistics of the validity of the simulator on the three validity components and speed of acquaintance with the simulator and the two workstations, measured on a 5-point Likert scale.

	N	M	SD
Structural validity	21	3.5	.92
Process validity	20	3.6	.66
Psychological reality	22	3.7	.71
Speed of acquaintance with the simulator	20	4.2	.83
Speed of acquaintance with the turn workstation	21	3.9	.62
Speed of acquaintance with the through workstation	21	4.0	.59

4.5.2 Goals

Figure 4.1 shows a relative moderate goal consistency among the train traffic controllers. Three controllers added two more performance indicators, but they were not included in the analysis.

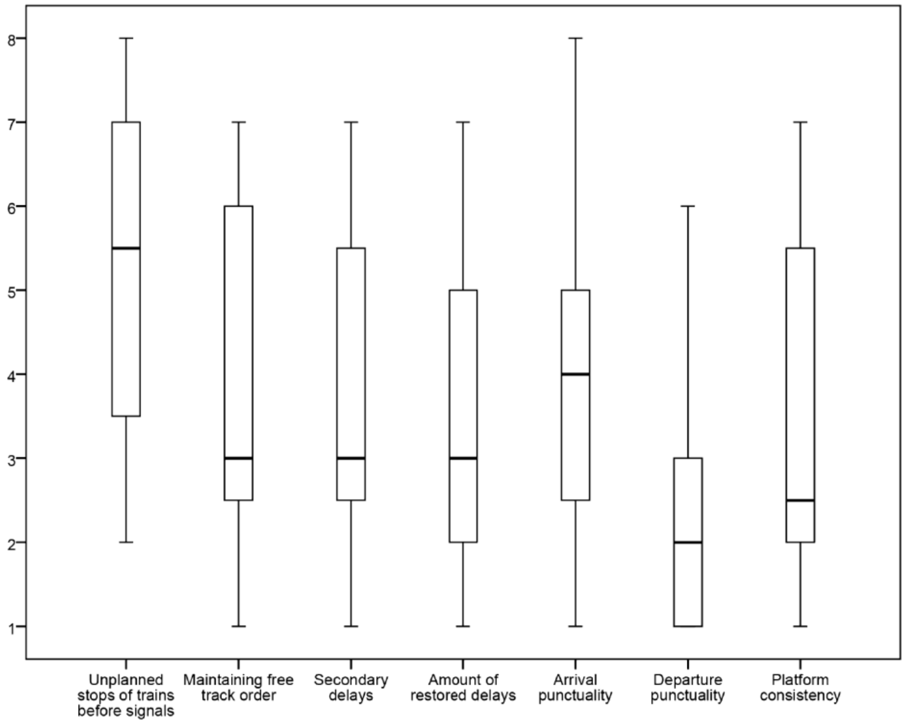


Figure 4.1: Median distribution of self-reported performance indicators (x-axis) with the ranking scale depicted on the y-axis (N = 20).

In the assessment of primary preferred performance indicators (N = 17), however, departure punctuality was consistently perceived as most important (36%). This indicator was followed by achieving high platform consistency (18%), arrival punctuality (14%), maintaining free track order (i.e., track use between stations in the planned order; 9%), the number of restored delays and secondary delays (both 5%), and the avoidance of unplanned stops of trains before signals (0%). As such, these results show a very fragmented preference with regard to primary key performance indicators.

4.5.3 Strategic mental models

The operators' strategic mental models were analyzed to obtain insights into the diversity of their individual completion strategies. The overall findings show that participants handled on average 61% of the completion strategies in the preferred manner (*SD* = 31.5). Those who enacted Scenario 1 twice handled 53% of the completion strategies in a deviating manner (*SD* = 21.1). Participants who enacted Scenarios 1 and 2 handled on average 37% of the completion strategies in the preferred manner (*SD* = 14.2) and 65% in an alternative manner (*SD* = 24.2) (see Figure 4.2). Based on Figure 4.2, a qualitative assessment supports the variation in completion strategies with regard to the operators' individual completion strategies.

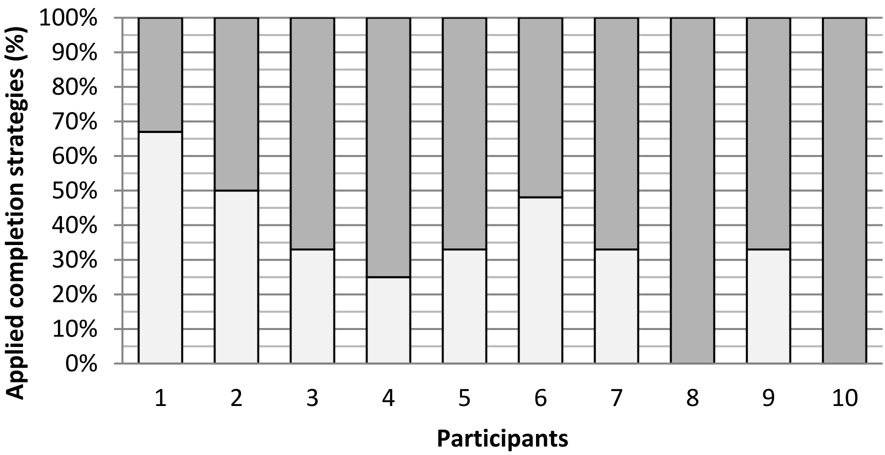


Figure 4.2: Applied completion strategies per participant for operators who enacted Scenarios 1 and 2. A white band indicates a preferred completion strategy being followed, and a gray band indicates alternative completion strategies. Even numbers represent participants from the through workstation; odd numbers, those from the turn workstation.

An analysis of the level of variation in completion strategies for each conflict point revealed diversity based on between one and three different completion strategies for four conflicting points in Scenario 1 and on five different variations of completion strategies for three conflicting points in Scenario 2 (see Figure 4.3). A qualitative assessment would show that there is a level of variation in the completion strategies with regard to different conflict points and that this variation differs between scenarios: Operators dealt with these conflict points with more diverse completion strategies in the moderately disrupted scenario than in the lightly disrupted scenario. Further, it is notable that preferred completion strategies were implemented more frequently in Scenario 1.

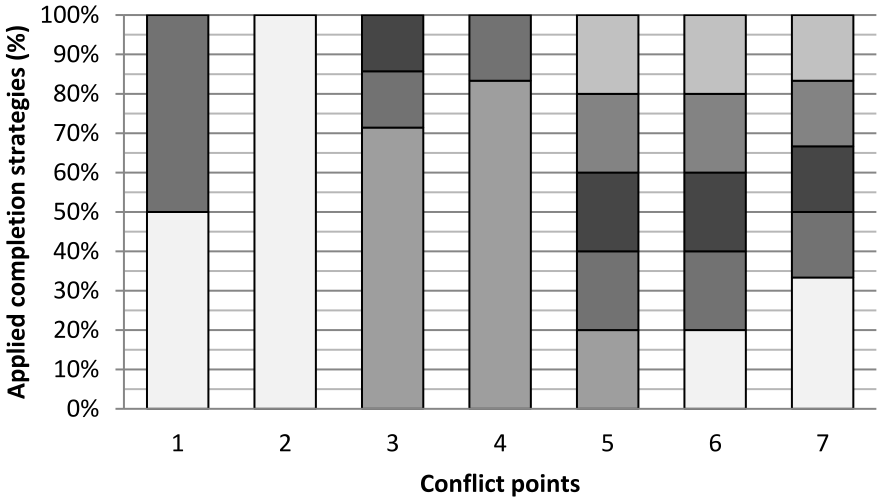


Figure 4.3: Applied completion strategies per conflict point for Scenario 1 (1–4) with N = 14 and Scenario 2 (5–7) with N = 10. A white color indicates a preferred completion strategy being followed, whereas different shades of gray indicate different completion strategies. Numbers 1, 2, 5, and 6 represent conflict points from the through workstation, and numbers 3, 4, and 7 represent conflict points for the turn workstation.

4.5.4 Performance

Spearman correlation tests were performed to test whether there is a congruence between the self-reported relative importance of performance indicators and objective performance (see Table 4.2). Although Scenario 1 does not reveal any significant correlations, Scenario 2 does, namely, a strong positive correlation between self-reported departure punctuality and objective arrival delay. Also a strong negative correlation was found between self-reported arrival punctuality and objective departure delay. A trend for a negative correlation between self-reported platform consistency and objective arrival delay was also found.

Although unexpected, these results provide interesting insights into goal competition, as they suggest that arrival punctuality and departure delay, departure punctuality and arrival delay, and platform consistency and arrival delay are competing goals.

A Spearman correlation test was also performed between the applied preferred and alternative completion strategies and performance. No significant relations were found.

Table 4.2: Correlation between self-reported performance indicators and objective performance indicators for scenario 2.

Self-reported performance indicator	Objective performance indicator	N	r	p
Departure punctuality	Arrival delay	10	.79	.007**
Arrival punctuality	Departure delay	10	-.73	.018*
Platform consistency	Arrival delay	9	-.59	.097

*p ≤ .01, **p ≤ .05

4.5.5 Experience

It was expected that the more experienced controllers would outperform the less experienced controllers due to their better organized mental models. The analyses showed a significant tendency in Scenario 1 for controllers with less experience in their current function to have a higher arrival punctuality score than the more experienced controllers (see Table 4.3). An opposite tendency was found in Scenario 2: The controllers with more experience in their current function have a higher arrival punctuality score than the controllers with less experience in their current function.

It was also investigated whether the applied completion strategies and performance indicators preference differed between the high- and the low-experience group. No significant difference was found for the variation in applied completion strategies, indicating that both more and less experienced operators show diversity in their completion strategies. For the different primary performance indicators, a trend was found for a difference in the importance of maintaining free track order ($U = 17.0$, $p = .073$) and unplanned stops of trains before signals ($U = 20.5$, $p = .095$). More experienced operators indicated these goals as being more important compared to the less experienced operators.

This finding is possibly because both goals are felt to be of importance to achieve a good performance in their train traffic operations, whereas less experienced controllers do not yet feel so.

Table 4.3: Differences in objective performance between more and less experienced train traffic controllers in their current function.

Scenario	Objective performance indicator	Experience	Mean Rank	N	U	p
1	Arrival punctuality	Low	13.8	11	13.0	.005*
		High	6.4	9		
2	Arrival punctuality	Low	3.9	5	4.5	.09
		High	7.1	5		

*p ≤ .01, **p ≤ .05

4.6 Discussion and conclusion

We investigated the following research questions: To what extent do organizational and individual goals correspond? And what is the level of diversity in the goals and strategic mental models of train traffic operators given operators' work experience, and how does it relate to their performance?

First, the level of correspondence between organizational and individual goals was explored. This correspondence appeared to be moderate when looking at the median distribution. However, when assessing the preference for arrival punctuality, this goal ranked in the third position, with 14% of the controllers adhering to the primary organizational goal. Operators indicated that they valuate departure punctuality (36%) and platform consistency (18%) as more important than the primary organizational goal, arrival punctuality. The low absolute percentages spread over multiple goals revealed a strong diversity in operators' goal preference. A diversity between operators in completion orders was also found: As many as five different completion strategies were identified in the moderately-to-severely disrupted scenario. It is notable that in this study, the level of diversity in strategic mental models could not be related to worse or better performance.

The valuation of the controllers' goals was not reflected in their performance. The results show that in a moderate-to-severe traffic condition, controllers who highly value arrival punctuality showed more departure delay. Controllers who focused on departure punctuality had less arrival delay, and those who focused on a high level of platform consistency had less departure delay. Although these results do not confirm the expectations, they are in line with the fact that individual goals do not always lead to the system performance that corresponds to personal goals. In fact, the presence of multiple and competing goals can be seen as characteristics of complex, ill-structured environments, as they have to be weighed and prioritized, and compromises have to be made (Amelung & Funke, 2013; Funke, 1991; Hong, 1998). To obtain resilience, performance requires certain goals to take precedence over other goals (Woods, 2006). The moderate- to-severe traffic condition was a more complex situation, and the controllers possibly had to make more compromises. These goals were probably not as conflicting in the less complex situation because the scenario did not cause a conflict between arrival and departure goals.

This study also revealed a difference in the valuation of the goals "maintaining free track order" and "unplanned stops of trains before signals" between the more and the less experienced operators: The former considered these goals to be more important. As such, more experienced operators appear to be more comfortable about satisfying lower-prioritized organization goals. A trend was found for the level of experience affecting performance: Less experienced controllers showed better arrival punctuality than experienced controllers when no complex disruptions were introduced (Scenario 1). In contrast, the opposite trend was found when more train delays were introduced (Scenario 2). The results of Scenario 2 are in agreement with previous studies, following the line that more experienced controllers perform better in complex situations because of their well-developed mental models (Bogard et al., 2013).

Some limitations to this study should be mentioned. Although the simulator problems were controlled for, they nonetheless necessitated a small sample size. Also, given the length of each scenario, the number of conflict points per workstation was rather small. A limitation of this study in terms of goals trade-off consequences is that the level of violations was not assessed; we did not assess when a certain goal was violated during the simulator study or what the implications were of prioritizing one goal over another in these scenarios. These points should be taken into consideration in future studies.

Further, in line with the measured individual markers of resilience in this study, authors of future research could investigate the diversity of strategic mental models in the actual work environment. The level of diversity of completion strategies between workstations could also be further investigated.

In sum, the primary organizational goal was not reflected at the operational level. An explanation for this finding might be the difference in the realization of operator's goal versus the evaluation of operator's performance. In an exemplary case, train traffic controllers may recognize arrival punctuality as both a primary organizational goal and an individual goal; however, due to external factors influencing the train traffic flow, a high arrival punctuality cannot be guaranteed by the operator alone. As such, operators may develop different preferences and coping mechanisms to better reflect their performance.

Although variability in cognition and behavior is both healthy and allowed, it can be argued that the revealed goals and strategic mental models of operators are too diverse and therefore are unpredictable and most probably weaken the resilience at the system level. These results could be used as an indicator of brittle points that prevent the creation of a resilient organization (Gomes et al., 2009). It is observable that there are gaps between the work that is expected and the work that is done. Especially with the upcoming and planned large-scale changes in the railway system, it could be undesirable to continue with the redesign without involving the operational layer. Participatory design could be used as a joint approach to shape these changes (Falzon, 2008), enabling a new generation to work in a restructured work environment and to resonate these changes throughout all the levels.

References

Algemene Rekenkamer. (2012). *Aanbesteden door NS-Railinfrabeheer* [Procurement by NS-Railinfrabeheer]. Retrieved from http://www.rekenkamer.nl/Publicaties/Onderzoeksrapporten/Introducties/2001/06/Aanbestedingen_door_NS_Railinfrabeheer_1995_2000

Amelung, D., & Funke, J. (2013). Dealing with the uncertainties of climate engineering: Warnings from a psychological complex problem solving perspective. *Technology in Society, 35*, 32–40.

Back, J., Furniss, D., Hildebrandt, M., & Blandford, A. (2008). Resilience markers for safer systems and organisations. In M. Harrison & M.-A. Sujan (Eds.), *Computer safety, reliability, and security* (Vol. 5219, pp. 99–112). Berlin, Germany: Springer.

Bary, E. (2015). *Procesleiding op het Spoor* [Traffic management in the railways]. Retrieved from http://www.nicospilt.com/ index_procesleiding.htm

Bogard, T., Liu, M., & Chiang, Y. H. V. (2013). Thresholds of knowledge development in complex problem solving: A multiple-case study of advanced learners' cognitive processes. *Educational Technology Research and Development, 61*, 465–503.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience and school.* Washington, DC: National Academy Press.

Burnard, K., & Bhamra, R. (2011). Organisational resilience: Development of a conceptual framework for organisational responses. *International Journal of Production Research*, 49, 5581–5599.

Chialastri, A., & Pozzi, S. (2008). Resilience in the aviation system. In M. Harrison & M.-A. Sujan (Eds.), *Computer safety, reliability, and security* (Vol. 5219, pp. 86–98). Berlin, Germany: Springer.

Crawford, E., Toft, Y., & Kift, R. (2014). Attending to technology adoption in railway control rooms to increase functional resilience. In D. Harris (Ed.), *Engineering psychology and cognitive ergonomics* (Vol. 8532, pp. 447–457). Berlin, Germany: Springer.

Dekker, S. (2006). Resilience engineering: Chronicling the emergence of confused consensus. In E. Hollnagel, D. D. Woods, & N. Leveson (Eds.), *Resilience engineering: Concepts and precepts* (pp. 77–92). Aldershot, UK: Ashgate.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*, 32–64.

Falzon, P. (2008). Enabling safety: Issues in design and continuous design. *Cognition, Technology & Work, 10*, 7–14.

Funke, J. (1991). Solving complex problems: Exploration and control of complex systems. In R. J. Sternberg & P. A. Frensch (Eds.), *Complex problem solving: Principles and mechanisms* (pp. 185–222). Hillsdale, NJ: Lawrence Erlbaum.

Glaser, R., & Chi, M. (1988). Overview. In M. Chi, R. Glaser, & M. Farr (Eds.), *The nature of expertise* (pp. xv–xxxvi). Hillsdale, NJ: Lawrence Erlbaum.

Golightly, D., Sandblad, B., Dadashi, N., Andersson, A. W., Tschirner, S., & Sharples, S. (2013). A socio-technical comparison of rail traffic control between GB and Sweden. In N. Dadashi, A. Scott, J. R. Wilson, & A. Mills (Eds.), *Rail human factors: Supporting reliability, safety and cost reduction* (pp. 367–376). London, UK: Taylor & Francis.

Gomes, J. O., Woods, D. D., Carvalho, P. V. R., Huber, G. J., & Borges, M. R. S. (2009). Resilience and brittleness in the offshore helicopter transportation system: The identification of constraints and sacrifice decisions in pilots' work. *Reliability Engineering & System Safety, 94*, 311–319.

Hale, A., & Heijer, T. (2006). Is resilience really necessary? The case of railways. In E. Hollnagel, D. D. Woods, & N. Leveson (Eds.), *Resilience engineering: Concepts and precepts* (pp. 125–147). Aldershot, UK: Ashgate.

Hollnagel, E. (2008). Preface: Resilience engineering in a nutshell. In E. Hollnagel, C. P. Nemeth, & S. Dekker (Eds.), *Resilience engineering perspectives: Remaining sensitive to the possibility of failure* (pp. xi–xiv). Aldershot, UK: Ashgate.

Hollnagel, E. (2014). Resilience engineering and the built environment. *Building Research & Information, 42*, 221–228.

Hollnagel, E., & Woods, D. D. (2005). *Joint cognitive systems: Foundations of cognitive systems engineering*. Boca Raton, FL: CRC Press.

Hong, N. S. (1998). *The relationship between well-structured and ill-structured problem solving in multimedia simulation* (Doctoral dissertation). Retrieved from http://www.cet.edu/pdf/structure.pdf

Klimoski, R., & Mohammed, S. (1994). Team mental model: Construct or metaphor? *Journal of Management, 20*, 403–437.

Knieps, G. (2013). Competition and the railroads: A European perspective. *Journal of Competition Law and Economics, 9*, 153–169.

Latham, G. P., & Locke, E. A. (1991). Self-regulation through goal setting. *Organizational Behavior and Human Decision Processes, 50*, 212–247.

Lengnick-Hall, C. A., & Beck, T. E. (2005). Adaptive fit versus robust transformation: How organizations respond to environmental change. *Journal of Management, 31*, 738–757.

Lengnick-Hall, C. A., Beck, T. E., & Lengnick-Hall, M. L. (2011). Developing a capacity for organizational resilience through strategic human resource management. *Human Resource Management Review, 21*, 243–255.

Lo, J. C., Sehic, E., & Meijer, S. A. (2014). Explicit or implicit situation awareness? Situation awareness measurements of train traffic controllers in a monitoring mode. In D. Harris (Ed.), *Engineering psychology and cognitive ergonomics* (pp. 511–521). Cham, Switzerland: Springer.

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., & Cannon-Bowers, J. A. (2000). *The influence of shared mental models on team process and performance. Journal of Applied Psychology, 85*, 273–283.

Meijer, S. A. (2012). Introducing gaming simulation in the Dutch railways. *Procedia: Social and Behavioral Sciences, 48*, 41–51.

Meijer, S. A., Van der Kracht, P., Van Luipen, J. J. W., & Schaafsma, A. A. M. (2009). Studying a control concept for high-frequency train transport. *Infrastructure Systems and Services: Developing 21st Century Infrastructure Networks* (pp. 1–6). Piscataway, NJ: IEEE.

Mohammed, S., Ferzandi, L., & Hamilton, K. (2010). Metaphor no more: A 15-year review of the team mental model construct. *Journal of Management, 36*, 876–910.

Mohammed, S., Klimoski, R., & Rentsch, J. R. (2000). The measurement of team mental models: We have no shared schema. *Organizational Research Methods, 3*, 123–165.

Nederlandse Spoorwegen. (2015). *Jaarverslag 2014* [Annual report 2014]. Retrieved from https://www.nsjaarverslag.nl/FbContent.ashx/pub_1001/downloads/v170225111117/annual-report-2014.pdf

Nemeth, C., Wears, R., Patel, S., Rosen, G., & Cook, R. (2011). Resilience is not control: Healthcare, crisis management, and ICT. *Cognition, Technology & Work, 13*, 189–202.

Popova, V., & Sharpanskykh, A. (2011). Formal modelling of organizational goals based on performance indicators. *Data & Knowledge Engineering, 70*, 335–364.

ProRail. (2013). *Laatste sporen aangesloten op verkeersleiding ProRail* [Final tracks connected to ProRail control center]. Retrieved from https://www.prorail.nl/persberichten/laatste-sporen-aangesloten-op-verkeersleiding-prorail

ProRail. (2015a). *Jaarverslag 2014* [Annual report 2014]. Retrieved from https://www.prorail.nl/sites/default/files/jaarverslag_2014.pdf

ProRail. (2015b). Programma Hoogfrequent Spoorvervoer [High-frequency railway transport program]. Retrieved from https://www.prorail.nl/programma-hoogfrequent-spoorvervoer

Ramaekers, P, De Wit, T., & Pouwels, M. (2009). *Hoe druk is het nu werkelijk op het Nederlandse spoor? Het Nederlandse spoorgebruik in vergelijking met de rest van de EU-27* [How busy is it really on the Dutch railway tracks?]. The Hague, Netherlands: Centraal Bureau voor de Statistiek.

Raser, J. C. (1969). *Simulations and society: An exploration of scientific gaming*. Boston, MA: Allyn & Bacon.

Righi, A. W., Saurin, T. A., & Wachs, P. (2015). A systematic literature review of resilience engineering: Research areas and a research agenda proposal. *Reliability Engineering & System Safety, 141*, 142–152.

Rijksoverheid. (2015). *Mansveld: Boetes voor NS en ProRail* [Mansveld: Penalties for NS and ProRail]. Retrieved from http://www.rijksoverheid.nl/nieuws/2015/02/12/mansveld-boetes-voor-ns-en-prorail.html

Roth, E. M., & Patterson, E. S. (2005). Using observational study as a tool for discovery: Uncovering cognitive and collaborative demands and adaptive strategies. In H. Montgomery, R. Lipshitz, &

B. Brehmer (Eds.), *How professionals make decisions* (pp. 379–393). Mahwah, NJ: Lawrence Erlbaum.

Salas, E., Stout, R. J., & Cannon-Bowers, J. A. (1994). The role of shared mental models in developing shared situational awareness. In R. D. Gilson, D. J. Garland, & J. M. Koonce (Eds.), *Situational awareness in complex systems* (pp. 297–304). Daytona Beach, FL: Embry-Riddle Aeronautical University Press.

Schaffernicht, M., & Groesser, S. N. (2011). A comprehensive method for comparing mental models of dynamic systems. *European Journal of Operational Research, 210*, 57–67.

Sheridan, T. B. (2002). *Humans and automation: System design and research issues*. Santa Monica, CA: Wiley.

Steenhuisen, B. (2009). *Competing Public Values: Coping Strategies in Heavily Regulated Utility Industries*. Enschede: Gildeprint Drukkerijen.

Steenhuisen, B., & De Bruijne, M. (2009). The brittleness of unbundled train systems: Crumbling operational coping strategies. *Paper presented at the Second International Symposium on Engineering Systems*, Cambridge, MA.

Sulmann, H. (2000). *Railverkeersleiding: Van sein tot sein* [Railway traffic control: From signal to signal]. Unpublished manuscript.

Veeneman, W. W. (2006). Punctualiteit, wat heet [Punctuality: What does it mean?]. *Verkeerskunde: Vaktijdschrift over Verkeer en Vervoer, 57*(5), 58–59.

Veeneman, W. W., & Van de Velde, D. M. (2006). The value of bus and train: Public values in public transport. In H. Maier (Ed.), *Proceedings of the European Transport Conference* (pp. 1–12). London, UK: Association for European Transport.

Webber, S. S., Chen, G., Payne, S. C., Marsh, S. M., & Zaccaro, S. J. (2000). Enhancing team mental model measurement with performance appraisal practices. *Organizational Research Methods, 3*, 307–322.

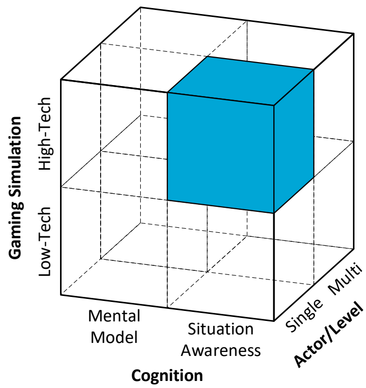
Wilson, J. R. (2000). The place and value of mental models. *Proceedings of the Human Factors and Ergonomics Society 44th Annual Meeting* (pp. 49–52). Santa Monica, CA: Human Factors and Ergonomics Society.

Wilson, J. R., Farrington-Darby, T., Cox, G., Bye, R., & Hockey, G. R. J. (2007). The railway as a socio-technical system: Human factors at the heart of successful rail engineering. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 221*, 101–115.

Woods, D. D. (2006). Essential characteristics of resilience. In E. Hollnagel, D. D. Woods, & N. Leveson (Eds.), *Resilience engineering: Concepts and precepts* (pp. 205–233). Aldershot, UK: Ashgate.

5 Explicit or implicit situation awareness? Measuring the situation awareness of train traffic controllers

Apart from a minor update, this chapter has been published as:
Lo, J. C., Sehic, E., Brookhuis, K. A., & Meijer, S. A. (2016). Explicit or implicit situation awareness? Measuring the situation awareness of train traffic controllers. *Transportation Research Part F: Traffic Psychology and Behaviour, 43*, 325-338.



Abstract
This study uses a single-actor simulator to investigate situation awareness levels of train traffic controllers. Two studies have been conducted, in which implicit, intuitive, unconsious levels of situation awareness are measured. The focus on implicit levels of situation awareness, builds further on previous research that have mainly focused on explicit, reasoned and conscious levels of situation awareness. A novel approach was used to uncover levels of implicit SA through a set of three analyses: 1) fairly low SAGAT values with correlations between SAGAT scores and multiple performance indicators; 2) negative correlations between work experience and SAGAT scores; and 3) structurally lower level-1 SA (perception) scores in comparison to level-2 SA (comprehension) scores in accordance with Endsley's three-level model. Two studies were conducted: A pilot study – which focused on SA measurements with TTCs in a monitoring mode (N=9) – and the main study, which involved TTCs from another control center (N = 20) and three different disrupted conditions. In the pilot study, SA was measured through the situation-awareness global assessment technique (SAGAT), perceived SA, and observed SA, and performance was measured through punctuality and unplanned stops of trains before red signals. In the main study, SA was measured through SAGAT, and perceived SA and multiple performance indicators, such as arrival and departure punctuality and platform consistency, were assessed. In both studies, the set of three analyses showed consistent and persistent indications of the presence of implicit SA. Endsley's three-level model and related SAGAT method can be constrained by the presence of these intuitive, unconscious processes and inconsistent findings on correlations between SAGAT scores and performance. These findings provide insights into the SA of TTCs in the Netherlands and can support the development of training programs and/or the design of a new traffic management system.

Keywords: Tacit knowledge; implicit situation awareness; explicit situation awareness

5.1 Introduction

The cognitive concept of situation awareness (SA) has been widely investigated by the human factors community in the past two decades and across different domains (Endsley, 2015; Sneddon, Mearns, & Flin, 2006). SA can be ascribed to practitioners in complex, dynamic systems that have perceptual and cognitive demanding tasks that are pressured by safe, effective and timely decisions (Endsley, 1995a). The notion of SA is in line with the limits of bounded rationality and bounded awareness, in which individuals are cognitively restrained by, for example, their dependency on sensory (perceptual) input, "computational powers," and situational circumstances (Chugh & Bazerman, 2007; Lipshitz, Klein, Orasanu, & Salas, 2001; Simon, 1983). Despite numerous discussions on situation awareness definitions (e.g. process versus product) and frameworks (e.g. SA residing in the mind versus the system), Endsley's three-level model of SA has received broad support in the human factors community (e.g. Dekker, Hummerdal, & Smith, 2010; Parasuraman, Sheridan, & Wickens, 2008; Sarter & Woods, 1991; Stanton et al., 2006). It is defined as (1) the perception of elements in the environment, (2) the comprehension of these elements, and (3) the projection of these elements in the near future (Endsley, 1988a). The development of SA is a process, which is reflected throughout the three levels and which can also be referred to as situation assessment (Endsley, 1995a). Situation awareness itself is the product from this process. Individual factors such as goals, objectives and expectations influence the situation assessment. Additionally, also task or system factors, such as interface design, stress and workload and automation impact the process of situation awareness. The model draws from traditional information-processing theories, in which a well-developed understanding of the system's dynamics (also known as mental model) is necessary to develop a good situation awareness (Endsley, 2001). Another characteristic of the model is that situation awareness is formulated as an indicator of decision-making, which in turn can predict the level of performance of actions.

The operationalization of the three-level model has so far mainly focused on explicating knowledge (e.g., Salmon, Stanton, Walker, & Green, 2006). For instance, the situation-awareness global assessment technique (SAGAT), which focuses on extracting operators' explicit knowledge through probes during simulator freezes, shows a correlation with performance and has received general acceptance in the human factors community (Salmon et al., 2009). Through SAGAT, a 'snapshot' of the operator's mental model of the situation is captured, a direct measurement of the pilot's knowledge of the situation is obtained and objectively collected (Endsley, 1988b), However, this focus on solely explicating knowledge can be seen as conflicting in accordance to the naturalistic decision-making field. In this research area, an emphasis has been put on investigating operators in their daily work environment, in which this line of research indicates that operators might use their intuition to conduct pattern matching in certain situations (Klein, 2008). Operators may use unconscious processes in order to take rapid decisions. As such, focusing on measuring explicit levels of situation awareness may not be a good reflection of operator's actual cognitive processes.

5.2 Explicit versus implicit situation awareness

A previous literature review on explicit and implicit situation awareness has been conducted by the current authors (Lo, Sehic, & Meijer, 2014). In this review SA has been found, in line with Adams, Tenney, and Pew (1995), as a dynamic mental model of the situation, in which explicit and implicit levels of knowledge can be distinguished. The active knowledge that resides in the working memory can be related to explicit knowledge, while the less active knowledge which cannot be inferred from queries or knowledge probes can be related to implicit knowledge (Croft, Banbury, Butler, & Berry, 2004; Endsley, 1997; Gugerty, 1997). Furthermore, implicit knowledge is considered as unintentional, unconscious, and intuitive. In accordance to Croft et al., (2004) and Durso and Sethumadhavan (2008), implicit SA can also be viewed as implicit processes in SA. In situations of competing attentional demands, these implicit processes are characterized as extremely durable and more robust, and related to an increase in expertise. The relation between expertise and implicit processes is also considered an aspect of the skill, rule, knowledge framework of Rasmussen (1983), which relates little conscious attention or control to the skill-based level, on the contrary to the knowledge-based level.

Previous examples of the operationalization of implicit SA have been through comparisons of recalling probes (such as the SAGAT method) with performance-based or speed/accuracy measurements, such as hostile or friendly aircraft recognition (Croft et al., 2004; Endsley, 2000a; Gugerty, 1997).

In a more general psychological context, these unconscious, automatic cognitive processes are also referred to as "system 1 versus system 2," which operate on a conscious level but more slowly (Kahneman, 2012). Although the role of unconscious processes in terms of both neuropsychological and cognitive mechanisms is recognized within fundamental streams in psychology, researchers are yet to develop a deeper understanding and controversial findings are impeding their progress (e.g., Dijksterhuis, Bos, Nordgren, & van Baaren, 2006; Newell & Shanks, 2014; Reber, 1989).

5.3 Train traffic control

Following the widespread breakup of the railway sector across Europe in the 1990s into multiple commercialized and governmental organizations, there has been a steady increase in research on rail human factors (Knieps, 2013; Van de Velde, 2001; Wilson & Norris, 2005). The de-bundling of the railway sector led to a rather rapidly changing domain in terms of technical requirements, namely the implementation of higher levels of automation, such as automatic route setting and traffic management systems (e.g., Ferreira & Balfe, 2014; Sharples, Millen, Golightly, & Balfe, 2011). There is, however, a notable difference across countries in the level of automation that has actually been implemented (Golightly et al., 2013). In addition, the organization and management of the railway system itself has also undergone a major transformation, with changes leading to, for example, frequent adaptions of operating procedures, organizational goals, and train traffic capacity that could be in conflict with one another, impacting on operator’s cognitive strategies (Lo, Pluyter, & Meijer, 2016; Steenhuisen, 2009).

Different cognitive strategies have been identified across domains, such as in aviation and the military. Findings from a study showed that air traffic controllers devoted 90% of their time to processing information (Kaempf, Klein, Thordsen, & Wolf, 1996), another that 95% of the tactical commanders in the military domain used a recognition decision strategy (Kaempf & Orsanu, 1997). Although no quantitative numbers are known, train traffic controllers (TTCs) are expected to spend a significant amount of time where they monitor the train traffic flow. TTCs in the Netherlands have a traffic management system that automatically sets planned routes. They operate in a fashion that requires active involvement to adapt the traffic management system when ad-hoc train routes are planned or delays are affecting the train traffic flow. Findings from an ethnographical study on the decision-making of railway traffic and network operators (i.e., train traffic and regional network controllers) in the Netherlands revealed that operators do not explore different consequences in-depth and have difficulty making their reasoning explicit, thus showing indications of tacit, implicit knowledge (Steenhuisen, 2009). It has been inferred that the implicit knowledge that is held by experienced railway operators is typical of railway culture and may also have been facilitated by the de-bundling of state-owned railway organizations (Steenhuisen, 2009, 2014; Wilson & Norris, 2005).

The aim of the present research was to contribute to the body of knowledge on SA and rail human factors research. In essence, this research focused on uncovering implicit SA levels of TTCs by presenting a novel approach on the identification of implicit situation awareness, and applying this approach in two studies at different railway traffic control centers. The following section describes our methodology for the identification of levels of implicit SA. The subsequent sections focus on SA measurements conducted in an individual human-in-the-loop simulation environment with a lightly disrupted train traffic flow. Given the low sample size of the first study and the experience of a less disrupted condition than was designed for, this study was coined as a pilot study. The second (main) study replicated the first study in terms of its research question and investigated levels of implicit situation awareness in both a light and medium disrupted condition at a different control center. The implications of both studies are described in the general discussion and conclusion section.

5.4 A novel approach for the identification of implicit situation awareness

It is not easy to identify implicit levels of SA in a maturing human factors field such as the railway sector, as in-depth domain knowledge is required to create and maintain control in scenarios designed to obtain controlled measurements of implicit SA. Based on the discussed literature in the previous section, we therefore used a novel set of analyses that could indicate the presence of implicit SA. An elaborate description of the operationalization of the related variables – that is, of SAGAT and the performance of TTCs (e.g., punctuality) – is provided in the method section of each study.

Firstly, the widely used SAGAT method was applied to measure SA, as it measures levels of explicit SA. These SA scores are represented in percentages of correctly answered SA probes, in which a high percentage reflects a high level of explicit SA. We posited that low absolute SAGAT values (i.e., poor explicit SA) indicate an absence of explicit SA while implicit SA may still be present. In order to test the latter, the presence of a positive correlation between the low SAGAT scores and performance could serve as a confirmation for implicit SA.

Secondly, based on previous studies that found an increase in implicit, tacit knowledge in more experienced operators, we posited that a negative correlation between SAGAT scores and work experience across TTCs could serve as a second indicator of the presence of implicit SA.

Thirdly, we posited that the presence of implicit SA would be observable in all three levels of SA as per Endsley's three-level model. According to this model, as illustrated by the described accident assessment scores for each SA level, level-1 SA (perception) scores should normally be higher in absolute values in comparison to the other two levels, as each subsequent level builds upon the current input. As such, equal or lower level-2 SA (understanding) scores should be expected, as should equal or lower level-3 SA (projection) scores. A deviation of this trend could indicate levels of implicit SA. This type of analysis can be operationalized by ascribing a probe to a certain SA level and thereby calculating the SA scores for each level of SA. A qualitative assessment for deviations in the scores across the three SA levels can be made to evaluate the extent a similar trend is followed, as is theoretically expected.

5.5 Pilot study

The pilot study focused on measuring SA when TTCs were faced with minor train traffic delays. However, the TTCs perceived the delays as rather undisruptive, and therefore took on a monitoring role during the scenarios. This study has also been described in Lo et al., (2014).

5.5.1 Method

5.5.1.1 Experimental setting

With the opportunity to conduct human-in-the-loop studies at railway traffic operators, organizational questions are often accompanied, influencing the design of the study. In the present study, the overall purpose was to investigate the impact of a human-in-the-loop simulation session on the quality control processes of a new train timetable. Table 5.1 describes the characteristics of the simulator; see Lo et al., (2014) for a more elaborate description.

Table 5.1: Characteristics of the simulator design in the pilot study (Lo et al., 2014).

Characteristic	**Description**
Purpose	Studying the impact of a simulator session on the quality control processes of a new train timetable
Scenarios	Two for each participant: 1) 2013 train timetable, 2) 2014 train timetable
Simulated world	Detailed infrastructure; detailed timetable; limited options in number of actions; larger area of train traffic operations (merged workstations of Zwolle station east-side and Zwolle station west-side)
# of participants	1 per session
Roles	Train traffic controller
Type of role	Similar to their own roles
Objectives	Execution of tasks – similar to those in their daily work
Constraints	Exclusion of roles outside the defined infrastructure area, exclusion of train driver, no major disruptions
Load	Average train delays in both rounds
Situation (external influencing factors)	Presence of individual observers seated next or near the participant, facilitators
Time model	Continuous

It should be noted that during the simulator runs, it became clear that the scenario load was not perceived as invasive as initially designed. The operators were to experience a minor disruption; however, during the sessions they did not perceive the train delays as sufficiently problematic to make manual changes to the traffic management system. This was possibly due to the automatic route setting (automatische rijweginstelling (ARI) in Dutch), which on a few occasions during the sessions automatically deactivated when a delay reached a certain threshold. TTCs could therefore remain in a monitoring mode of working.

5.5.1.2 Participants

Eleven TTCs from the regional traffic control center in Zwolle participated in this study. Extra operators were scheduled in the day and evening shifts by the personnel planner so that operators were able to take over the train traffic control task from participants that were willing to participate in the study and who were licensed to operate the current workstation in the simulator.

5.5.1.3 Materials

Work experience, *perceived competences* and *motivation*. Before each session, a number of background questions were asked, namely operator's work experience in the railway sector, work experience in the current job function, perceived experience of the workspace, perceived competencies in comparison to peers, and motivation to participate in the session. The first two items were open-ended questions, whereas the latter three were measured on a 5-point Likert scale, ranging from "strongly unexperienced" or "fully disagree" to "strongly experienced" or "fully agree."

Situation awareness probes. Three types of SA measurement techniques were selected to triangulate measurements of SA. As validation of the situation awareness global assessment technique (SAGAT) has received major attention (Salmon et al., 2006), it was selected as a query technique. Probes were developed based on a concept of a goal-directed task analysis (GDTA) for TTCs (Endsley, Bolté, & Jones, 2003) and developed in collaboration with a subject matter expert. Examples of the SAGAT questions are shown in Table 5.2. Probes were presented in a multiple-choice answering format in line with Strater, Endsley, Pleban, and Matthews (2001).

Table 5.2: Examples of the SAGAT queries (Lo et al., 2014).

SA level	Probe/query example	Answering category example
1	At which track does train 13828 arrive at Zwolle station?	Track 14, track 15, track 16
2	Which train leaves Zwolle station first according to planning?	[Train number] 12522, 3629, 9119
3	How is the track capacity at 07:46 at Zwolle station?	6 tracks free, 5 tracks free, 4 tracks free, other, namely: ..

During the session, the participants received 22 queries that were presented during three freeze interruptions, each of seven or eight probes. The simulator freezes occurred immediately after possible conflicting choices in the train traffic flow. In the analysis, 19 SAGAT queries were used for each scenario, in which percentages of correct answers were calculated.

Perceived situation awareness self-ratings were measured through the Mission Awareness Rating Scale (MARS) (Matthews & Beal, 2002) at the end of each scenario. These three items were equivalent to the three SA levels as identified by Endsley (1988a) and were scored on a 4-point scale, ranging from "fully disagree" to "fully agree."

Observed situation awareness was measured based on items identical to perceived SA self-ratings following the MARS questions. Scores were administered by a subject matter expert, who was present during all sessions. To support the evaluation, an observation sheet used during training session was provided as a guideline.

Performance is assessed in the railway sector at a system level through performance indicators such as punctuality. The individual performance of a TTC is currently not assessed due to the complexity of external influencing factors. In consultation with the performance and analytics department, the performance indicators "punctuality" and "unplanned stops of trains before signals" were identified. Here, "punctuality" is defined as the entry and exit times of trains for a workstation that is responsible for a specific allocated area. The "unplanned stops of trains before signals" performance indicator relates to unexpected changes of signals, which might lead to train drivers encountering an unplanned red signal. This specific performance indicator may be linked to safety issues, as it might trigger a possibility for a train to pass a red signal. Simulator log files were used to retrieve the performance data. Punctuality is calculated based on the three-minute threshold the Dutch railway infrastructure organization has determined for defining delay. As such, the punctuality of trains is measured in terms of percentages and unplanned stops of trains before signals in terms of absolute values, within the given scenario.

Simulator validity was measured in order to obtain an indication of the validity of the human-in-the-loop simulator given the task at hand. Three components of simulator validity in line with Raser (1969) were identified: 1) structural validity (the degree of similarity in structure, such as of physical objects in the simulated and the reference system); 2) processes validity (the degree of similarity in processes, such as communication between the simulated and reference system); 3) and psychological reality (the degree to which the participants perceive the simulated system as realistic). Structural validity (Cronbach's alpha, α = .43) was measured through three items, for example "I can apply the information from the information sources in the simulator in a similar way as in the real world." Similarly, three items measured process validity (α = .75), for instance, "The train traffic flow in the simulator is similar in its processes to the real-world train traffic flow." Seven items were used to measure the third component, psychological reality (α = .73), for example, "The train model is sufficiently realistic for the current task." A 5-point Likert scale, ranging from "*fully disagree*" to "*fully agree*," was used to measure the items.

Mental workload. Five workload items based on the NASA-TLX (Hart & Staveland, 1988) were administered after each scenario (α = .81). The NASA-TLX item on physical demand was not included, as physical demands are not applicable in the work of a TTC. A 5-point Likert scale was applied in line with previous items.

Learning in terms of *mental model development* with regard to the new timetable, i.e. to what extent did the participant learn, was checked with the items "the new timetable is more challenging than the current timetable" and "I quickly got used to the new timetable." All these items were measured on a 5-point Likert scale.

The *learning effects* of the SAGAT queries were assessed based on comparisons of the SA probe scores for each simulator freeze.

5.5.1.4 Procedure

The session started with a general introduction on the purpose of the simulator session and a description of the simulator's functionalities (see Figure 5.1). Participants then completed a pre-questionnaire and gave their permission to make video recordings during the simulator session. Subsequently, participants conducted two 35-min scenarios with two freezes for SA probes and a final SA probe at the end of the round. During the scenarios, the facilitators asked about the usability of the simulator, and the TTCs' way of working when they were not occupied with their task, in order to simulate conversations as in normal work conditions. At the end of the second round, participants completed a post-questionnaire.

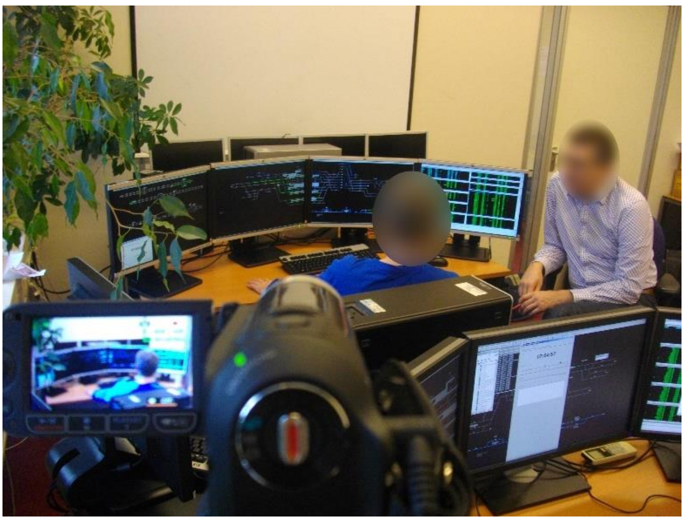


Figure 5.1: Simulator setup with the simulator set with the new timetable (foreground) and the set with the current timetable (background)

5.5.2 Results

In total, 11 TTCs (10 male and 1 female) took part in the sessions. Two male TTCs were omitted from the analyses due to non-responses or deviation from the probes' instructions. The average work experience in the current job function was 15.0 years (*SD* = 8.08). The level of average work experience overall in the railway sector was slightly higher, *M* = 20.4, *SD* = 9.94. Participants perceived the level of their competency in their current workspace as high (*M* = 4.2, *SD* = .67). A high level of interest was expressed in participating in the simulator session (*M* = 4.4, *SD* = 1.01).

5.5.2.1 Mental model development

Participants indicated that they had quickly got used to the new timetable (*M* = 4.4, *SD* = .73), indicating that operators could rapidly absorb the characteristics of the new timetable. They also indicated that the new train timetable was not more challenging than the current train timetable (*M* = 1.7, *SD* = .71). Qualitative data obtained during the session supports both results. As such, it can be inferred that operators had a sufficiently developed mental model of the changed railway system, which indicated that the measured SA values may be less affected by the new situation.

5.5.2.2 Learning effects

SAGAT probes drawn from the three measurement moments were compared for significant differences using a Friedman test. No significant difference were found, indicating that participants' scores did not significantly deviate, thus ruling out learning effects on the SA probes.

5.5.3 Simulator validity

The findings show that participants have a rather positive to a positive perception of the simulator (see Table 5.3), which supports the perceived validity of the simulator for its current purpose. Notable is the slightly higher score for experienced validity of the simulator (i.e. psychological reality). The quantitative findings can also be supported by the qualitative data, in which participants indicated to be able to carry out their task as a train traffic controller in the presented scenarios. In a more severe disrupted condition, this might have been otherwise as the limited functionalities of the simulator may set constraints. This is also slightly reflected in the current structural validity scores. As such, findings in the simulated environment with regard to cognitive and behavioral indicators, such as SA and performance, can be generalized to a regular work environment with regards to this task.

Table 5.3: Validity dimensions of the simulator for the current task.

	N	M	SD
Structural validity	9	3.7	.45
Process validity	9	3.9	.47
Psychological reality	9	3.9	.40

5.5.3.1 Situation awareness and performance

Table 5.4 presents the measurements of SA, performance, and mental workload. A comparison between the scores in scenario 1 and scenario 2 using the Wilcoxon test showed a significant difference for the observed SA scores (Z = -2.33, p = .02). The level of SA rated by the observer was higher in scenario 2 in comparison to scenario 1.

Table 5.4: Measurements of situation awareness, performance and mental workload.

	Scenario 1				Scenario 2			
	N	Mean Rank	M	SD	N	Mean Rank	M	SD
SAGAT (%)	9	4	44.4	17.68	9	4	37.11	11.07
Perceived SA (1-4)[1]	6	0	3.1*	.53	9	1	3.3*	.43
Observed SA (1-4)[1]	9	0[a]	3.4	.41	8	3.5[a]	3.8*	.35
Punctuality (%)	9	3.3	99.4*	1.12	9	2.0	98.6*	1.65
Unplanned stops	9	4.5	2.0*	.87	9	4.5	2.2	1.30
Mental workload (1-5)	N/A	N/A	N/A	N/A	8	N/A	1.7	.52

* Significant on the Kolmogorov–Smirnov normality test $p < .05$.
[a] $p \leq .05$.
[1] based on a Likert scale ranging from 1 to 4.

In terms of absolute SAGAT scores, the mean values are rather low (44% in scenario 1 and 37% in scenario 2). However, perceived and observed SA scores show high to very high levels of SA. Furthermore, in terms of the performance indicator punctuality, the findings indicate near optimal performance achievements, which can be explained by the introduced low impact delays.

5.5.3.2 Implicit situation awareness

To investigate the extent of implicit or explicit SA, we applied the novel set of analyses presented in section two. Firstly, the relation between SA and the two performance indicator types was assessed using the effect size scales in correlation analysis by Cohen (1988). A large effect size was found for the relation between the SAGAT scores and punctuality in scenario 1 (ρ = .64, p = .06). Thus, a higher SA probe score (i.e., explicit SA) leads to a higher level of train punctuality. A similar but moderate effect size was found in the correlation between the SAGAT scores and punctuality in scenario 2, ρ = .32, p = .42. Furthermore, moderate effect sizes were found between perceived SA and punctuality in scenario 1 and between perceived SA and unplanned stops in scenario 2, respectively (ρ = .37, p = .33; ρ = .42, p = .27). A higher level of perceived SA is related to a higher level of punctuality, and a higher level of perceived SA is related to fewer unplanned stops of trains. Unexpected correlations were found between observed SA and unplanned stops in scenarios 1 and 2, respectively ρ = .68, p = .04; ρ = .66, p = .08. These findings indicate that a higher level of observed SA is related to more unplanned stops. An unexpected correlation was also found between observed SA and punctuality in scenario 2; ρ = .57, p = .14, in which a higher level of observed SA can be related to a worse performance in punctuality.

Apart from the unexpected negative relation between observed SA scores and performance, there is a tendency for both perceived SA and SAGAT scores to show a positive relation with performance, in line with expectations. However, absolute values of SAGAT probes seem to be rather low in the current monitoring mode. Given the relation between SA probes and performance, the lower SAGAT values might be explained by the presence of implicit SA.

Secondly, implicit SA would become more apparent when work experience increases. In line with this implication, a trend was found in scenario 1 for a large negative correlation between the work experience in the railway domain and the percentage of correct SAGAT answers, ρ = .54, p = .14 and a moderate negative correlation in scenario 2, ρ = .46, p = .22; more experience in the railway domain can be related to a lower level of explicit SA. Moderate to large correlations were also found between work experience in their current role as a TTC and perceived situation awareness in scenario 1, respectively ρ = .60, p = .12 and ρ = .68, p = .06. Similar correlations were also found in scenario 2, respectively ρ = .36, p = .28 and ρ = .64, p = .03. More work experience in either the current function or the railway sector is related to a lower level of perceived SA. Between both types of work experience and perceived mental workload, only small effect sizes were found in both scenarios.

Finally, the implication of implicit SA in TTCs may be further supported by the findings in Table 5.5, which lists the SA probe scores calculated per SA level. The findings indicate that the level-1 SAGAT probe scores were fairly low in terms of absolute values (e.g., 37% in scenario 1) and lower scores in comparison to level-2 SAGAT scores (e.g., 65% in scenario 1). In accordance with the three-level model, SA probe scores would be highest at level 1 (perception of elements) and drop with each subsequent SA level. Thus, operators might not process all level-1 SA information explicitly, but instead rely on filtering mechanisms that enable them to understand and make predictions about future states of the traffic flow.

Table 5.5: SA probes per SA level (also reported in Lo et al., 2014).

	Scenario 1		**Scenario 2**	
	total # items (N=9)	**% correct**	**total # items (N=9)**	**% correct**
Level 1	99	37	108	39
Level 2	54	65	45	42
Level 3	18	39	18	17

5.5.4 Discussion

Given the low sample size, this study was considered as a pilot study. However, the findings provided initial indications of the presence of TTCs' implicit SA through the application of a novel set of three analyses. A first result that supports the notion of implicit SA is reflected by the low SAGAT scores. Low SAGAT scores can be ascribed to undeveloped SA, for instance if there is not yet a solid mental model base, as is the case with novices. Another explanation is the presence of implicit SA; that is, operators are not consciously aware of environmental cues. Given the moderate positive correlations between SAGAT scores, there is support for implicit SA, as a higher SAGAT score can be related to better performance. Although effect sizes could be stronger, consistent findings were found for multiple performance indicators.

A second result that supports the notion of implicit SA is the negative correlation between both work experience in the current function and railway sector and SAGAT scores. The findings are in line with the phenomenon that expert operators exhibit more tacit knowledge, which results in implicit SA. Thirdly, support for implicit SA is also provided by the finding of lower level-1 SAGAT scores in comparison to the level-2 SAGAT scores. The lower level-1 SAGAT score might be a result of the fact that although operators scan for cues in the train traffic flow, they do not actively process the perceived information.

The pilot study had a few limitations: a large number of the correlations did not reach significance, probably because of the small sample size. As such, the generalizability of these findings is rather limited. Also, more accurately observed SA should be determined by involving more observers and establishing inter-rater reliability. Finally, further investigation is needed to exclude the possibility that the low SAGAT scores in terms of absolute values can be ascribed to the (passive) monitoring mode of operations that TTCs exhibited during the simulator sessions (Endsley & Rodgers, 1997).

5.6 Main study

The pilot study focused on measuring SA in TTCs in a passive, monitoring mode during a lightly disrupted train traffic condition. Building on and verifying the findings from the pilot study, the main study investigated levels of implicit SA of TTCs in both a lightly disrupted train traffic condition and a moderately disrupted train traffic condition at a different control center with different train traffic controllers.

5.6.1 Method

5.6.1.1 Experimental setting

We again used the human-in-the-loop simulator for TTCs, but this time with expanded functionalities with regard to manual changes in the traffic management system during more serious disruptions. Two TTCs at two workstations, sharing responsibility for the train traffic flow and infrastructure capacity at Utrecht Central Station, participated in each session. The workstations focused on either corridor-steered train traffic (the "through" workstation) or turning train traffic (the "turn" workstation). The scenarios were designed by two senior TTCs in such a way that the introduced delays did not affect both areas of responsibility. As such, collaboration between the TTCs was not necessary.

Regarding the load in the scenarios, the participants conducted three classes of scenarios, making three conditions: (1) known minor delays, (2) unknown minor delays, and (3) unknown moderate delays. Table 5.6 describes the design characteristics of the simulator session.

Table 5.6: Characteristics of the simulator design in the main study

Characteristic	**Description**
Purpose	Familiarization with the changed railway infrastructure area due to the removal of 66 switches around Utrecht Central Station
Scenarios	Two for each participant: First the "through" workstation and then the "turn" workstation, and vice versa for the other participant
Simulated world	Detailed infrastructure; detailed timetable; options focused on logistic implementations in the traffic management system
# of participants	2 per session
Roles	Train traffic controller at "turn" and "through" workstations
Type of role	Similar to their own roles
Objectives	Execution of tasks – similar to those in their daily work
Constraints	Exclusion of roles outside the defined infrastructure area, exclusion of train driver
Load	Minor delays in round 1, moderate delays in round 2
Situation (external influencing factors)	Presence of individual senior TTCs as trainers, facilitators
Time model	Continuous

5.6.1.2 Participants

Twenty-two TTCs from the regional control center in Utrecht participated in the simulator sessions. A personnel planner scheduled all active TTCs that were authorized to operate both workstations. Due to the availability of the twenty-two TTCs and the simulator, the measurements had to be spread over five days.

5.6.1.3 Materials

Similar variables were measured with identical scales as in the pilot study, with regard to *work experience*, *perceived competences and motivation*, *perceived situation awareness*, *performance*, and *mental workload* (α = .83).

Situation awareness probes were administered in the second simulator run with in total 19 items. The SA probes were developed by a subject matter expert and evaluated by two senior TTCs. A number of items were removed depending on simulator issues experienced by some participants and items that were retrospectively too difficult to identify as correct or incorrect. An example of an item that was identified as too difficult to identify was related to the situation in 10– 15 min. Occurrences of simulator issues were identified as invasive for SA acquirement when an interruption by one of the experimenters occurred two minutes prior to administration of SA probes. This two-minute limit is identified as the threshold for interruptions during SAGAT measurement (Endsley, 1995b; Kaber, Perry, Segall, McClernon, & Prinzel, 2006).

Performance. Punctuality was divided into departure punctuality and arrival punctuality; the latter is measured by the railway infrastructure organization as an official key performance indicator for the railway system. Both departure and arrival punctuality are calculated based on a three-minute threshold; delays shorter than three minutes are not included. Departure and arrival delay are used as performance indicators, operationalized by the amount of train delays in seconds. Platform consistency was determined as a performance indicator related to service level, and defined as the number of trains that deviated from the planned platform. For the analysis, these numbers were normalized and calculated in percentages. For an elaborate description of these performance indicators see Lo et al., (2016).

Simulator validity was measured similarly with three items for structural validity (Cronbach's alpha, α = .65 after the removal of one item) and process validity (α = .60). Psychological reality was measured with three items (α = .67).

Learning or *mental model development* was assessed for the speed of getting used to the workstations: "I was able to quickly get accustomed to the changes on the 'turn' workstation," "I was able to quickly get accustomed to the changes on the 'through' workstation."

Learning effects between scenarios were checked by the item "the second scenario was easier because of experiences with the first scenario." These items were measured on a 5-point Likert scale. Also, SAGAT scores of each simulator freeze were compared to identify possible learning effects in answering SA probes.

5.6.1.4 Procedure

The sessions were spread over five days as all authorized operators for the current workstations were required to be scheduled for a two-hour lasting simulator session. Two operators each participated in two rounds (see Figure 5.2). At the beginning of each session, facilitators familiarized the operators with the functionalities of the simulator, and senior TTCs held a brief training about the possibilities and limitations of the available infrastructure. These senior TTCs remained present during the session. Permission was obtained from the operators to make video recordings during the session. Questionnaires were handed out to participants before and after each round. During the second simulator run, two short pauses were introduced to obtain SA probes. In order for participants to reflect on the system's performance and to increase the motivation of operators in their participation, actual scores for that infrastructural area were displayed on their screens in terms of arrival and departure punctuality and platform consistency. Although the highest score obtained was kept anonymous, it was listed on a whiteboard. To increase the similarity between the simulator setting and the usual work setting, conversations were allowed, as TTCs are used to conversing while at work. The insights acquired during these conversations were used as qualitative data.

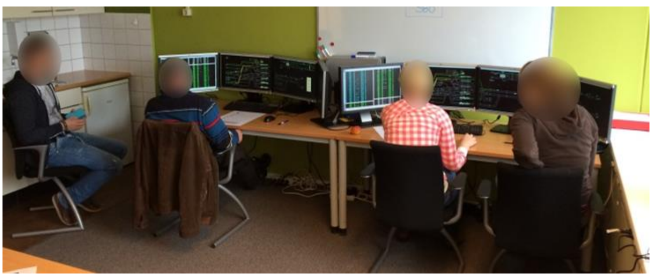


Figure 5.2. Simulator setup with the "turn" and "through" TTCs and two facilitators

5.6.2 Results

Twenty-two TTCs (18 male and 4 female) participated in the simulator sessions. Two participants were omitted from the analysis as they twice received the moderately disrupted scenarios. On average, participants had 10.3 years of work experience (*SD* = 9.24) as a TTC, 12.2 years of work experience in the railway sector (*SD* = 12.09), and 8.1 years of work experience with the current (Utrecht Central) workstations (*SD* = 8.33). Of the participants, 65% also occasionally functioned as planners during disruptions. The participants expressed a positive interest in participating in the simulator sessions (*M* = 3.9, *SD* = 1.04).

5.6.2.1 Mental model development

Operators indicated the speed at which they became familiar with the simulator as positive (*M* = 4.2, *SD* = .86). They also indicated that they quickly got used to the simulator at both the "turn" workstation (*M* = 3.8, *SD* = .60) and the "through" workstation (*M* = 3.9, *SD* = .57). This indicates that there was hardly any distinction with regard to differences in perceived difficulty between workstations, next to a sufficiently developed mental model of the changed railway system. As such, measured SA values may be less affected by the new situation.

5.6.2.2 Learning effects

On average, participants were neutral about the second round in the session being easier after experiencing the first scenario (*M* = 3.2, *SD* = .92). No significant difference was found between the three disruption conditions for this item, $\chi^2(2) = 2.654$, $p = .27$. As an illustration, when comparing conditions 1 and 3 as most differentiating groups, the result indicates that participants in condition 1 who were aware of the train delays in scenario 1, did not find the second round more easy than participants in condition 3, who received a more difficult scenario in the second round.

To check for learning effects occurring over the three SA probes pauses, a Friedman test was conducted to explore significant differences between the three SA scores. No significant differences were found between the three measurement points, indicating that there were no significant deviations between the SA scores and therefore no significant learning effects.

5.6.2.3 Simulator validity

Participants were neutral to slightly positive about the simulator functionalities in terms of the structural and process validity of the simulator (see Table 5.7). In terms of their experience of the simulator in comparison to their regular work environment, participants were slightly more positive. Qualitative data indicated that participants were content with the functionalities of the simulator for the current purpose of the simulator.

Table 5.7: Validity dimensions of the simulator for the current task.

	N	**M**	**SD**
Structural validity	20	3.2	.77
Process validity	20	3.5	.70
Psychological reality	20	3.7	.71

5.6.2.4 Situation awareness and performance

Variables related to SA, mental workload, and performance are given in Table 5.8 in terms of mean ranks and means to provide an overview of the normal and non-normal distributed values.

A non-parametric Kruskal–Wallis was conducted due to the sample differences between conditions and low sample size. Significant differences between conditions were found for arrival and departure punctuality and departure delay, respectively $\chi^2(2) = 14.48$, $p = .001$, $\chi^2(2) = 14.36$, $p = .001$ and $\chi^2(2) = 7.44$, $p = .024$. These differences were mainly found between conditions 1 and 3 and conditions 2 and 3, whereas condition 3 had lower mean ranks for arrival and punctuality and higher mean ranks for departure delays. In general, this implies that some performance indicators score worse with more train delays, but not in terms of SA. Also a significant difference was found for perceived mental workload; $\chi^2(2) = 6.22$, $p = 0.045$. Participants in condition 3 had a significant greater workload compared to participants in conditions 1 and 2, in which a clear distinction can be drawn between the lightly and the moderately disrupted train traffic condition.

Further analyses were conducted to explore the SAGAT score differences in groups with regard to workstation and experience as a planner. A trend was found for differences in SAGAT scores between the "turn" and the "through" workstation: $U = 18.0$, $p = .09$. The "through" workstation had a higher SAGAT score (mean rank = 11.3) in comparison to the "turn" workstation (mean rank = 7.0). It is notable that the perceived difficulty between both workstations did not significantly differ from each other. Also no significant difference between the workstations was found for perceived SA.

It was also expected that planners would have higher SAGAT scores, as a planning role requires a more careful future train traffic flow assessment. However, no significant differences in SAGAT scores were found between the two groups. However, a significant difference was found between non-planners and planners on perceived SA: $U = 13.5$, $p = .013$. Non-planners indicated a higher level of perceived SA (mean rank = 14.1) in comparison to planners (mean rank = 7.6).

Table 5.8: Measurements of situation awareness and performance.

	Condition 1				Condition 2				Condition 3			
	N	**Mean rank**	**Mean**	**SD**	**N**	**Mean rank**	**Mean**	**SD**	**N**	**Mean rank**	**Mean**	**SD**
SAGAT (%)	4	11.5	58.2**	27.54	4	7.9	41.1**	11.27	9	8.4	43.8	19.57
Perceived SA (1-4)1	6	9.6	3.0*	.21	4	11.9	3.2**	.33	9	9.4	3.0*	.31
Arrival punctuality (%)	6	15.8[b]	88.8	5.98	4	15.0[a]	86.5**	7.68	10	5.5[ab]	70.7	4.45
Departure punctuality (%)	6	13.3[b]	91.0	6.60	4	13.1[a]	89.0**	1.63	10	7.8[ab]	83.3	7.94
Arrival delay (s)	6	5.7	.74	.29	4	5.3	.75**	.30	10	15.5	1.7	.11
Departure delay (s)	6	6.7[a]	.52	.35	4	7.3	.57**	.15	10	14.1[a]	1.2	.53
Platform consistency (%)	6	12.4	96.5*	6.01	4	11.3	94.5**	6.56	10	9.1	91.1	8.52
Perceived MWL (1-5)	5	6.5[a]	1.7	.52	3	5.5[a]	1.5**	.61	10	12.2[aa]	2.4*	.59

* Significant on the Kolmogorov-Smirnov normality test $p < .05$.
** Normality test not computable.
[a] $p \leq .05$.
[b] $p \leq .001$.
[1] based on a 4-point Likert scale

In terms of learning effects, it is notable that although participants were knowledgeable of the specific train delays in condition 1, no significant differences in the various performance indicators were found in condition 2. This indicates that having knowledge of the train delays in this scenario did not entail a better SA, performance, or mental workload. Spearman correlations were drawn to investigate the relationship between perceived SA, SA probes, and mental workload. Between these variables, a trend for a negative correlation was found between perceived SA and perceived mental workload; $\rho = .43$, $p =.09$. As such, a higher level of perceived SA is related to a lower perceived mental workload.

5.6.2.5 Implicit situation awareness

Given the differences in levels of performance and SA, Spearman's correlations were drawn within each condition. In conditions 1 and 3, negative correlations were found between arrival punctuality and SAGAT scores; $\rho = .80$, $p = .20$, respectively $\rho = .49$, $p = .18$. Contrary to expectations, these findings show that a higher level of arrival punctuality leads to a lower SA probe score. Again contrary to expectations, moderate to large negative correlations were found for departure punctuality and SA in condition 1; $\rho = .40$, $p = .09$, condition 2; $\rho = .63$, $p = .37$, and condition 3; $\rho = .68$, $p = .05$. For arrival delay and SAGAT scores, a positive correlation was found in condition 1; $\rho = .80$, $p = .20$, namely a longer arrival delay leads to a higher level of SA. Likewise a moderate positive correlation was found between departure delay and SAGAT scores in condition 1; $\rho = .40$, $p = .60$ and in condition 3; $\rho = .44$, $p = .24$, with contradicting results in condition 2; $\rho = .64$, $p = .37$. Finally, a moderate negative correlation was found in condition 3 between platform consistency and SAGAT scores; $\rho = .49$, $p = .18$. In sum, there is an unexpected general tendency across conditions for a negative relation between SA probes and multiple performance indicators; for instance, a higher score on SA probes is related to a lower level of arrival and departure punctuality performance. Likewise, a higher performance score on arrival and departure is related to lower SA probe scores. However, it can be posited that the lower absolute SAGAT values in comparison to the high punctuality values indicate a level of implicit SA.

Secondly, building on findings from earlier studies and the pilot study, we expected to find a correlation between work experience as a TTC and SAGAT scores. A trend for a moderate negative correlation was found, $\rho = .41$, $p = .10$; in which greater work experience is related to lower SA probe scores. Similarly, a moderate effect size was found between work experience in the railway sector and SAGAT scores; $\rho = .33$, $p = .20$. Both results support the indication of the presence of TTCs' implicit SA. Additionally, in line with the previous findings, negative correlations were found between the two types of work experience (i.e., as a TTC and in the railway domain) and perceived SA; respectively $\rho = .38$, $p = .11$, $\rho = .33$, $p = .14$. Very small effect sizes were found for the relation between the two types of work experience and perceived mental workload.

The third analysis focused on the analysis of SA queries for each SA level (see Table 5.9). The values for the "turn" and the "through" workstation were also included to investigate the development of SAGAT values across SA levels. In line with the results of the pilot study, a qualitative assessment of the SA scores indicates structurally lower level-1 SA scores in comparison to level-2 SA scores. Subsequently, level-3 SA scores were lower than level-2 SA scores. As such, the relatively low level1 SA scores might support the notion of implicit SA.

Table 5.9: SA probes per SA level.

	"Turn" workstation		**"Through" workstation**		**Total**	
	total # items (N=9)	**% correct**	**total # items (N=8)**	**% correct**	**total # items (N=17)**	**% correct**
Level 1	62	45	66	35	128	40
Level 2	37	76	48	54	85	64
Level 3	23	52	24	29	47	40

5.6.3 Discussion

In line with the approach using a set of three analyses to identify levels of implicit SA, the results of the main study, with regard to the first analysis, lend support for the presence of implicit SA in TTCs. In both lightly and moderately disrupted conditions, fairly low SAGAT values were found with no significant difference between condition 2 (minor delays) and condition 3 (moderate delays), and with only a very small difference in absolute values. This finding indicates that a more active role of operators does not imply lower SAGAT scores.

The presence of correlations between SAGAT scores and multiple performance indicators shows that the absolute low SAGAT values cannot be attributed only to the absence of awareness of the situation. However, the negative relationship – namely a higher SAGAT score is related to worse performance – is contrary to expectations. This outcome can possibly be used as an illustration of the sensitivity of SAGAT as a measurement tool, calling into doubt its predictive validity (e.g., Salmon et al., 2006).

Secondly, negative correlations, although moderate in effect size, were found between both work experience as TTC and SAGAT scores, and between work experience in the railway sector and SAGAT scores. These correlations indicate that more work experience in general is related to lower SAGAT scores, which is in line with the presence of implicit SA in expert operators.

Finally, the existence of similar trends for lower level-1 SA scores in comparison to level-2 SA scores as in the pilot study, might indicate unconscious processing of the perception during situation assessment, thus supporting the presence of implicit SA.

A number of limitations in the current study can be identified, such as the small sample size in the three conditions and the inclusion of a dashboard that indicated the system's performance. Although the dashboard was introduced for operators to reflect on their performance and to motivate an active participation, this could have influenced their actual behavior.

5.7 General discussion and conclusion

5.7.1 Summary of the findings

In the present research, a set of three analyses was used to investigate the presence of implicit SA. In line with expectations, the findings of both the pilot and the main study support the indications of implicit SA through the fairly low absolute values of the SAGAT probes, the identification of correlations between SAGAT scores and multiple performance indicators, the negative relation between work experience and SAGAT scores, and deviations in SA levels (level-1 SA scores were lower than level-2 SA scores). Although the effect sizes of the correlations could be larger, the persistent and consistent trends underline this implication.

As similar results were found with regards to the SAGAT scores in both the pilot and main study with regards to the light disrupted condition, the findings of the main study exclude low SAGAT scores resulting from the operator's monitoring mode. Also the fact that traffic controllers at two different regional control centers had low SAGAT scores provides mild support for the generalizability of the findings, although care should be taken because of the relatively small sample sizes and the likely related non-significant correlations.

The inconsistent correlations between SAGAT scores and performance, and the indications of levels of implicit SA, found in these studies may also lead to remarks about and call into question measurements of explicit SA, as underlined in Endsley's three-level model using the information-processing paradigm as a foundation. Critical remarks can be aimed at the use of the SAGAT method to capture the cognitive strategies of operators in terms of SA as a product. Recent discussions on SA theory propose a paradigm shift leading to distributed cognition, in which the role of the working memory – which is thus responsible for conscious, explicit SA – as a main component in the development of SA is called into question (e.g., Chiappe, Vu, Rorie, & Morgan, 2012; Stanton, Salmon, & Walker, 2014). As such, methods that focus on situation assessment could be more sensitive to the role of implicit processes and could provide richer input in the development of an understanding of TTCs' situation awareness. Alternative methods could be the use of an eye-tracker (van de Merwe, van Dijk, & Zon, 2012) or real-time query techniques combined with accuracy and response time based measurements, such as the situation present assessment method (SPAM) (Durso, Dattel, Banbury, & Tremblay, 2004).

5.7.2 Limitations

It should be remarked that official performance indicators for TTCs are yet to be identified and formalized by the railway infrastructure organization. Another limitation that needs to be addressed is the evaluation of the SAGAT probes: probes that should have straightforward answers could be interpreted in different ways. For instance, a correct level-3 SA answer for the predicted status of a certain train will be (e.g., on time, delayed arrival but punctual departure, delayed arrival and delayed departure, etc.) depends on the beholder. Punctuality has a threshold of three minutes, as defined by the railway infrastructure organization. However, it is only assumed that the same definition is held by participants. This issue makes the use of queries not only contentious, but also somewhat complicated. Finally, more serious scenarios could be developed to investigate the SA development in these circumstances.

5.7.3 Balancing explicit and implicit situation awareness

Operators' awareness of what is happening in a dynamic environment has been associated with levels of safety (Sarter & Woods, 1991; Stanton, Chambers, & Piggott, 2001). For instance, accident investigations involving major air carriers found that 71% of the accidents can be classified as human error causes, of which 88% can be attributed to SA issues (Endsley, 1995c). Within these investigations, 72% of the cases were related to level-1 SA errors, and 22% and 6% were related to level 2 and level 3, respectively. From a managerial point of view, it may be hard to establish whether operators made safety critical errors that led to injuries or possibly death by not having identified, understood, or foreseen critical events.

Another hurdle to gaining organizational acceptance of unconscious processes, is that it is very difficult for operators to explicate their reasoning due to their unconscious processes. Therefore, a method has been proposed to train operators in developing a good explicit SA, in addition to efforts to improve design and automation issues (Endsley, 2000b). An example of a widely supported SA training program is the Enhanced Safety Through Situation Awareness Integration (ESSAI), a program funded by the European Commission to train pilots to improve their SA (ESSAI, 2000). Although the program assessed and recognized the importance of implicit SA, this was not incorporated in the training program due to its limited applied use and novel advancements, which so far remain rather poorly investigated.

To conclude that train traffic control operations need to follow a similar route to explicate situation awareness may be a simplified approach. As the naturalistic decision-making field also states through the recognition-primed decision model; a blend of intuition and analysis is desired. If operators would to adopt explicit, analytical cognitive processes, their performance would be too slow, while a pure intuitive cognitive process would be too risky (Klein, 2008).

5.7.4 Future work

The current findings provide initial insights into the situation awareness of TTCs, which can be put into relation with the different skill classifications (e.g. 'competently (un)aware' vs. 'incompetently (un)aware') that is used by the train traffic learning center. As such, the learning center recognizes the existence of 'competently unaware' train traffic controllers, i.e. train traffic controllers with implicit situation awareness, in which the current findings can be used evaluate its learning program. More research may be needed to investigate on which situational aspects operators should have an explicit SA, while in other conditions a fast, intuitive (implicit SA) is preferred. Following this line, interface design or decisionsupport systems can be optimized or designed for.

Further research is needed to cope with the demanding changes in the railway sector by a denser and complex train traffic infrastructure. Here, the increasing role of automation in a new traffic management system and use of decision support systems are expected to play a significant role. Insights into current ways of SA development – for instance, with regard to abrupt transitions whereby operators directly need to switch from a passive monitoring mode to an active role in dealing with different degrees of disruptions – are yet to be garnered. In addition, differences between regional control centers, each having their own unique infrastructural characteristics and organizational culture, also need thorough investigation. Nonetheless, the limitations found with the current SA approach provide a foundation for research on SA on larger units of analyses, that is, at team or network level. Furthermore, the presence of implicit SA at the operational level and its policy or political implications for strategic decision makers could be investigated. Also the application of the recognition-primed decision model could be further investigated for practical implications.

References

Adams, M. J., Tenney, Y. J., & Pew, R. W. (1995). Situation awareness and the cognitive management of complex systems. *Human Factors, 37*(1), 85–104.

Chiappe, D., Vu, K.-P. L., Rorie, C., & Morgan, C. (2012). A situated approach to shared situation awareness. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 56*(1), 748–752.

Chugh, D., & Bazerman, M. (2007). Bounded awareness: What you fail to see can hurt you. *Mind & Society, 6*(1), 1–18.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Croft, D. G., Banbury, S. P., Butler, L. T., & Berry, D. C. (2004). The role of awareness in situation awareness. In S. Banbury & S. Tremblay (Eds.), *A cognitive approach to situation awareness: Theory and application* (pp. 82–103). Bodmin, Cornwall: MPG Books Ltd.

Dekker, S. W. A., Hummerdal, D. H., & Smith, K. (2010). Situation awareness: Some remaining questions. *Theoretical Issues in Ergonomics Science, 11*(1–2), 131–135.

Dijksterhuis, A., Bos, M. W., Nordgren, L. F., & van Baaren, R. B. (2006). On making the right choice: The deliberation-without-attention effect. *Science, 311*(5763), 1005–1007.

Durso, F. T., Dattel, A. R., Banbury, S., & Tremblay, S. (2004). SPAM: The real-time assessment of SA. In S. Banbury & S. Tremblay (Eds.), *A cognitive approach to situation awareness: Theory and application* (pp. 137–154). Bodmin, Cornwall: MPG Books Ltd.

Durso, F. T., & Sethumadhavan, A. (2008). Situation awareness: Understanding dynamic environments. *Human Factors, 50*(3), 442–448.

Endsley, M. R. (1988a). Design and evaluation for situation awareness enhancement. *Proceedings of the 32nd human factors society annual meeting* (Vol. 32, pp. 97–101).

Endsley, M. R. (1988b). Situation awareness global assessment technique (SAGAT). *Proceedings of the IEEE 1988 national aerospace and electronics conference (NAECON)* (Vol. 32).

Endsley, M. R. (1995a). Towards a theory of situation awareness in dynamic systems. *Human Factors, 37*(1), 32–64.

Endsley, M. R. (1995b). Measurement of situation awareness in dynamic systems. *Human Factors, 37*(1), 65–84.

Endsley, M. R. (2001). Designing for situation awareness in complex systems. *Proceedings of the Second International Workshop on Symbiosis of Humans, Artifacts and Environment.*

Endsley, M. R. (2015). Situation awareness misconceptions and misunderstandings. *Journal of Cognitive Engineering and Decision Making, 9*(1), 4–32.

Endsley, M. R., Bolté, B., & Jones, D. G. (2003). *Designing for situation awareness.* New York, NY: Taylor & Francis Group.

Endsley, M. R. (2000a). Direct measurement of situation awareness: Validity and use of SAGAT. In M. R. Endsley & D. J. Garland (Eds.), *Situation awareness analysis and measurement* (pp. 147–174). Mahwah, NJ: Lawrence Erlbaum Associates.

Endsley, M. R. (2000b). Training for situation awareness. In M. R. Endsley & D. J. Garland (Eds.), *Situation awareness analysis and measurement* (pp. 349–365). Mahwah, NJ: Lawrence Erlbaum Associates.

Endsley, M. R. (1995c). A taxonomy of situation awareness errors. In R. Fuller, N. Johnston, & N. McDonald (Eds.), *Human factors in aviation operations* (pp. 287–292). Aldershot: Avebury Aviation.

Endsley, M. R., & Rodgers, M. D. (1997). *Distribution of attention, situation awareness, and workload in a passive air traffic control task: Implications for operational errors and automation* No. DOT/FAA/AM-97/13 ed. Federal Aviation Administration Office of Aviation Medicine.

Endsley, M. R. (1997). The role of situation awareness in naturalistic decision making. In C. E. Zsambok & G. Klein (Eds.), *Naturalistic decision making* (pp. 269–283). Mahwah, NJ: Lawrence Erlbaum Associates.

ESSAI (2000). *WP1 orientation on situation awareness and crisis management.* Workpackage Report. ESSAI/NLR/WPR/WP1. EC DG-TREN. Contract No.: 2000GRD1-10450.

Ferreira, P., & Balfe, N. (2014). The contribution of automation to resilience in rail traffic control. In D. Harris (Ed.). *Engineering psychology and cognitive ergonomics* (Vol. 8532, pp. 458–469). Springer International Publishing.

Golightly, D., Sandblad, B., Dadashi, N., Andersson, A. W., Tschirner, S., & Sharples, S. (2013). A socio-technical comparison of rail traffic control between GB and Sweden. In N. Dadashi, A. Scott, J. R. Wilson, & A. Mills (Eds.), *Rail human factors: Supporting reliability, safety and cost reduction* (pp. 367–376). London: Taylor & Francis.

Gugerty, L. J. (1997). Situation awareness during driving: Explicit and implicit knowledge in dynamic spatial memory. *Journal of Experimental Psychology: Applied, 3*, 42–66.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). Amsterdam: North-Holland.

Kaber, D. B., Perry, C. M., Segall, N., McClernon, C. K., & Prinzel, L. J. III, (2006). Situation awareness implications of adaptive automation for information processing in an air traffic control-related task. *International Journal of Industrial Ergonomics, 36*(5), 447–462.

Kaempf, G. L., Klein, G. T., Thordsen, M. L., & Wolf, S. (1996). Decision making in complex naval command-and-control environments. *Human Factors, 38*(2), 220–231.

Kaempf, G. L., & Orsanu, J. (1997). Current and future applications of naturalistic decision making in aviation. In C. E. Zsambok & G. Klein (Eds.), *Naturalistic decision making* (pp. 81–90). Mahwah, NJ: Lawrence Erlbaum Associates.

Kahneman, D. (2012). *Thinking, fast and slow*. London: Penguin Books Ltd.

Klein, G. (2008). Naturalistic decision making. *Human Factors, 50*(3), 456–460.

Knieps, G. (2013). Competition and the railroads: A European perspective. *Journal of Competition Law and Economics, 9*(1), 153–169.

Lipshitz, R., Klein, G., Orasanu, J., & Salas, E. (2001). Taking stock of naturalistic decision making. *Journal of Behavioral Decision Making, 14*(5), 331–352.

Lo, J. C., Pluyter, K. R., & Meijer, S. A. (2016). Individual markers of resilience in train traffic control: The role of operators' goals and strategic mental models and their implications for variation, expertise and performance. *Human Factors, 58*(1), 80–91.

Lo, J. C., Sehic, E., & Meijer, S. A. (2014). Explicit or implicit situation awareness? Situation awareness measurements of train traffic controllers in a monitoring mode. In D. Harris (Ed.). *Engineering psychology and cognitive ergonomics* (Vol. 8532, pp. 511–521). Springer International Publishing.

Matthews, M. D., & Beal, S. A. (2002). *Assessing situation awareness in field training exercises*. Orlando, FL: U.S. Army Research Institute for the Behavioral and Social Sciences.

Newell, B. R., & Shanks, D. R. (2014). Unconscious influences on decision making: A critical review. *Behavioral and Brain Sciences, 37*, 1–19.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making, 2*(2), 140–160.

Raser, J. C. (1969). *Simulations and society: An exploration of scientific gaming*. Boston, MA: Allyn & Bacon.

Rasmussen, J. (1983). Skills, rules, and knowledge; Signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man and Cybernetics, 13*(3), 257–266.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology, 118*(3), 219–235.

Salmon, P., Stanton, N., Walker, G., & Green, D. (2006). Situation awareness measurement: A review of applicability for C4i environments. *Applied Ergonomics, 37*(2), 225–238.

Salmon, P. M., Stanton, N. A., Walker, G. H., Jenkins, D., Ladva, D., Rafferty, L., & Young, M. (2009). Measuring situation awareness in complex systems: Comparison of measures study. *International Journal of Industrial Ergonomics, 39*(3), 490–500.

Sarter, N. B., & Woods, D. D. (1991). Situation awareness: A critical but ill-defined phenomenon. *The International Journal of Aviation Psychology, 1*(1), 45–57.

Sharples, S., Millen, L., Golightly, D., & Balfe, N. (2011). The impact of automation in rail signalling operations. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 225*(2), 179–191.

Simon, H. A. (1983). *Reason in human affairs.* Stanford, CA: Stanford University Press.

Sneddon, A., Mearns, K., & Flin, R. (2006). Situation awareness and safety in offshore drill crews. *Cognition, Technology & Work, 8*(4), 255–267.

Stanton, N. A., Chambers, P. R. G., & Piggott, J. (2001). Situational awareness and safety. *Safety Science, 39*(3), 189–204.

Stanton, N. A., Salmon, P. M., & Walker, G. H. (2014). Let the reader decide: A paradigm shift for situation awareness in sociotechnical systems. *Journal of Cognitive Engineering and Decision Making*.

Stanton, N. A., Stewart, R., Harris, D., Houghton, R. J., Baber, C., McMaster, R., ... Green, D. (2006). Distributed situation awareness in dynamic systems: Theoretical development and application of an ergonomics methodology. *Ergonomics, 49*(12–13), 1288–1311.

Steenhuisen, B. (2009). *Competing public values: Coping strategies in heavily regulated utility industries*. Enschede: Gildeprint Drukkerijen.

Steenhuisen, B. (2014). Cutting dark matter. Professional capacity and organizational change. *Journal of Organizational Ethnography, 3*(2), 152–168.

Strater, L. D., Endsley, M. R., Pleban, R. J., & Matthews, M. D. (2001). *Measures of platoon leader situation awareness in virtual decision-making exercises*. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

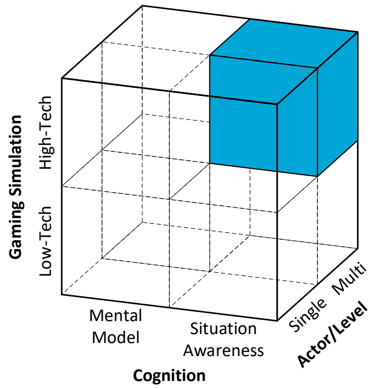
van de Merwe, K., van Dijk, H., & Zon, R. (2012). Eye movements as an indicator of situation awareness in a flight simulator experiment. *The International Journal of Aviation Psychology, 22*(1), 78–95.

Van de Velde, D. M. (2001). The evolution of organisational forms in European public transport. *7th conference on competition and ownership in land passenger transport* (pp. 25–28).

Wilson, J. R., & Norris, B. J. (2005). Rail human factors: Past, present and future. *Applied Ergonomics, 36*, 649–660.

6 Balancing organizational and academic research: Investigating train traffic controller's geographical workspace design and team situation awareness using gaming simulations

Apart from a minor update, this chapter has been published as:
Lo, J. C., Sehic, E., & Meijer, S. A. (2019). Balancing organizational and academic research: Investigating train traffic Controller's geographical workspace design and team situation awareness using gaming simulations. *Journal of Rail Transport Planning & Management, 10*, 34-45.



Abstract
This chapter focuses on the topic of the practical value of scientific research, specifically in relation to team situation awareness investigations. This study focuses on providing insights into the use of a human-in-the-loop simulator in which an organizational research question investigates the impact of multiple geographical workspace designs, while in parallel human factors research is conducted to investigate the concept of team situation awareness from an academic research interest. Finding a balance between the practical and academic implications in one research design and its findings does not rely on a trivial approach. The current article aims to contribute on several levels: (1) to illustrate the balance between research for practice and research for academia through the applications of gaming simulations; (2) to illustrate the use of gaming simulations for railway traffic operations and (3) to provide insights in team SA development in railway traffic operations using gaming simulations.

Keywords: Balancing research and practice; organization; gaming simulation; research design; traffic control; team situation awareness

6.1 Introduction

In many European countries, there is a steady increase in passenger as well as freight train traffic volumes (5th EC rail market monitoring report, 2016). In a number of countries this increase leads to issues with capacity allocation, i.e. network saturation. The railways in the Netherlands are amongst the busiest in Europe in terms of density and frequency of utilization with very high network saturation values of 50.000 train-kilometers per each line-kilometer per year (Ramaekers, De Wit & Pouwels, 2009; 5th EC rail market monitoring report, 2016). In line with these developments, long-term governmental and organizational targets have been set to facilitate the growing demand on diversified and more dense train schedules, while maintaining its core business to deliver reliable and punctual train services together with railway undertakings (Goverde and Meng, 2011; Van de Velde, 2013). The Dutch railway infrastructure organization ProRail is responsible to provide this increase in capacity of their infrastructure network, in which solutions are sought in near and far future process innovations that directly impact the task-space of railway traffic operators (Meijer, 2012).

The need to test the impact of these designs beforehand is recognized by the Dutch railway sector, in which the use of gaming simulation environments is an applied method. To simulate alternative modes of the system, different types of simulation environments have been used as a research tool, in which different future configurations (e.g. infrastructural, role, and/or procedural changes) can be compared (Meijer, 2015). These simulated environments are designed to resemble an operating model of reality that exhibit high validity levels in terms of structural and process validity and psychological reality (Ryan, 2000; Raser, 1969). They vary from low-tech multi-actor table-top environments to high-tech single- and multi-actor human-in-the-loop simulators (e.g. Kortmann and Sehic, 2011; Lo & Meijer, 2013; Albers, Lo, Sehic, Van 't Woudt, 2018) . Altogether this spectrum of environments can be denoted as gaming simulations, as gaming concepts - such as immersion - and gaming simulation design concepts - such as roles, rules etc. - lay the fundaments in all aforementioned environments (Meijer, 2012). Different typologies of gaming simulations exist where the main distinctions are made between simulation/games for research, training, design and policy-making (Grogan and Meijer, 2017; Peters, Vissers & Heijne, 1998).

Hybrid forms of gaming simulations can exist as the typologies for gaming simulations do not adhere to strong categorical boundaries. Besides from the fact that gaming simulations can be used as a tool to support strategic decision-makers, they can also be used as a research environment. In the current study, the question posed by the infrastructure manager as organization is focused on which distribution of the geographical workspace designs, i.e. the geographical area a train traffic controller is responsible for, would be most manageable and preferable for railway traffic operations. Due to changes in the infrastructure, it is possible that certain geographical areas have less switches, therefore needing less supervisory control. In terms of workload balance and mitigating options, it might therefore be preferable to reallocate geographical areas. The implication of identified geographical workspace design options is explored for train traffic controllers and the regional network controller.

Although gaming simulations often strive to be representational and rich in their situational context and environmental cues in relation to the reference system, they still represent controlled environments. As such, they also can function as a test environment to conduct descriptive, naturalistic and/or experimental social scientific studies. Especially from an engineering psychology or human factors approach, investigations in the cognitive process of operators are key in the research and design of complex socio-technical systems such as the railways (Wilson, Farrington-Darby, Cox, Bye, & Hockey, 2007). Especially the notion of the cognitive concept of situation awareness (SA) has been a profound concept within the human factors literature as an indicator of good decision-making of operators, such as in railway traffic operations (Endsley, 1995; Tschirner, Sandblad, & Andersson, 2014). To perceive and being able to comprehend and make predictions of elements in the system are all components that fall within one of the SA definitions (Endsley, 1988). However, especially in complex systems such as the railways, the command and control structure is composed of technical components and multi-actor interactions. For instance, frequent returning questions following a severe disruption are: what could we have done better in terms of operational efficiency? Where did the situation get out of control, i.e. where did operators lose their ability to maintain a sufficient situation awareness? Where did technical systems fail to support operators? Answering these questions aids in understanding what the dynamics in the operational setting are and is key for further optimizations in the system's design (Roth, Multer & Raslaer, 2006; Farrington-Darby and Wilson, 2009).

The current study focuses on two research questions that are investigated in parallel; the organizational research question: what is the impact of each of the four options having differently distributed geographical areas for train traffic management? The second question relates to an academic research question: how does team situation awareness develop at a regional control center? Investigating multiple and different types of research questions in one study may be challenging when it comes to finding a balance between practical relevance and context for the organization and academic rigor for the academic community. This study uses a framework by Hevner, March and Park (2004) to identify and address the different components when working from both an organizational as academic perspective.

In sum, the current article aims to contribute on several levels: (1) to provide insights in the balance between research for practice and research for academia through the applications of gaming simulations; (2) to provide insights in the use of gaming simulations for organizational research and (3) to provide insights in academic research question on team SA development in railway traffic operations using gaming simulations.

6.2 Balancing organizational and academic research

Organizations are often limited in resources, such as time and personnel, however wish to have useful findings with academic rigor. An approach in finding this balance is discussed by Hevner et al., (2004), in which they present a framework that amongst others focuses on nowledge for decision-making in organizations, while combining this with design- and behavorial-science paradigms (see Figure 6.1).

In Figure 6.1, the 'environment' refers to the problem space and business needs of organizations. 'Relevance' is obtained by framing organizational research activities that address business needs. On the other side, the 'knowledge base' refers to theories, frameworks, methodologies and/or instruments in the operationalization of a research study. 'Rigor' is achieved by applying the knowledge base in an appropriate manner. For this knowledge base often design-science or behavioral-science paradigms are used, in which behavioral paradigms are typically rooted in empirical and data collection techniques and design paradigms are focused on computational, mathematical and empirical methods. Both the 'environment' as well as the 'knowledge base' space influence the 'research' that is conducted. Depending on the goal of the 'research' space, which may be oriented on exploratory or explanatory research ('theories') or utility ('artifacts'), different research methods such as simulation, can be applied to assess and possibly refine the research.

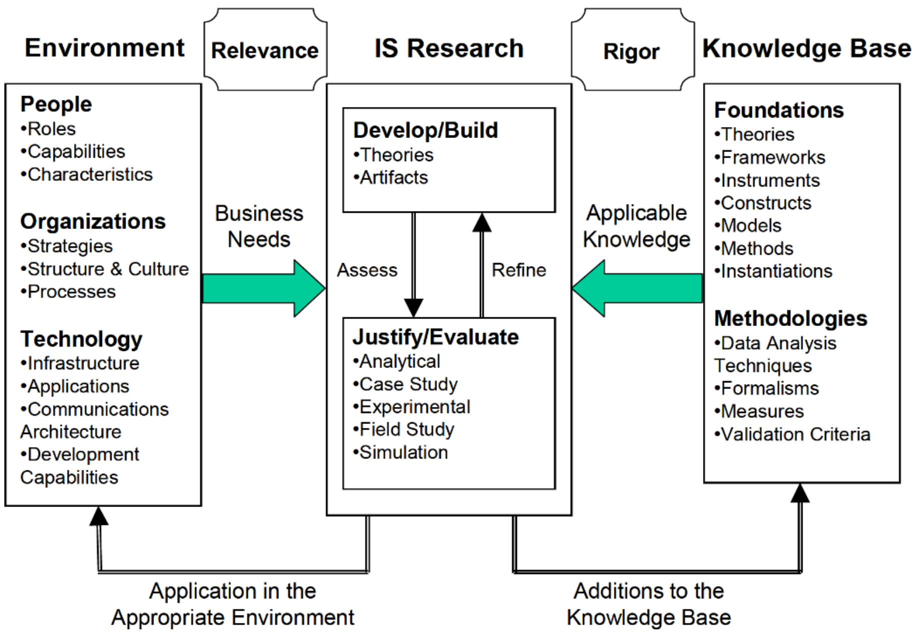


Figure 6.1: Information systems research framework by Hevner et al., (2004).

When applying this framework to the research study in this article, the environment is represented by one of the railway regional traffic control centers, where management has the need to investigate the impact of the four options that have different distributed geographical areas for train traffic management. The knowledge base consists of the psychological paradigm, experimental research methods and methods on gaming simulation design. Additionally, the development of team cognition amongst railway traffic operators is investigated from an academic research interest. Gaming simulation is used as a research method to conduct the study; more specifically with human-in-the-loop simulators.

6.3 Team situation awareness in railway traffic operations

Unraveling the black box of operator's cognition is complex and laborious work. However, the benefits of understanding operator's cognitive processes and decision-making can contribute to different fields, such as selection and assessment, training, but also the development of modeling and the design of an operational system (e.g. Corman, D'Ariano, Hansen, & Pacciarelli, 2011; Kauppi, Wikström, Sandblad, & Andersson, 2006; OnTime, 2013; Samà, Meloni, D'Ariano, & Corman, 2015).

As briefly introduced, situation awareness is seen as a key cognitive concept within the setting of human-machine interaction and socio-technical systems. Few cognitive engineering studies have investigated the (shared) situation awareness of railway traffic operators (e.g. Golightly, Wilson, Lowe, & Sharples, 2010; Lo & Meijer, 2013; Lo & Meijer, 2019; Roth et al., 2006). The results from observations and human-in-the-loop simulator experiments showed indications for tacit knowledge and implicit levels of individual situation awareness when the popular situation awareness global assessment tool (SAGAT) was applied to assess the SA of train traffic controllers (Lo, Sehic, Brookhuis, & Meijer, 2016; Steenhuisen, 2009). Similar findings have frequently been reported with expert operators and with other SAGAT related measurement techniques such as the Situation Present Assessment Technique (SPAM) and the Situation Awareness Rating Technique (SART) (Croft, Banbury, Butler, & Berry, 2004; Golightly et al., 2010). All in all, these studies indicated weak support for traditional information-processing theories on SA, such as the three-level model by Endsley (1988), and stronger support for macrocognitive theories and methods when assessing team and network SA.

According to macrocognitive approaches, team cognition goes beyond the sum of individuals and should be measured on a team level (Cooke et al., 2013; Stanton et al., 2015). Following this theory, team SA could therefore be defined as the timely and adaptive response through team interactions (communication and coordination) when roadblocks are encountered (Gorman, Cooke, Pederson, Olena, & DeJoode, 2005). As severe railway disruptions hardly can be contained locally and controlled by single operators, the current article zooms in on the team level in order to develop an understanding of the collaborative nature between operators within a regional control center. More specific, the current investigation includes the development of situation awareness between railway traffic operators and potential social-psychological factors that might impact this development.

Following Figure 6.2, elements beyond the cognitive processes of an individual come into attention in team collaborations. That is, team cognition is not determined by aggregates of individual-level properties (Cooke et al., 2013). Factors such as cohesion and trust influence the group dynamics, which play an important role within the control room environment (Farrington-Darby and Wilson, 2009; Forsyth, 2001). In terms of command and control structures such as in railway traffic control, operator communication and cooperation influence the development of situation awareness on a team and on a network level (Lo & Meijer, 2013).

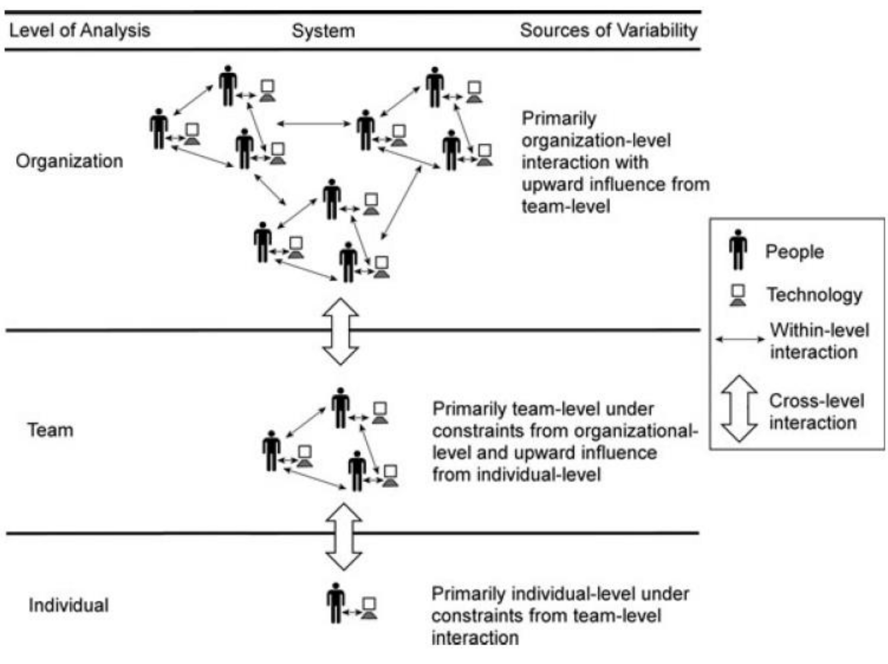


Figure 6.2: Sources of variability at each level of analysis (Cooke, Gorman, Myers, & Duran, 2013).

Simulation environments for research

Within ProRail, different simulation environments are used to answer research questions. These questions are predominantly focused on investigations of future changes impacting railway traffic operations. Different tools are used to support decision-making, such as computer simulations and human-in-the-loop simulation. Multiple tools have been developed and used for the (Dutch) railways to investigate optimizations in the railway system, e.g. for infrastructure maintenance, passengers, driver advisory systems but also for the allocation of geographical areas to traffic controllers (Galapitage, Albrecht, Pudney, Vu & Zhou, 2018; Ghaemi, Zilko, Yan, Cats, Kurowicka & Goverde, 2018; Van Aken, Bešinović & Goverde, 2017; Zhao, Wang & Peng, 2018). However, more research is still needed with regards to the cognitive modeling of railway traffic operators (e.g. Aydoğan, Lo, Meijer & Jonker, 2014). As such, the assessment of operators, especially their knowledge and skills, are deemed important in the decision-making of intended changes that affect the management of the railway system. Human-in-the-loop simulation is therefore used by the Dutch railway infrastructural manager to facilitate this need.

6.3.1 Human-in-the-loop simulation: PRL game and FRISO

The human-in-the-loop simulation environment that is used in this study consists of two components. The interface that closely represents that of the actual train traffic control system is called PRL (in Dutch: PRocesLeiding) Game, while the engine of the simulator is a computer simulation tool that is used within ProRail to perform operations research. FRISO (Flexible Rail Infrastructure Simulation of Operations) is a simulation tool with discrete, dynamic, stochastic and deterministic properties (Middelkoop, Meijer, Steneker, Sehic, & Mazzarello, 2012). Key features of the tool are the automated construction of a simulation model, which is accomplished through connections with an existing infrastructure database, a flexible infrastructure editor and means to perform single and multiple (stochastic) simulation experiments, and more recently the connection with a train traffic control module (Middelkoop & Loeve, 2006; Middelkoop et al., 2012). In essence, FRISO is developed to simulate the interaction between and behavior of trains that are given a certain scenario (such as disruption in the infrastructure, rolling stock delay etc.) to assess timetable robustness and to identify bottlenecks. The level in which FRISO simulates components in relation to the safety system is on an abstract level. Additionally, there are several train driver simulation profiles that can be used for different train driver's driving behavior. A number of studies have validated FRISO (e.g. Middelkoop and Loeve, 2006; Roungas, Meijer & Verbraeck, 2018). Figure 6.3 depicts an overview of various modules of FRISO.

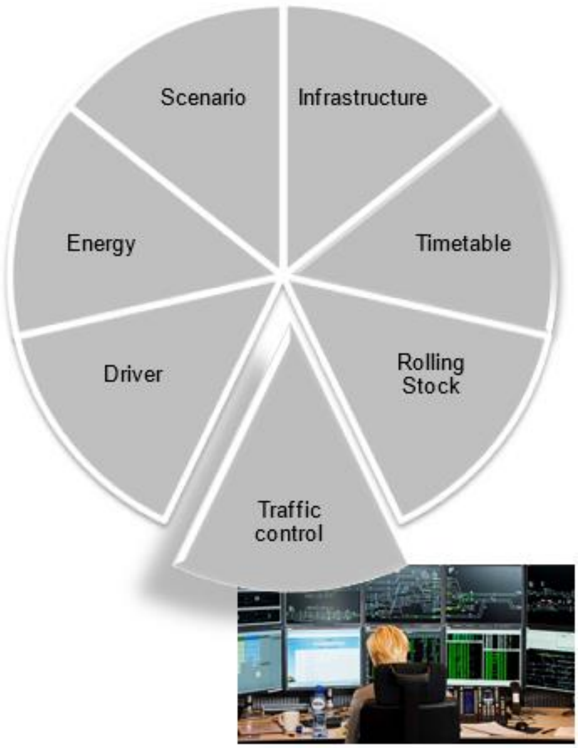


Figure 6.3: Modules of FRISO where the train traffic control module in FRISO serves as the engine for the human-in-the-loop simulator.

In terms of interactions between FRISO and PRL Game, train traffic controllers can make changes, such as changes in route setting, revoke signal status and apply disruption mitigation procedures. Human-in-the-loop simulation sessions are particularly applied in conditions where the impact on railway traffic operations is expected to be substantial and where variation in operator's behavior is predicted. Also, in an earlier stage of an infrastructural modification project, operator's feedback on adaptations in the infrastructure, timetables or operational procedures is seen as valuable. For research questions in a very early stage of a design process, computer simulation studies may be applied as a tool to conduct the investigation, whereas in more mature stages of a design process human-in-the-loop simulation sessions are applied. An example of a human-in-the-loop simulation with individual operators is a workload investigation of train traffic controllers with a newly planned infrastructural layout. Next to studies that focus on the impact of certain changes, especially the debriefing after the simulation sessions can have an important role. The debriefing can be used to bridge the gap in understanding between railway traffic operators, timetable planners and disruption mitigation protocol designers.

6.4 Method

Two human-in-the-loop simulator sessions were conducted with operators from the regional control center that covers the simulated geographical area.

6.4.1 Simulator design

For the simulated environment a human-in-the-loop simulator was used, which can be described in accordance with the following gaming simulation design characteristics (see Table 6.1).

Table 6.1: Characteristics of the simulator design.

Core aspect	**Description**
Purpose	Studying the impact of different distributed geographical areas for train traffic controller roles
Scenarios	Four
Simulated world	Detailed infrastructure; detailed timetable; larger area of train traffic operations
# of participants	Four, of which one in an observing role
Roles (#)	Train traffic controller (2), train traffic planner (1), and regional network controller (1) in an observing role
Type of role	Similar to their own function
Objectives	Execution of tasks – similar as to their daily work
Constraints	Exclusion of roles outside the defined infrastructure area, train drivers were role-played by a facilitator
Load	Medium severe disruptions
Situation (external influencing factors)	Presence of individual observers seated next or near the participant
Time model	Continuous

6.4.2 Participants

In total eight operators participated in this session. Two teams each consisting of two train traffic controllers and one train traffic planner, actively participated in the scenario. In each team, one regional network controller participated in an observing role in the session.

Operators with specific knowledge of the investigated geographical areas were selected by the management of the control center to participate in the study. Their participation remained voluntarily however. The role of the regional network controller was included to obtain additional insights. However, as it was not within the scope of this study to investigate a severe disruption that would require an active role of the regional network controller, the regional network controller took upon an observing role.

6.4.3 Materials

The study made use of a convergent mixed methods research design (Creswell & Plano Clark, 2011), in which multiple measurement techniques were applied, such as qualitative data from the debriefing, data from questionnaires after each scenario and from observations during the human-in-the loop simulator session and/or video recordings.

Simulator validity was measured through an existing questionnaire; three items each for structural and process validity and psychological reality (Lo et al., 2016; Raser, 1969; van Lankveld, Sehic, Lo, & Meijer, 2017). Structural validity relates to the extent that physical objects are comparable between the simulated and reference system. Process validity relates to the extent that processes, such as communication, are comparable between the simulated and reference system. Finally, psychological reality relates to the extent to which participants experience the simulated environment as realistic. Examples of items were: "I can apply the information from the information sources in the simulator in a similar way as in the real world" for structural validity, "the train traffic flow in the simulator is similar in its processes to the real-world train traffic flow" for process validity and "the train model is sufficiently realistic for the current task" for psychological reality. Items were measured on a five-point Likert scale, varying from 'strongly disagree' to 'strongly agree'. The selection of scenarios was also discussed regarding their validity. In the debriefing session, the scenarios were reflected in terms of their comparability to participants' actual work environment.

Geographical workspace design was mainly assessed through obtained feedback during the debriefings after each tested option. Examples of questions were: what are the advantages of this specific geographical workspace design? What do you identify as bottlenecks with the simulated geographical workspace design option? An example of one of the geographical workspace design options is depicted in Figure 6.4. One train traffic controller is responsible for Utrecht 'turn' (KEER, in green) and the other for Utrecht 'through' (DOOR, in orange). Both train traffic controllers control parts of the area of Utrecht Central station (UTCS, in blue). The train traffic planner on the other hand is responsible for the planning of both geographical areas until 15 min before operations.

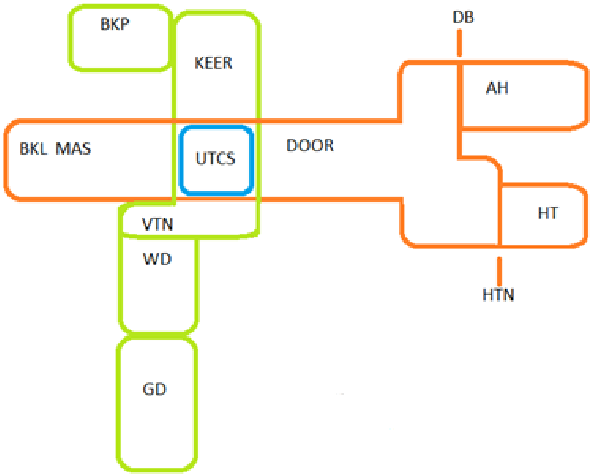


Figure 6.4: Example of an option of the geographical workspace design for the area of Utrecht Central Station.

Team situation awareness as defined and operationalized by Gorman, Cooke and Winner (2006) was measured during the scenarios through observations and video recordings. Questions on their perceived coordination and communications, such as (1) which colleague needs to be informed with the disruption in the given scenario, (2) with which colleague(s) collaboration was/were needed during the scenario, were included in questionnaires and addressed during the debriefing after each scenario. Team situation awareness was measured through analysis and categorization of the information exchange in a number of communication categories, such as statements related to uncertainty, actions, planning and factual statements (Bowers, Jentsch, Salas, & Braun, 1998). Examples of questionnaire items following each scenario were, e.g. to what extent the participant was satisfied with the communication and coordination between colleagues, and to what extent they were satisfied with the collaboration in the scenario. These items were measured on a five-point Likert scale, varying from ‘strongly dissatisfied’ to ‘strongly satisfied’ and an additional answering category ‘not applicable’. Questions for the debriefing were also focused on the development of the collaboration, for instance how operators experienced the collaborations and if operators were able to build up mutual awareness of the situation.

Trust and cohesion as influencing social-psychological factors were exploratively investigated and measured in the closing questionnaire, following its importance in the control room environment (Farrington-Darby & Wilson, 2009). Items measured were: to what extent do you perceive as important (1) trust in skills (e.g. decision, actions) of colleagues, (2) good understanding between colleagues, (3) able to count on colleagues and (4) good ambiance between colleagues. These items were measured on a five-point Likert scale, varying from 'strongly disagree' to 'strongly agree'. In the debriefing session it was also discussed what makes a team successful to understand the broader implications of teams.

6.4.4 Procedure

Each team conducted four scenarios in one day: all scenarios consisted of the current or newly (three variations) defined allocated areas of responsibility. The disruptions introduced were of a medium severe impact, designed such that it would require collaboration between operators. However, the primary design goal of the scenarios was to investigate advantages and bottlenecks in the geographical workspace design options. Each scenario took about 1 hour. Before the start of the session and subsequently after each scenario, questionnaires were provided to the participants. A debriefing was held after each scenario. Video and audio recordings were made during each scenario and during the debriefings. Figure 6.5 depicts the human-in-the-loop simulator set.

Figure 6.5: Impression of the human-in-the-loop simulation session.

6.5 Results

6.5.1 Participants

In total four train traffic controllers participated in the simulation sessions with an average professional experience in their current function of *M* = 11.0 years, *SD* = 9.42. The average professional experience of the two train traffic planners was *M* = 5.0 years, *SD* = 0.00. All train traffic controllers who participated also worked as train traffic planners as well. The professional experience of one regional network controller was noted, in which the participant had 3 years of professional experience in the current function and 24 years of professional experience in total in the railway sector.

6.5.2 Validity of the simulator session

Train traffic controllers and planners assessed the validity of the simulator in which structural validity received an average score of 2.7, *SD* = 0.86. Operators perceived the physical objects of the simulator as not positive, neither negative in its validity. This is in line with the observations and findings from the debriefing; the validity of the simulator was influenced by a number of technical issues and the absence of a number of system functionalities. Process validity received an average score of 3.8, *SD* = 0.42. In terms of psychological reality, operators scored an average of 3.6, *SD* = 0.15. On both these validity components, operators assessed the validity of the simulator as slightly positive.

During the debriefing, participants indicated that scenarios were representative to disruptions in the actual operational environment. However, they indicated that it would be very important to introduce scenarios in other areas than those in the simulated workspace design, especially with the workspace design options that were geographically extended. Also, the inclusion of severe disruptions is preferred, specifically in relation to the role of the regional network controller.

6.5.3 Geographical workspace design

In the debriefing session, train traffic controllers indicated to have experienced the simulation sessions with different geographical workspace design options and scenarios as positive; they obtained tangible insights in the consequences of the different options. Examples of such insights were the experienced workload increase in one of the design options where the train traffic controller gained a larger geographical area to manage. In another design option an additional geographical area was also added. However, operators experienced that this specific area was valuable, as it provided increased flexibility for disruption mitigation actions. The reason for this is that due to the removal of many switches at Utrecht Central station, it is difficult to apply mitigation strategies in this location. Mitigation strategies would be only efficient if they are applied at a longer range on the train corridor towards Utrecht Central station.

6.5.4 Team situation awareness

Due to multiple simulation technical issues, data from the game session and video recordings were deemed as not sufficiently valid to include in the data analysis and conduct quantitative analyses on team situation awareness. As multiple measurement techniques were used, other data could still be used albeit on a higher abstract level in the explorative study.

Operators varied in their satisfaction of the coordination and communication between colleagues from slightly positive (*M* = 3.7, *SD* = 0.52) to positive (*M* = 4.2, *SD* = 0.41) in the four scenarios. Findings from the questionnaire item on the perceived coordination and communication with other colleagues showed that participants predominantly indicated to have collaborated with both train traffic controllers and/or planner. In the debriefing session, operators indicated that collaboration between the three roles is a regular way of working as team, which cannot be compared to collaboration with other colleagues within the regional control center. The reason for this is that operators share the area around Utrecht Central station, which may be divided between the two traffic controllers. However, as they still share the same area information on their displays and systems, they can create a shared situation awareness. As the planner is responsible for managing the train traffic control until 15 min before operations, also having the exact same area information and systems covering both geographical areas of the two train traffic controllers, he/she is also part of the team. Other train traffic controllers are informed on a need to know basis only, having their own geographical areas to manage. The operators all agreed that the regional network controller is not part of this team. However, in case of severe disruptions, close collaboration with the regional network controller is necessary. Data from observations during the game session as well as video recordings show that the three roles closely work with each other, sharing detailed information or consulting on their actions. In line with the observations that information exchange is a key action in their collaboration, operators also suggested that an element of being a more valuable colleague is knowing that one receives support or assistance when the workload would get too high. Another example is that by a think-aloud way of working, others know what one is working on, which is another aspect of being considered a team player.

6.5.5 Trust and cohesion

Findings from the questionnaires indicated that train traffic operators and planners perceive trust in colleagues' skills (*M* = 4.3, *SD* = 0.52) and to be able to count on colleagues (*M* = 4.3, *SD* = 0.52) as most important characteristics for a good collaboration. Also, a good understanding between colleagues (*M* = 4.2, *SD* = 0.41) and a good ambiance (*M* = 4.0, *SD* = 0.89) were perceived as important components in the collaboration with co-workers. The findings from the questionnaires may seem expected and wished for. A richer context of the data was provided through qualitative findings in the debriefing. In the debriefing session, operators pointed out that there are many colleagues and therefore many team configurations, in which some are clearly more successful than others. Teams would be more successful when colleagues can count on each other's qualities and can rely on them, indicating the importance of a specific component of trust. Although there are differences between operators in how they operate, a clear separation in tasks between the three roles is necessary.

6.6 Lessons on balancing between organizational and academic research approaches

In order to achieve a successful outcome, both for the organization and the academic research, several parameters were identified based on the research team's experience. The experiences and lessons learned on conducting research with organizational and academic purposes are described in line with the three components (environment, knowledge base and research) in the earlier discussed framework of Hevner et al., (2004) and are discussed with respect to challenges and obtained insights.

6.6.1 Environment

6.6.1.1 Involving stakeholders

Close collaboration with the project team that uses the research findings for decision-making is important for stakeholder management. Involving the project team in the gaming simulation and research design leads to a better understanding between two parties. Project team members who are especially new with the research method learn more about the possibilities and constraints of gaming simulations, whereas gaming simulation designers and researchers learn more about the issues in the project scope. Experience in the current and previous studies indicated that the relevance of gaming simulation is best explained through five criteria, i.e. complexity, communication, creativity, consensus and commitment, as described by Duke and Geurts (2004). These five criteria can be identified in simulating with participants, in which *complexity* is made more explicit and therefore also more tangible, *creativity* is stimulated when simulations are experienced as enjoyable and productive, and through *communication* between participants a better understanding is achieved which leads to more *consensus* and a higher *commitment*.

In the current study, simulation limitations led to the inclusion of only three train traffic control roles instead of the initial intended seven train traffic control roles (i.e. the entire regional control center). However, retrospectively the three train traffic control roles were a good selection for the scope of this study. For the project and research team, these are considerations that need to be made in close collaboration to develop a shared ownership and success.

Also, a shared understanding and alignment between the project and research team provides a base for flexibility in the formulation of the organization research question and its research design. As a result, the current investigation was able to conduct a research study with an academic focus, next to the primary organizational research study.

Another important stakeholder is the operator. During the gaming simulation design, a number of operators were involved to develop the scenarios and to test the simulation environment. Although two of the facilitators in this study have extensive knowledge
of traffic control operations, knowledge of specific local characteristics is needed to increase validity of the simulation environment. As such, the simulation environment is fine tuned and tested by and for operators.

6.6.1.2 Making the organizational research question explicit

An obvious but possibly challenging factor is extracting the underlying research question that project teams actually require. From experience with the current study, which is also built upon previous human-in-the-loop simulation studies, the formulation of an organizational research question by stakeholders is not necessarily straightforward. For example, the research question being formulated in a general manner with multiple ways of operationalizations. In the current study, the project team expected that the findings will pinpoint one solution, i.e. one geographical design option. However, participants provided additional insights of conditions or solutions that fine tuned each design option, such as reassigning certain geographical areas. In the current case, the project team found the findings valuable and would use them as input to formulate another set of geographical design options to iteratively achieve the optimal design.

6.6.1.3 Adaptations to the research maturity of the organization

The extent to which stakeholders in the organization are used to conduct research can possibly be identified as the most important factor that puts most of the load on the balance between the extent of practical application and academic rigor. This level of ‘research maturity’ holds for both the project team and participants. As mentioned in the previous sections 6.7.1.1 and 6.7.1.2, project teams may need to be informed about the possibilities and constraints of gaming simulation as a tool and the implications of certain choices in the research design, e.g. limiting the number of sessions. On the other hand, participants may need to get used to them being the object of study, in which their decision-making and performance during the simulation and their feedback from the debriefing would be analyzed to answer their organization's research questions. With little experience in participation of studies, more senior operators may be selected by management who are willing and interested to participate. Thus, this selection may be biased in its sample. The extent to which operators are used to participate in studies also determines the research design and its methods. For instance, in the current study it was chosen to keep the questionnaire as short as possible, i.e. to five to ten min, in order to reduce the level of invasiveness of the study for operators. With the increasing familiarity of stakeholders to research studies, it is expected that these limitations will decrease and that a larger pool of operators may volunteer. However, it is expected that resources in terms of number of sessions may only be increased occasionally with more critical issues.

6.6.1.4 Simulating railway traffic operations with human-in-the-loop

The occurrence of technical issues with a human-in-the-loop simulator in development may be unavoidable, as was the case in the current study. However, the ability of operators to (partially) test different geographical design options was considered more valuable than its limitations. The organizational research question was answered due to its design and emphasis on the collected input from operators in debriefing session. With improving technical stability and features in future human-in-the-loop studies, it is expected that more quantitative data can be obtained from the simulator loggings and statistical analyses can be employed, for example on network punctuality. In the context of collecting and analyzing quantitative data, the validity of the simulation environment will especially be of importance. However, obtaining qualitative data from the debriefing session will remain a unique and valuable research approach, in which operators are able to provide input on the fine tuning of designs.

It is planned to conduct larger multi-actor human-in-the-loop simulations to also investigate the impact of severe disruptions. For these, a larger scale of the system needs to be involved that deals with specific operator roles that are involved during disruption management (e.g. Albers et al., 2018).

6.6.2 Knowledge base

6.6.2.1 Matching research topics

Another parameter in the balance between organizational and academic research approaches involves the research design. The challenge there is to align paradigms of the organizational and academic research approach. The organizational research question ‘what is the impact of each of the four options having differently distributed geographical areas for train traffic management?’ can be answered using the behavioral paradigm in the ‘knowledge base’ of Hevner et al., (2004) (see section 2). Similarly, the second research question ‘how does team situation awareness develop at a regional control center’ also follows the behavioral paradigm, in which they both use similar research approaches, designs and measurement techniques.

Despite the simulator issues in this study, the second (academic focused) research question provided interesting initial insights on team situation awareness development. The combination of an organizational research question together with a second, human factors related, research question is not only identified as a good fit, but also valuable in terms of descriptive research on railway traffic management. Further and future research in this field can support the development of operator's cognitive models in the use of a computer simulation context.

6.6.2.2 Mixed methods approach

The current study makes use of a mixed methods approach, in which data is obtained from the debriefing, questionnaires and from observations during the human-in-the loop simulator session and video recordings. The triangulated input from multiple data sources provides overlapping and complementary support to the findings. The questionnaires provide quantitative data, while data from the debriefing and observations provide contextual rich information. Additionally, some of the collected data can be useful to answer both research questions. For instance, the data on the validity of the simulator is useful for both the organizational as the academic research question.

6.6.3 Research

6.6.3.1 Flexibility in the research procedure

Conducting a study in an applied setting limits experimental control that can be exercised compared to a laboratory setting, especially in an organization with a low 'research maturity' (see 6.7.1.3). Consequently, the challenge here is to balance the degree of freedom in the research procedure, in which participants can create a comfortable and safe environment for themselves, while maintaining experimental protocols in the designed study. Thereby, natural behavior of the participants may be induced. However, participants may therefore deviate from the procedures in terms of informal interaction during the study or questions during the debriefing may be asked in a different order or skipped in relation to the situation at hand. Also, based on the input from participants when technical issues occurred, it was decided whether to put more effort in solving an issue, or ending a scenario. The impact on the quality of the debriefing input from operators was assessed in particular (see also 6.7.1.4).

6.6.3.2 Using facilitators with a diverse and overlapping skill set

In order to conduct the current study, facilitators with (technical) knowledge of the human-in-the-loop simulator, gaming simulation design expertise, expertise of railway traffic operations and (human factors) research expertise were required. This study was carried out with three facilitators that had complementary and partially overlapping expertise covering the above knowledge set. During the event of technical issues in the simulator, facilitators were able to use their knowledge and skills to make adaptations to the research procedure. For instance, next to participants responses, two facilitators would discuss if the simulated scenario so far provided sufficient content in terms of events, decisions and performance, that could be used as input for in the debriefing with operators.

It should be noted that the involvement of facilitators in both the development and execution of the research design could be seen as a limitation as there might be a certain level of bias involved that could affect the study. As such, the inclusion of multiple facilitators with overlapping knowledge could also support more objective decisions throughout the investigation.

6.6.3.3 Obtaining academic rigor with practical relevance

Another aspect is to create a balanced research design that often adheres towards organizational constraints and minimal requirements for academic rigor. Ideally, organizations would want to conduct a similar study with as few teams as possible and obtain answers that can be applied to a broad range of questions. Ideally, academic studies would prefer research conditions with, in this case, a large number of diverse teams that conduct the study according to a four (scenarios) by four (geographical workspace design options) research design. This issue strongly relates to the generalization of the results and methodological approach. Maintaining academic rigor while keeping track of the impact of the factors mentioned above, specifically 'research maturity' and 'stakeholder involvement' may be biggest challenge in finding the balance between the two approaches.

6.7 Discussion and conclusion

This article discusses the fine balance between a combined research study: one that has an organizational purpose on geographical workspace design with in parallel another research study that focuses on team situation awareness from an academic interest. Both research questions were investigated simultaneously with a human-in-the-loop simulator and uncovered relevant insights that contribute to the knowledge base of the Dutch railway sector.

The study on the development of team situation awareness is an explorative investigation with an emphasis on qualitative measurements, given the constraints of the data. The findings provide insights in understanding the organizational dynamics through a better understanding of how team situation awareness amongst operators is developed. Although the findings may be obvious in retrospective, the expectation was that the entire regional control center would be identified as a team. As such, in the preparations of the simulation, attempts were made to conduct a gaming simulation session that comprised all train traffic controller areas. Due to restrictions from a simulator technical perspective, the scope was reduced to a number of indicated areas by the management of the regional control center. The findings on the development of team situation awareness took the absent train traffic controlling areas into account. However, the characterizations of the workspace, i.e. shared displays with predominantly identical information and the overlapping geographical area, appeared to be components that strongly bound these three roles together. As such, insights were obtained in the nature of 'sharedness' of situation awareness, in which this could be seen as shared (overlapping information) as opposed to distributed (unique and complementary information for each operator). Another finding is that train traffic controllers are aware of different sorts of team configurations, in which some teams are more successful than others. Trust in colleagues' skills and cohesion of the team are acknowledged variables that influence team collaboration.

In conducting research on the geographical workspace design options, analyses were mainly done with data derived from the debriefing sessions. Debriefing is an intrinsic part of research games, in which operators can make their tacit knowledge explicit after they experienced a future contextual setting. Debriefing is also an efficient research approach to determine the quality of both the reliability and validity, as it opens the black box of the game session (Van den Hoogen, Lo & Meijer, 2016).

In order to find a balance in the scientific rigor with multiple research questions, several parameters needed to be fine tuned. These parameters are on the level of stakeholder management (i.e. project team from the organization), research approach (i.e. operationalization of the research questions and level of generalization of the findings) and research design (i.e. number of participants, measurement tools, procedure and flexibility of the design). Choices made with these parameters often have their own advantages and disadvantages. For instance, the current study is of a qualitative nature, which limits causal analyses. However, the use of specific qualitative measurement techniques such as debriefing also provides the opportunity to discuss topics in-depth, therefore producing knowledge for both the organization and a discipline such as human factors. Particularly challenging to manage is the balance between the maintenance of certain scientific research principles, without becoming too intrusive from the organization's perspective.

This approach combines different team SA methodologies, varying from an ethnographical approach using explorative research methods through the obtained qualitative data in debriefing sessions and observations (e.g. Farrington-Darby and Wilson, 2009), and a traditional controlled human-in-the-loop simulation environment (e.g. Gorman et al., 2006). An additional unique characteristic of the current experimental research design is that both teams conducted simulations with all four geographical workspace design options. If the data in simulator sessions were to be analyzed, the team SA of these teams could also be assessed between all four options, which could provide an indication of the robustness in team SA. Further on, although the two research purposes (organization and academic) have been discussed separately, they can also be connected. In general, academic research often involves a more theoretical oriented emphasis while research studies within organizations include context and relevance. The practical value of having conducted an academic oriented study on team situation awareness is that its results can be incorporated in the design of future gaming simulations for the organization. For instance, in relation to participant selection and/or number of workspaces. For the organization, the findings could be useful in projects such as the design of operator systems or task distribution, especially with the descriptive nature of the current study.

The current study also has limitations, such as technical issues, the limited scope of the scenarios and the limited number of simulator sessions. Although the latter two were conscious choices, they still are weaknesses in the research design. However, with regards to the population of train traffic controllers and planners it can be argued that this population is fairly limited to approximately 22 operators. As such, the current number of operators represents about 25% of the regional control center's train traffic controller and planner population. In terms of addressing the technical issues, future software developments of the simulator should support a more stable simulation environment. Additionally, when new scenarios are developed, a longer testing period should be included in the preparation time to ensure technical stability.

Future research could build on the current qualitative knowledge from this study and focus on more quantitative analyses, in which simulator data (i.e. from the simulation tool FRISO) could be used. Additionally, future studies could look at modelling properties of human cognitive or team processes, which could be applied in the simulation tool to simulate certain operators.

References

Aydogan, R., Lo, J. C., Meijer, S. A. & Jonker, C. M. (2014). Modeling Network Controller Decisions based upon Situation Awareness through Agent-Based Negotiation. In S. Meijer & R. Smeds (Eds.), *Frontiers in Gaming Simulation* (pp. 191-200). Berlin: Springer.

Albers, S., Lo, J., Sehic, E. & Van 't Woudt (2018). Development & use of multi-actor simulation environment for Dutch railways. *Winter Simulation Conference (WSC'18).*

Bowers, C. A., Jentsch, F., Salas, E., & Braun, C. C. (1998). Analyzing communication sequences for team training needs assessment. *Human Factors, 40*(4), 672-679.

Cooke, N. J., Gorman, J. C., Myers, C. W., & Duran, J. L. (2013). Interactive Team Cognition. *Cognitive science, 37*(2), 255-285.

Corman, F., D'Ariano, A., Hansen, I. A., & Pacciarelli, D. (2011). Optimal multi-class rescheduling of railway traffic. *Journal of Rail Transport Planning & Management, 1*(1), 14-24.

Creswell, J. & Plano Clark, V. (2011*). Designing and Conducting Mixed Methods Research*. Thousand Oaks, CA: SAGE Publications Inc.

Croft, D.G., Banbury, S.P., Butler, L.T., & Berry, D.C. (2004). The role of awareness in situation awareness. In S. Banbury & S. Tremblay (Eds.). *A cognitive approach to situation awareness: Theory and application* (pp. 82-103). Aldershot: Ashgate Publishing.

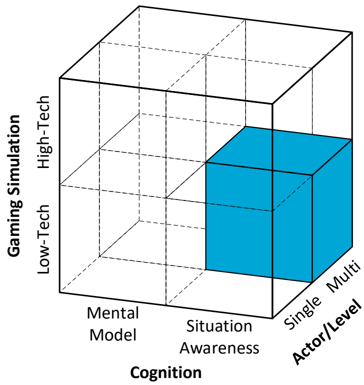
Duke, R. D., & Geurts, J. L. A. (2004). *Policy Games for Strategic Management*. Dutch University Press: Amsterdam.

Endsley, M. R. (1988). Design and Evaluation for Situation Awareness Enhancement. *Proceedings of the 32nd Human Factors Society Annual Meeting*, 97-101.
Endsley, M. R. (1995). Measurement of situation awareness in dynamic systems. *Human Factors, 37*(1), 65-84.
Farrington-Darby, T., & Wilson, J. R. (2009). Understanding social interactions in complex work: a video ethnography. *Cognition, Technology & Work, 11*(1), 1-15.
Forsyth, D. R. (2001). *Group dynamics*. Pacific Grove, CA: Brooks/Cole.
Galapitage, A., Albrecht, A. R., Pudney, P., Vu, X., & Zhou, P. (2018). Optimal real-time junction scheduling for trains with connected driver advice systems. *Journal of Rail Transport Planning & Management, 8*(1), 29-41.
Ghaemi, N., Zilko, A. A., Yan, F., Cats, O., Kurowicka, D., & Goverde, R. M. P. (2018). Impact of railway disruption predictions and rescheduling on passenger delays. *Journal of Rail Transport Planning & Management, 8*(2), 103-122.
Golightly, D., Wilson, J. R., Lowe, E., & Sharples, S. (2010). The role of situation awareness for understanding signaling and control in rail operations. *Theoretical Issues in Ergonomics Science*, 11(1-2), 84-98.
Gorman, J. C., Cooke, N. J., & Winner, J. L. (2006). Measuring Team Situation Awareness in Decentralized Command and Control Environments. *Ergonomics*, 49(12-13), 1312-1325.
Gorman, J. C., Cooke, N. J., Pederson, H. K., Olena, O. C., & DeJoode, J. A. (2005). Coordinated Awareness of Situation by Teams (CAST): Measuring Team Situation Awareness of a Communication Glitch. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 49*(3), 274-277.
Goverde, R. M. P., & Meng, L. (2011). Advanced monitoring and management information of railway operations. *Journal of Rail Transport Planning & Management, 1*(2), 69-79.
Grogan, P. T., & Meijer, S. A. (2017). Gaming Methods in Engineering Systems Research. *Systems Engineering, 20*(6), 542-552.
Hevner, A. R., March, S. T., Park, J., & Ram, S. (2008). Design science in information systems research. *Management Information Systems Quarterly, 28*(1), 75-106.
Kauppi, A., Wikström, J., Sandblad, B., & Andersson, A. W. (2006). Future train traffic control: control by re-planning. *Cognition, Technology & Work, 8*(1), 50-56.
Kortmann, R., & Sehic, E. (2011). The Railway Bridge Game – usability, usefulness, and potential usage for railways management. In M. Beran (Ed.), *Changing the world through meaningful play* (pp. 119 – 125). Spokane, WA.
Lo, J. C., & Meijer, S. A. (2013). Measuring Group Situation Awareness in a Multi-Actor Gaming Simulation: A Pilot Study of Railway and Passenger Traffic Operators. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 57*(1), 177-181.
Lo, J. C., Sehic, E., Brookhuis, K. A., & Meijer, S. A. (2016). Explicit or implicit situation awareness? Measuring the situation awareness of train traffic controllers. *Transportation Research Part F: Traffic Psychology and Behaviour, 43*, 325-338.
Lo, J. C., & Meijer, S. A. (2019). Assessing network cognition in the Dutch railway system: insights into network situation awareness and workload using social network analysis. *Cognition, Technology & Work.*
Meijer, S.A. (2012). Introducing Gaming Simulation in the Dutch Railways. *Procedia - Social and Behavioral Sciences, 48*, 41-51.
Meijer, S. A. (2012). Gaming simulations for railways: Lessons learned from modeling six games for the Dutch infrastructure management. In X. Perpinya (Ed.), *Infrastructure Design, Signaling and Security in Railway*. Croatia: IntechOpen.
Meijer, S. (2015). The power of sponges: Comparing high-tech and low-tech gaming for innovation. *Simulation & Gaming, 46*(5), 512-535.
Middelkoop, A.D. and L. Loeve (2006). Simulation of traffic management with FRISO. *Computers in Railways X: Computer System Design and Operation in the Railway and Other Transit Systems*.
Middelkoop, D., Meijer, S., Steneker, J., Sehic, E., & Mazzarello, M. (2012). Simulation backbone for gaming simulation in railways: A case study. In C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, & A.M. Uhrmacher, (Eds.), *Proceedings of the Winter Simulation Conference (WSC'12)* (pp. 3262-3274). Berlin, German: IEEE.
OnTime (2013). *Functional and technical requirements specification for perturbation management*. Technical report FP7 - SCP0 – GA – 2011 – 265647.

Peters, V., Vissers, G., & Heijne, G. (1998). The validity of games. *Simulation & Gaming, 29*(1), 20-30.
Ramaekers, P, De Wit, T., & Pouwels, M. (2009). *Hoe druk is het nu werkelijk op het Nederlandse spoor? Het Nederlandse spoorgebruik in vergelijking met de rest van de EU-27*. [In Dutch]. Centraal Bureau voor de Statistiek.
Raser, J.C. (1969). *Simulations and society: An exploration of scientific gaming*. Allyn & Bacon, Boston.
Roth, E. M., Multer, J., & Raslear, T. (2006). Shared situation awareness as a contributor to high reliability performance in railroad operations. *Organization Studies, 27*(7), 967-987.
Ryan, T. (2000). The role of simulation gaming in policy-making. *Systems Research and Behavioral Science, 17*(4), 359-364.
Roungas, B., Meijer, S. & Verbraeck, A. (2018). Validity of railway microscopic simulations under the microscope: Two case studies. *International Journal of System of Systems Engineering, 8*(4), 346-364.
Samà, M., Meloni, C., D'Ariano, A., & Corman, F. (2015). A multi-criteria decision support methodology for real-time train scheduling. *Journal of Rail Transport Planning & Management, 5*(3), 146-162.
Stanton, N. A., Salmon, P. M., & Walker, G. H. (2015). Let the reader decide: A paradigm shift for situation awareness in sociotechnical systems. *Journal of Cognitive Engineering and Decision Making, 9*(1), 44-50.
Steenhuisen, B. (2009). *Competing public values: Coping strategies in heavily regulated utility industries*. Enschede: Gildeprint Drukkerijen.
Tschirner, S., Sandblad, B., & Andersson, A. W. (2014). Solutions to the problem of inconsistent plans in railway traffic operation. *Journal of Rail Transport Planning & Management, 4*(4), 87-97.
Van Aken, S., Bešinović, N., & Goverde, R. M. P. (2017). Solving large-scale train timetable adjustment problems under infrastructure maintenance possessions. *Journal of Rail Transport Planning & Management, 7*(3), 141-156.
Van den Hoogen, J., Lo, J., & Meijer, S. (2016). Debriefing research games: Context, substance and method. *Simulation & Gaming, 47*(3), 368–388.
Van de Velde, D. M. (2013). Learning from the Japanese railways: Experience in the Netherlands. *Policy and Society, 32*(2), 143-161.
Van Lankveld, G., Sehic, E., Lo, J. C., & Meijer, S. A. (2017). Assessing Gaming Simulation Validity for Training Traffic Controllers. *Simulation & Gaming, 48*(2), 219-235.
Wilson, J. R., Farrington-Darby, T., Cox, G., Bye, R., & Hockey, G. R. J. (2007). The railway as a socio-technical system: Human factors at the heart of successful rail engineering. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 221*(1), 101-115.
Zhao, J.,Wang, D. & Peng, Q. (2018). Optimizing the Train Dispatcher Desk Districting Problem in High-Speed Railway Network. *Transportation Research Board 97th Annual Meeting*. Washington DC, United States.

7 Assessing network cognition in the Dutch railway system through communication: Insights into network situation awareness and workload using social network analysis

This chapter has been published as:
Lo, J. C., & Meijer, S. A. (2020). Assessing network cognition in the Dutch railway system: insights into network situation awareness and workload using social network analysis. *Cognition, Technology & Work, 22*(1), 57-73.



Abstract
This chapter investigates network situation awareness and workload of railway traffic operators using a group cognition perspective. A table-top simulation environment is used to conduct the study, in which its design principles are elaborated upon. Network cognition is operationalized through communication content and flow and studied through social network analysis (SNA). SNA centrality metrics, such as degree, closeness and betweenness, are assessed in these networks. As part of the study, two cases are compared where operational procedures for disruption mitigation are varied. The dependent variables are the different types of communication network structures that are conceptualized for communication flow and semantic network structures for communication content. Although the quantitative comparisons between the two operational procedures regarding their communication flow and semantic networks showed no significant differences, this study provides a methodology to compare different conditions.

Keywords: Network situation awareness; social network analysis; communication; group cognition; railway traffic operations; workload

7.1 Introduction
When train traffic and network operations become disrupted, railway traffic control is mostly a job for humans, sometimes with the help of decision-support tools. Railway traffic controllers are challenged since the interpretation of a situation implies coupling a large number of often fuzzy indications, of which the consequences are combinatorially explosive. Next to these complex and ill-

defined situations, the work that railway traffic controllers carry out is under high time pressure and with high stakes, often in close collaboration with other operators in different locations (Farrington-Darby, Wilson, Norris, & Clarke, 2006; Funke, 2001).

The application of computer simulation is rather limited when one wishes to make claims regarding the impact of certain innovations, namely changes in the railway system on, e.g., operations in railway traffic control (Van den Hoogen & Meijer, 2014). An important indicator in the assessment of these changes is the implication for the cognition and decision making of railway traffic controllers. The introduced innovations are mostly related to process optimizations to solve railway track capacity issues in a highly dense and space-constrained country such as the Netherlands. As there is a need to test the impact of alternative modes of the system, the Dutch railway infrastructure organization ProRail turns to single-actor human-in-the-loop and multi-actor table-top simulation environments as a platform to test future configurations of the system and to train personnel to work with them. Multi-actor table-top simulation environments are the most commonly used due to their short development time and low development costs and were used in the current study (e.g., Lo, Van den Hoogen, & Meijer, 2013; Meijer, 2012).

Understanding the cognition of operators in complex socio-technical systems is crucial for training, safety, performance management, but also for the design of the system in terms of level of automation and interface design (Farrington-Darby & Wilson 2009; Wilson & Norris 2006). More specifically, cognitive constructs such as situation awareness (SA) and workload can be seen as focal concepts. SA was originally introduced as a predictor of good decision making. It is defined in a broader sense as the ability to see the 'big picture' and to 'know what is going on', which is a result of an individual's cognitive processes (Endsley, 1988, 1995; Tenney & Pew, 2006). As with other complex control tasks such as in air traffic control, the development and maintenance of SA in railway traffic operations are crucial for operators (Farrington-Darby et al., 2006; Golightly, Wilson, Lowe, & Sharples, 2010). In contrast to studies on SA, however, in the railway sector a stronger focus has been put on the role of workload and its understanding, due to its strong link with safety and performance (Pickup, Wilson, Sharpies, Norris, Clarke, & Young, 2005; Young, Brookhuis, Wickens, & Hancock, 2015).

Research on SA has also been facing challenges in finding a convergent measurement approach, particularly when it comes to analysis on a team or network level. These challenges are particularly caused by fundamentally different theoretical and methodological approaches for individual and team cognition (Cooke, Gorman, & Kiekel, 2008). As such, it has been argued that cognitive structures, such as SA in a group setting, do not necessarily reside solely in the individual, but rather as a whole in the team (e.g., team or group cognition) or additionally including non-human artefacts in the system (e.g., distributed situation awareness following distributed cognition) (Endsley, 1995; Cooke & Gorman, 2006; Letsky & Warner, 2008; Salmon, Stanton, Walker, & Jenkins, 2009; Stanton, Salmon, Walker, & Jenkins, 2010). It has also been stated that interactions between operators in a network are more relevant than for instance the maintenance of an operator's situation awareness (Salmon et al., 2009). Therefore, team process behaviors as communication and coordination are identified as possible indicators of cognition within the team or system (e.g., Cooke & Gorman, 2009; Letsky & Warner 2008). As such, following macrocognition as theoretical paradigm, cognition should be measured at its respective unit of analyses. This implies that beyond the individual level itself, cognition resides on higher abstraction levels. For instance, communication can be an indicator of cognition on a team level (interrelations between co-workers) or on a network level (interrelations of a set of teams). When analyzing on a system level, non-human artefacts such as automation or decision support tools should also be considered as cognition on this level. In the current study, multiple dyadic teams and human interactions, not including artefacts, will be investigated, taking upon a team/group cognition theoretical stream and focusing on insights at a network level.

Although there have so far been only limited studies on network SA using social network analysis (SNA), the potential of SNA has been identified as a tool to study the network situation awareness by analyzing patterns of communication or content flow between actors within the system (Foltz & Martin, 2008; Houghton et al., 2006; Sorensen & Stanton, 2011; Stanton et al., 2006; Weil et al., 2008). Through SNA metrics such as 'centrality' or 'closeness', positions of individuals in a communication network can be analyzed. This can be achieved by identifying an individual's central position in the network based on the number of communication exchanges with other individuals in quantitative terms or in qualitative terms using the graphical representation of the network. For instance, with measures of 'centrality', certain individuals can, therefore, be pinpointed as key figures in a network of individuals, who maintains contact with many individuals in that network. As such, these findings provide insights into the interaction, the performance of teams and organizations and ultimately situation awareness (Houghton et al., 2006; Weil et al., 2008).

The present study utilized table-top or paper-based simulation environments in which the emphasis is on the exploration of the socio-cognitive dynamics of, for example, the network situation awareness of a team of professionals within a part of the Dutch railway system. These multi-actor table-top simulations are predominantly low-tech in the sense that they make use of analogue materials to represent components of the systems and they can also be found in emergency services simulations (e.g., Houghton et al., 2006). Operators often perform their own role in these simulated environments. Additionally, given the purpose of the table-top simulation to test different types of procedures in a subset of the railway system, the aim of the study was threefold: (1) to explore the cognition of the current railway network through different communication flow and content network structures, namely as indicators of network situation awareness and workload, (2) to explore the analysis of comparing two types of procedures through quantitative measures using SNA as a method, and (3) to provide insights into the use and design of table-top simulations.

7.2 Railway traffic and passenger traffic control in the Netherlands

Worldwide reforms in the governance of, amongst others, the railway sector took place during the 1990s with different implementations in terms of structure (horizontally and/ or vertically separated) and ownership (franchises, government, private, etc.) (Owens 2004). This diversity is not only reflected in the organization of train traffic operations, but also in the automation of control and interaction with control (Golightly et al., 2013). For instance, in Great Britain automatic route setting (ARS) is used widely, however not entirely across the country. As a result, the role of a train traffic controller in these areas is divided between a dispatcher and a signaller. More differences in railway traffic control characteristics exist between countries, but remain limited (e.g., Golightly et al., 2013; Schipper & Gerrits, 2018).

In 1995, the Dutch government de-bundled the national railways into rail infrastructure management – ProRail – and Dutch Railways (Nederlandse Spoorwegen; NS), the principle train service (Algemene Rekenkamer, 2012). Since then, ProRail has been a separate organization focusing on the rail network's governance (e.g., extension, maintenance, safety and capacity allocation). As such, one of its tasks is train traffic control. Railway traffic operations have a decentralized command and control structure as the hierarchical structure is informal, i.e., there is a distributed responsibility. The three main functions within the Dutch railway traffic control are (see also Figure 7.1) (e.g. Aydoğan, Lo, Meijer, & Jonker, 2014):

- Train traffic controller (TTC): Based at a regional control center and responsible for a sub-region. A TTC ensures the availability and safety of the infrastructure capacity in the current situation. In the Netherlands, the roles of signaller and train dispatcher are combined into the role of TTC. A TTC uses a traffic management system (TMS), which means that in normal conditions the train traffic flow is regulated automatically according to the planned time table, and the TTC only needs to monitor for deviations. The number of TTCs per regional control center depends on the size and complexity of the regional area.
- Regional network controller (RNC): Responsible for optimizing and managing train activities at a regional level through planning and coordination. The contact between an RNC and TTCs is, therefore, mostly related to ad hoc changes to the time table or train traffic flow (e.g., order requests) or disruptions in the railway network. In more

complex regional areas, two RNCs may be present to share the workload. Both TTCs and RNCs operate from the same operational control room in a regional control center. There are currently 13 regional control centers in the Netherlands.

- National network controller (NNC): Responsible for optimizing and managing train activities at a national level through planning and coordination. An NNC coordinates activities between RNCs in the case of failures, incidents and emergencies, or ad hoc requests from the railway network. An NNC also handles all long-distance trains within and beyond the country's borders. Two NNCs operate from the Operational Control Center Rail (OCCR), where they report to a directing NNC, who is the contact point for passenger and freight traffic operators. These parties all operate from the same operational control room. Although the organizational structure of railway traffic operations may seem hierarchical, these parties do not formally report to each other (Werkwijze Verkeersleider, 2012).

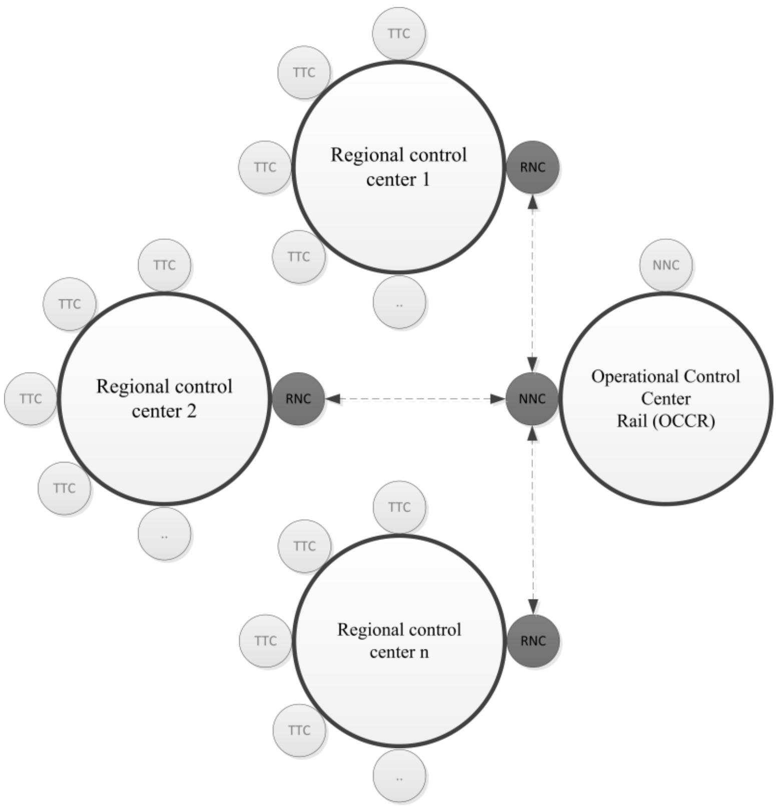


Figure 7.1: Illustration of three main roles within the Dutch railway traffic organization.

Other operational roles within ProRail are the operators that coordinate with emergency services in the control room (back office, BO) and the emergency coordinator (EC). The BO is in contact with the EC, who will be present at the physical location of the emergency, for instance in the case of a collision.

However, disruption management within the railways also implies the involvement of other parties due to the dispersion of responsibilities within the railway system. As such, Dutch Railways is responsible for the rolling stock and crew management. Previous research has investigated the consequences of debundling the railway sector, which resulted in, among others, an 'archipelago' of operators (Van den Top & Steenhuisen, 2009). In its current form, traffic control in the railway sector can be characterized in terms of multi-agency coordination (Salmon, Stanton, Jenkins, & Walker, 2011). Due to the historical ties between the two organizations, the organizational structure of Dutch Railways' passenger traffic control resembles that of the described roles within railway traffic control; that is, the TTC coordinates with the regional passenger traffic junction coordinator (RPTJC). The contact for the RNC is the regional passenger traffic monitor (RPTM) and the NNC coordinates with the national passenger traffic controller (NPTC). Similarly, the first two operators work in the same regional control center, of which there are five (NS, 2014). Both the NPTC and the NNC operate from the Operational Control Center Rail. The regional passenger traffic material and passenger coordinator (RPTMPC) is an additional role that coordinates with train drivers (TD) and coordinates activities related to the availability of rolling stock during disruptions. The passenger information dispatcher (PID) is part of the passenger traffic organization; however, he or she is co-located with the train traffic controllers to understand and timely inform about the situation and consequences for passengers.

7.3 Macrocognition in socio-technical work environments

The importance of studies on teams has become more pronounced as team-based systems are increasingly implemented (Cooke, Salas, Kiekel, & Bell, 2004). In addition, because work environments have become more complex and dynamic, teams more often work in virtual or geographically distributed environments and are more reliant on technology, and team compositions are more heterogeneous (Tannenbaum, Mathieu, Salas, & Cohen, 2012). Especially in decentralized command and control structures where team process behaviors, such as communication and coordination, are crucial for team performance, teams often consist of heterogeneous skilled operators (Gorman, Cooke, & Winner, 2006). As such, the traditional information-processing approach that uses aggregation methods on internalized (individual) knowledge to capture the cognitive structures and processes within teams may be seen as too simple (Cooke, Salas, Kiekel, & Bell, 2004; Cooke, Gorman, Myers, & Duran, 2013; Millot, 2015). Instead, measurement of mental processes at the team or higher levels of analysis can be linked to externalized representations, i.e., observable actions. This theoretical stream stresses the existence of macrocognitive processes, in which a number of theories are connected, such as distributed cognition, activity theory, and group cognition (Letsky & Warner, 2008). These theories differ from one another in the role that cognitive functions and knowledge play, e.g., distributed cognition states that knowledge can reside in non-human artefacts vs. knowledge as a holistic entity in a group of humans in accordance with group cognition.

Similar to group cognition theory, team cognition theory recognizes that cognition is more than the sum of individuals, and, therefore, should be measured and studied at the team level (Cooke et al., 2013). In essence, these theories share the same theoretical paradigm and, however, are used interchangeably by researchers (e.g., Letsky & Warner, 2008). For consistency reasons, the term group cognition is predominantly used throughout the article. Consequently, depending on how cognition in teams is approached, different measurement techniques can be applied (Wildman, Salas, & Scott, 2013). The group cognition perspective emphasizes that team cognition is team interaction and should, therefore, be directly measureable through dynamic communication and patterns in coordination (Cooke et al., 2004, 2008, 2013; Letsky & Warner, 2008; Stahl, 2006). Using communication analysis as a reflection of team cognition can be compared to using verbal protocol or think-aloud procedures to derive knowledge from individuals (Cooke et al., 2004; Cooke & Gorman, 2009). Communication can also be seen as an important indicator of team behavior that affects the development of team SA (Salas, Prince, Baker, & Shrestha, 1995). From a macrocognitive approach, team SA can be measured by observing the coordinated response of the team to a situational change, organization SA through network representations in terms of semantics and flow analysis, and system SA through the analysis of propositional networks (Gorman et al., 2006; Stanton et al., 2006; Weil et al., 2008).

As for the unit of analysis for railway traffic operations, railway traffic and passenger traffic control consist of many small, often dyadic teams. Because the table-top simulation environments involve many of these small teams, the level of analysis was conducted at the network level, thus investigating network cognition using a group cognition perspective.

7.3.1 Communication analysis

The operationalization of communication has resulted in a number of types in terms of communication content (what is said), communication flow (who talks to whom) and communication manner (how it is said in a verbal and non-verbal manner), in which the first two types have been primarily investigated through static or sequential communication aspects (Cooke & Gorman, 2009; Cooke et al., 2008). An example of static communication flow is the total amount of time that person A talks to person B, whereas a sequential communication flow can be illustrated by the number of times that person A talks, followed by person B.

Static communication content can be exemplified by the number of arguments, whereas sequential communication content would be the number of arguments followed by insults. In the current study, the focus is on the use of static characteristics of communications, through the use of SNA as a tool that can provide quantitative measures for the communication network holistically (e.g., Houghton et al., 2006; McMaster, Baber, & Houghton, 2005).

7.3.1.1 Communication flow

Although the analysis of the communication flow seems less rich compared to the analysis of communication content, there are preliminary indications that the former is as promising as the latter (Cooke & Gorman, 2009). Methods of analyzing communication flow are: dominance (speech quantity among team members), flow quantity (amount of speech to and from team members), flow sequence (sequential patterns of speech), stability (variations in speech quantity) and flow as a team process surrogate (an estimation of team process behavior through communication data) (Cooke, Gorman Kiekel, Foltz, & Martin, 2005; Kiekel, Gorman, & Cooke, 2004).

Haythornthwaite (1996) posits SNA as an approach and technique to investigate exchanges between actors (e.g., individuals, groups or organizations) regarding resources. Resources can be understood in terms of both tangible matters, such as money or services, and information. Thus, the exchange of patterns of information could be revealed as a social network, in which actors represent the nodes and the ties that connect the nodes represent information exchange. The connection between communicators in the network can be assessed in terms of direction (directed vs. undirected flow of information) and strength (e.g., frequency and duration of the contact) between nodes. A number of studies used the flow of information to study different actor-to-actor network structures and their performance using SNA, since this can also provide insights into the division of labour in a network (Baber, Stanton, Atkinson, McMaster, & Houghton, 2013; Houghton et al., 2006; McMaster et al., 2005; Weil et al., 2008). Herein, communication flow is captured through the communication between actors. Investigations on the communication flow in emergency service operations with SNA used social network metrics, such as degree and closeness as centrality measures (e.g., Houghton et al., 2006). Centrality measures can be related to, for example, the extent to which a certain operator contributes to the flow of communications.

7.3.1.2 Communication content

One way to analyze communication content is to use latent semantic analysis (LSA), in which indications have been found a strong relation with performance-based scores (Cooke & Gorman, 2009). Alternative methods that can be used for communication content evaluation include both word counts and keyword indexing (KWI). Word count looks into, e.g., the average number of words in transcripts or per utterance and is correlated with LSA vector length, whereas keyword indexing uses mathematical approaches to, for instance, compute vector lengths and distances between utterances in a transcript (Cooke et al., 2005). Following this method, indications for group cognition could be found in conceptualizations, such as the mean of the similarity matrix based on all utterances and similarities between subsequent utterances.

Another way to conceptualize communication content is to create concept maps that capture a network structure of the task knowledge holistically (Cooke et al., 2004). A number of studies apply network or social network analysis to the assessment of concept maps or propositional networks and semantic networks (or knowledge-to-knowledge networks), in which communication transcripts might be used (e.g., Weil et al., 2008; Sorensen & Stanton, 2011; 2012). The difference between the use of propositional networks and semantic networks is that the former entail propositions in terms of a basic statement and links between nodes are labelled (Salmon et al., 2009). The use of SNA on this type of network has shown to be a sensitive measure to assess distributed situation awareness, that is SA on a system level, by identifying differences between two scenarios (Sorensen & Stanton, 2011).

In the present study, the focus is on the selection of various SNA metrics to provide an in-depth analysis of the communication flow and communication content between railway and passenger traffic operators in two scenarios with different but equally severe disruptions. The emphasis also lays on analyzing group cognition as part of interactions between humans; i.e., through their communication and information exchange.

7.4 Design of the table-top simulation environment

Many of the studies that investigate the workload, situation awareness and decision making of operators in complex socio-technical systems have been heavily researched in highly realistic settings, such as human-in-the-loop simulators, or in naturalistic environments (e.g., Hauland, 2008; Klein, 2008; Mogford, 1997). The notion that close-to-real environments provide the ability to portray the naturalistic behavior of individuals has been a strong driving force for the development of simulators (Caro, 1973). However, human-in-the-loop simulators are often accompanied by high development costs. On the contrary, the development of table-top simulation environments is in general rather rapid and low cost. However, designing table-top simulation environments as an alternative to close-to-real simulators is no trivial path. This section provides a description of the design of a table-top simulation environment, as the design of the simulation environment is usually not that elaborately touched upon in studies (e.g., Houghton et al., 2006).

The focus in the present study was on designing a table-top simulation in which parts of the system would be changed and then tested with human operators.

The chal- lenge of using this type of simulation environment for research purposes lies in obtaining a high degree of struc- tural and process validity, such that participants experience the simulated environment as their normal work environment; that is, obtaining a psychological reality (Raser, 1969). Provided the fulfilment of these three validity types, a high predictive validity can be assumed. The use of indexical and symbolic simulation principles may capture the essence of the actual work environment such that participants experience a high psychological reality and the external validity of the simulation outcomes can be ensured (Dormans, 2011). Indexical simulation refers to the degree of the causal relation between rules of the simulated and the actual work environment. Symbolic simulation refers to the resembling mechanisms of the actual work environment in the simulation environment. Also non-tangible elements, such as (organizational) culture, can be captured in these types of simulated environments (Duke & Geurts, 2004; Meijer, Hofstede, Beers, & Omta, 2006). Subsequently, a number of practical guidelines for its development are followed:

- Identification of the purpose of the simulation - which parameters of the system, (e.g., infrastructure, roles, procedures) should be changed.
- Assessment of the impact of the changed parameters on the railway system –which part of the railway system should be included in the simulation and which operators are responsible for these parts of the system.
- Selection of scenario – which conditions can be identified to fulfil the requirements of the research question on testing changes in the system.
- Identification of the information needs of operators – what information do they need to build their situation awareness. The goal-directed task analysis (GDTA) is a type of cognitive task analysis that is specifically designed to uncover situation awareness requirements (Endsley, Bolté, & Jones, 2003). This technique maps operators' goals, their related decisions and their information needs. Therefore, this technique may help in identifying necessary information requirements related to the scenario.

For the table-top simulation, the following choices were made in collaboration with subject matter experts (SMEs) (see Table 7.1). About 8 weeks were needed between the initial meeting and the session to design and prepare the table-top simulation. The design of the simulated environment (e.g., setup of the room layout, scenarios, etc.) was intended to be as much representative to the actual work environment as possible in order to maximize its validity.

Table 7.1: Design aspects of the table-top simulation environment, slightly adapted from Lo & Meijer (2013).

Core Aspect	**Description**
Purpose	To study the impact of current and alternative procedures for the improvement of the speed and realization of railway infrastructure disruption mitigation
Scenarios	Two: 1. current procedure, 2. alternative procedure. The scenarios took place during peak hours and lasted 45 minutes
Simulated world	Railway system between Amsterdam Central Station and Alkmaar Station. Representation of train traffic flow on A0 foam board with schematic representation of the infrastructure, representation of train through pegs with information about train number and length of delay, automatic route setting simulated through facilitators. Train delays and status on national-wide corridors logged in a developed computer program. Timetable information provided on A4 sheets, Simulation of co-location by room separation
# of participants	12, excluding facilitator roles
Roles (#)	Train traffic controller (TTC) (4), regional network controller (RNC) (1), national network controller (NNC) (1), regional passenger traffic monitor (RPTM) (1), regional passenger traffic junction coordinator (RPTJC) (1), regional passenger traffic material and passenger coordinator (RPTMPC) (1), national passenger traffic controller (NPTC) (1), passenger information dispatcher (PID) (2). Facilitators took upon the roles of: train drivers (TD) responsible for passenger trains, train drivers responsible for shunting train, emergency coordinator (EC) and the back-office (BO)
Type of role	Similar or equal to their own roles
Objectives	Execution of tasks – same as in their daily work, only in scenario 2 with new procedures
Constraints	Inclusion of two regional traffic centers, exclusion of roles outside the defined infrastructure area, exclusion of train driver
Load	Two sequential medium impact disruptions; 1. train malfunction, 2. gas leak in a tunnel. These types of disruptions can be categorized as low to average in terms of frequency. Also, both disruptions may be interpreted within the same order magnitude / class of impact
Situation (external influencing factors)	Presence of individual observers seated next to or near the participant, facilitators, occasional attendance of observers from both railway organization
Time model	Continuous

Figure 7.2 shows the setup of the table-top simulation. It included four control centers comprising two regional control centers of the railway traffic organization, one regional control center of the passenger traffic organization and one national control center (OCCR). As also described in Table 7.1, automation of the train traffic flow was represented by facilitators, who moved the trains. The trains were represented by pegs bearing information about train number and length of delay. Operators received all the necessary information.

Some was translated into shared information displays on laptops, for example delays of trains on long-distance routes for the RNC and NNC, and logged communication on the status and details of the disruption. Operators interacted with the traffic flow by providing orders to the facilitators to for instance, hold, turn and/or cancel trains. Facilitators would dynamically adapt the status of the train traffic flow and of a single train by moving trains each minute and adding the amount of delay in relevant cases. As such, crucial functions of the train traffic management system could be translated and supported in an analogue simulated environment. For an elaborate description of the representation of various railway table-top simulations, see Meijer (2015).

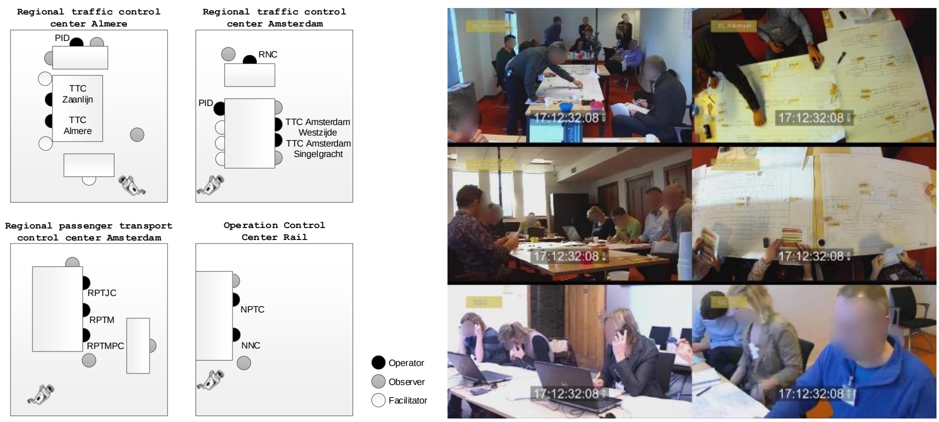


Figure 7.2: Left: setup of the simulation environment. Right: camera shots of the four control centers.

7.5 Method

7.5.1 Participants

Six operators from the railway infrastructure organization ProRail and six operators from the passenger traffic service organization participated in the study.

7.5.2 Materials

As the disruptions in the two scenarios were designed to be as much as similar as possible in severity and consequences for operational processes in the train traffic flow, the independent variable in this study was the type of procedure, namely the current procedure for dealing with disruption (scenario 1) and the alternative procedure for tackling the disruption (scenario 2). In essence, the alternative procedure for tackling the disruption differed in that there would be (1) a predefined protocol for the disruption management, (2) stronger emphasis on the operational process of isolating the disrupted area, (3) faster availability of the predefined disruption protocol and (4) general applicability of the predefined protocol on the infrastructure, rolling stock and personnel.

The dependent variables for communication flow were the different communication networks that are conceptualized through the:

- Undirected communication flow: total frequency of communication between operators.
- Directed communication flow: frequency per node of who contacted whom.
- Directed flow of failed communication attempts: frequency per node of failed contact. Failure in the ability to initiate communication contact is likely to be a result of a high communication load and thus can be linked to workload (Gregoriades & Sutcliffe 2006). This is measured through unanswered phone calls, which due to a busy line or the operator ignoring the call.
- Undirected average length of conversation: total duration of communication in seconds.

Another dependent variable was the communication content, which provides insights into network knowledge, represented by semantic networks that are created on the basis of transcribed communications between operators. Text files were created for a single operator on the basis of their verbal expressions. Firstly, the files were imported in AutoMap 3.0.10.36 and pre-processed with filters, a constructed generalization thesaurus and a deleted words list with prepositions, determiners, etc. (e.g., Freeman, Weil, & Hess, 2006; Weil et al., 2008). Only relevant concepts based on a form of scree plot were selected (e.g., Walker et al., 2010a; 2010b). A subject matter expert assessed the concepts for their relevance. Since the scenarios involved one large event as opposed to multiple large events, network situation awareness was qualitatively assessed by investigating connected concepts in the semantic network as a whole (Weil et al., 2008). The assumption for this approach is that the coordination stage after the disruption does not include a major event that affects a change in the situation awareness of the network.

Both communication flow and content were drawn from communication logs that were created on the basis of the video footage, in which verbal communication via telephone and within control centers was transcribed and coded. Multiple individuals in a co-located room were coded as recipients when an individual in that room was not explicitly addressed by his/her name or function, in line with the official communication protocol. Communication between participants and facilitators who performed multiple other roles were also included in the communication log files.

As the study focused on an in-depth analysis of the communication flow and content through the use of SNA, a number of frequently used centrality metrics were analyzed (Haythornthwaite, 1996). For this, the software program UCINET 6 was used, in which normalized calculations were reported on:

- Degree (Deg): the number of nodes that are connected to one specific node. For example, the amount of communication between one actor and all other actors in the network; in other words, which operator has the most contact with other operators in the railway network. For the directed connections, the degree in terms of 'inbound' and 'outbound' was used, in order to differentiate the initiating actor of the conversation from the receiving actor. For communication content, a high degree centrality implies a highly linked concept in the semantic network. This metric is comparable to 'sociometric status' (Houghton et al., 2006; Sorensen & Stanton, 2011)
- Closeness (Clo): the shortest path of communication between an actor and all other actors in the network, i.e., in how many steps information is transferred from one operator to another. Closeness centrality was calculated for the undirected communication flow. This metric is also comparable to 'Bavelas Leavitt centrality' (Houghton et al., 2006)
- Betweenness (Betw): the position of an actor between other actors in the network. Calculations of betweenness in an undirected communication flow provide insights into the structure of the communication network, in which the position of an actor is an indicator for the power an operator has over the flow of information.

7.5.3 Scenarios

Both scenarios were designed to take place during peak hours in the afternoon, starting at 16.40 and 16.25, respectively, for scenario 1 and scenario 2. In scenario 1, the Zaanlijn train traffic controller received a call at 16.48 from a train driver regarding engine problems. After 3 min, the driver confirmed the malfunction and reported that smoke was issuing from the engine. He advised that due to this, a number of tracks should be cleared and made unavailable at Uitgeest station.

In scenario 2, more time was allocated before the disruption was introduced, in case the operators needed to familiarize themselves with the newly introduced procedure. The TTC Zaanlijn received a call at around 17.02 from a train driver who reported smelling gas in the train tunnel. All train traffic was, therefore, put on hold until further notice.

7.5.4 Procedure

The simulation sessions were held on the same day and both were introduced and debriefed in plenary sessions. Prior to the second session, an in-depth explanation was provided of the similarities and differences between the old and the proposed disruption mitigation procedure. During the simulation sessions, video recordings were made of each control center and observers were present near participants, who were occasionally asked about their decisions or actions.

7.6 Results

Two passenger information dispatchers were excluded from the analysis as their role in the simulation environment was solely to investigate how the two procedures affected their work, which in this case was limited to that of an observer. Additionally, four roles – namely passenger train driver (TD passenger), driver for shunting trains at stations (TD shunting), one back office coordinator (BO) and an emergency coordinator (EC) – were performed by two facilitators and included in the analysis. The average work experience of the six railway traffic controllers and the four passenger traffic controllers was, respectively, 20.2 years, *SD* = 11.38, and 10.7 years, *SD* = 9.43.

7.6.1 Scenario 1

7.6.1.1 Communication flow

Figure 7.3 illustrates the communication network of railway and passenger traffic operators. The nodes are ordered in such a way to easily visualize the informal hierarchical structure in terms of operational levels, that is, operational duties in the field (train drivers, emergency coordinator), followed by control room operations at a local (TTC and RPTJC), regional (RNC and RPTM) or national level (NNC, NPTC and back office). The different node colors represent the different control rooms. The values in between the nodes indicate the undirected communication flow in terms of frequency. The values in brackets represent the undirected failed communication attempts.

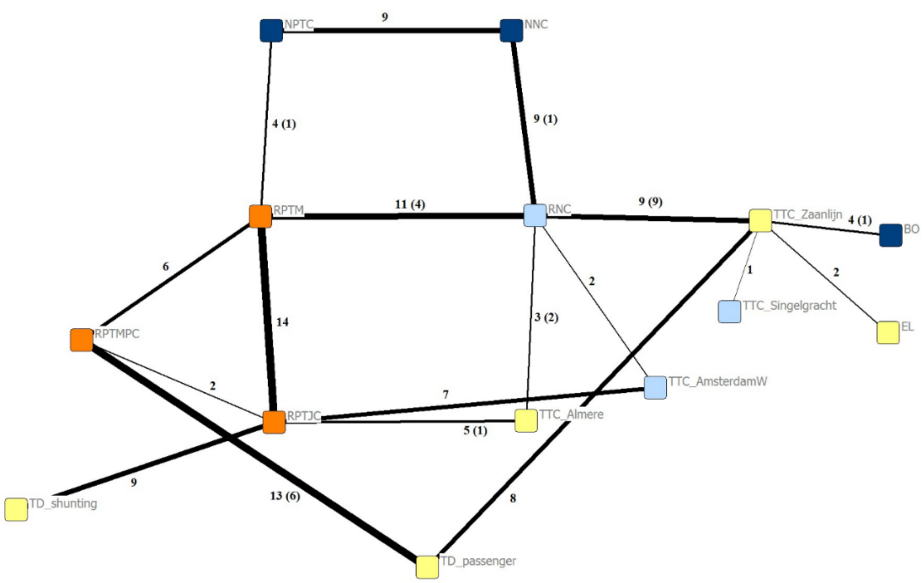


Figure 7.3: Undirected communication network in scenario 1.

The values highlighted in bold in Table 7.2 indicate central actors with regard to a specific centrality metric. The results for the undirected communication flow show that the RNC, RPTM and RPTJC are mostly in contact with different operators within the network and, therefore, have a high degree centrality. However, the RNC is the most central in how many steps within the network he or she needs to take to reach other operators in the network, i.e., is able to most efficiently obtain information (closeness centrality). Further on, the betweenness score shows that the RNC and TTC Zaanlijn are key actors in passing on information.

Regarding the directed communication flow, the RNC and the RPTJC are most central in initiating and receiving conversations within the communication network, as is the RPTM in contacting other actors. The RPTM, RPTJC and RPTMPC show relatively high centrality scores in this network, which might be explained by the collocation of these actors as underlined by the video recordings. It can also be noted from the values in the directed communication flow that the facilitators have a less active role in the communication network.

Further on, the findings on the directed flow of failed communication attempts show that the RNC and TTC Zaanlijn have high scores on the degree centrality in their incoming and outgoing communication network. Video recording observations and communication logs explain failed communication attempts, as other operators were in another telephone conversation or on some occasions were too busy to answer their phone. In some cases, an operator tried to reach the unresponsive operator until he or she was reached, but in the meantime also continued with their work. By looking at failed or unresponsive calls and linking this to the communication network, possible bottlenecks can be identified. This also can be seen as an indicator of an increased workload and inefficiency in the task work.

Finally, the results for the undirected average length of conversation indicate that the TTC Zaanlijn and RPTJC on average have longer conversations compared to other operators in the network.

Thus, the findings from the four communication networks provide unique insights into which operator plays a key role. More specifically, the RNC and TTC Zaanlijn are, given the current circumstances, the overall key operators: they act as gatekeepers for exchanging information, but are also potential bottlenecks as they are not always reachable. To illustrate the latter inefficiency, which may affect operators' workload, the relative value between failed communication attempts in relation to the actual conversations could provide an indicator of the workload of an operator, that is, 37% of incoming and 6% of outgoing communications attempts for the RNC fail, respectively to 25% of incoming and 33% of outgoing communications attempts for the TTC Zaanlijn. The RPTMPC is not included in this assessment as the train driver's communication overload can be explained by one facilitator performing three roles (also as train driver for shunting and as the EC).

Table 7.2: Centrality values for each communication network in scenario 1.

Role	**Undirected communication flow**			**Directed communication flow**		**Directed flow of failed communication attempts**		**Undirected average length conversations**
	Deg	**Clo**	**Betw**	**InDeg**	**OutDeg**	**InDeg**	**OutDeg**	**Deg**
NNC	9.9	70.1	4.1	4.5	7.1	0	1.3	5.1
RNC	**18.7***	**86.2**	**45.3**	**10.9**	**10.9**	**12.8**	**7.7**	16.1
TTC Zaanlijn	13.2	81.5	**46.2**	7.7	7.7	5.1	**7.7**	**21.1**
TTC Almere	4.4	75.4	4.1	2.6	2.6	1.3	2.6	10.3
TTC Singelgracht	.5	63.1	0	0.6	0	0	0	3.5
TTC AmsterdamW	4.9	75.4	4.1	4.5	1.3	0	0	8.7
NPTC	7.1	66.1	1.9	5.1	1.3	0	1.3	5.9
RPTM	**19.2**	81.5	17.3	5.8	**14.7**	2.6	3.8	12.3
RPTJC	**20.3**	75.4	22.0	**10.9**	**12.8**	1.3	0	**20.1**
RPTMPC	11.5	78.5	9.0	5.1	8.3	0	**7.7**	7.7
TD shunting	4.9	56.9	0	5.8	0	0	0	2.4
TD passenger	11.5	75.4	7.7	8.3	0	7.7	0	6.6
BO	2.2	63.1	0	1.9	0.6	1.3	0	6.0
EC	1.1	63.1	0	0	1.3	0	0	4.2

*Values in bold indicate a high centrality score

7.6.1.2 Communication content

Analysis of the communication transcripts resulted in the semantic network depicted in Figure 7.4.

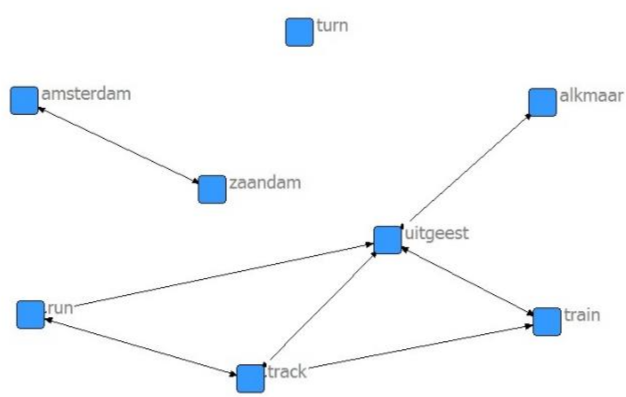


Figure 7.4: Semantic network of key concepts in scenario 1.

It is notable that the key concepts addressed relate to larger train stations in the area that is affected, rolling stock and its status, track availability and turning possibilities. Quantitative results for the degree centrality strength of nodes in the network are also depicted in Table 7.3. As expected from the visual representation of the network, the conversations between operators mainly focused on Uitgeest station.

Table 7.3: Degree centrality values for the semantic network in scenario 1.

Concept	**Deg**
uitgeest	57.143
track	42.857
train	28.571
run	28.571
amsterdam	14.286
zaandam	14.286
alkmaar	14.286
turn	0.000

To provide a qualitative approach to facilitate the understanding of the constructed semantic network, the transcripts were assessed to relate the entire coordination activity to concepts in the semantic network, therefore providing insights into network situation awareness. In the current network, three groups of connected nodes or clusters can be identified. It was observed that at the start of the disruption, six calls were needed throughout the network to inform all operators of the disruption. Operators then focused on the consequences of the train malfunction at Uitgeest station, by adapting the train traffic flow to fit with the reduced infrastructure capacity and available rolling stock and crew in the changed conditions, that is, mainly between Amsterdam and Zaandam. The portion of rolling stock that could not be allocated to a station track or shunted to a yard is the main challenge that operators have to deal with. A third cluster that can be identified possibly related to the turning of rolling stock between Amsterdam and Zaandam. As the concepts are not related, however, it might indicate that the operators did not explicitly mention the stations in their communication. It is notable from the qualitative assessment of the transcripts that except for operators at the national control center (NNC and NPTC), operators shared highly detailed information regarding, for example, newly assigned numbers of rolling stock, the availability of tracks and the location of certain rolling stock.

7.6.2 Scenario 2

7.6.2.1 Communication flow

The communication network structure for scenario 2 is shown in Figure 7.5.

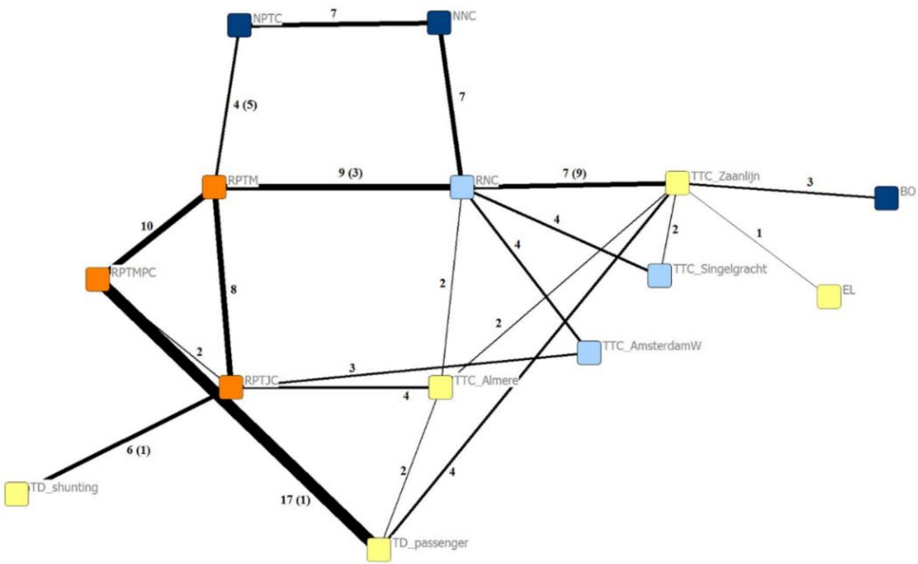


Figure 7.5: Undirected communication network in scenario 2.

The main actors with high values on degree centrality are the same as in scenario 1; that is, the RNC and the RPTM made the largest contribution to the flow of communications (see Table 7.4). The RPTMPC also appears to be a more central actor in terms of total number of interactions. The TTC Zaanlijn and the TTC Almere show a higher degree of closeness centrality in addition to the RNC, which is still the most central in efficiently obtaining information within the network. As in scenario 1, the RNC and TTC Zaanlijn control the information flow to other parts of the network; that is, they have a high betweenness centrality.

The RNC and the RPTMPC are central actors for outgoing communication, while the passenger train driver and the RPTM have a high degree centrality for receiving incoming communications. As in scenario 1, the RNC has the highest centrality when it comes to operators who want to reach him but cannot. In this subgroup of less available actors, the TTC Zaanlijn and NPTC are the most central with regard to failed outgoing connections. Regarding the duration of conversation, the RPTJC seems to be the main actor regarding the length of conversations with other actors.

Table 7.4: Centrality values for each communication network in scenario 2.

Role	**Undirected communication flow**			**Directed communication flow**		**Directed flow of failed communication attempts**		**Undirected average length conversations**
	Deg	**Clo**	**Betw**	**InDeg**	**OutDeg**	**InDeg**	**OutDeg**	**Deg**
NNC	6.4	65.4	4.1	2.3	4.1	0	0	4.9
RNC	**14.9***	**84.6**	**35.0**	5.9	**9.1**	**10.6**	1.0	13.3
TTC Zaanlijn	8.6	**82.7**	**34.3**	4.1	4.5	1.0	**7.7**	13.4
TTC Almere	4.5	**80.8**	13.7	2.7	1.8	0	0	6.7
TTC Singelgracht	2.7	69.2	0	1.8	0.9	0	0	2.5
TTC AmsterdamW	3.2	71.2	2.1	2.3	0.9	0	0	9.0
NPTC	5.0	59.6	1.9	2.7	2.3	0	**7.7**	3.6
RPTM	**14.0**	78.8	16.0	**7.2**	6.8	4.8	0	9.0
RPTJC	10.4	76.9	20.7	5.4	5.0	0	1.0	**16.9**
RPTMPC	**13.1**	73.1	4.2	2.3	**10.9**	0	1.0	6.5
TD shunting	2.7	53.8	0	2.7	0	1.0	0	1.7
TD passenger	10.4	73.1	5.1	**8.1**	2.3	1.0	0	4.7
BO	1.4	59.6	0	1.4	0	0	0	3.6
EC	.5	59.6	0	0	0.5	0	0	1.8

*Values in bold indicate a high centrality score

As in scenario 1, the RNC and the TTC Zaanlijn control the information flow. This position in relation to failed communication attempts indicate that the operators' workload is possibly affected by the 50% of failed incoming and 5% of failed outgoing communications attempts for the RNC, respectively to 10% of failed incoming and 44% of failed outgoing communications attempts for the TTC Zaanlijn. Operators also acted in a similar way as in scenario 1 when they could not reach another operator; occasionally they tried to reach the unresponsive operator until they reached him or her, while continuing their work.

7.6.2.2 Communication content

The semantic network in scenario 2 shows concepts that are identical to those in scenario 1 (see Figure 7.6). The current network, however, shows more interrelated nodes. This indicates a higher degree of the shared information behavior on these concepts between operators, although similarly focusing on turning possibilities and running rolling stock at major train stations.

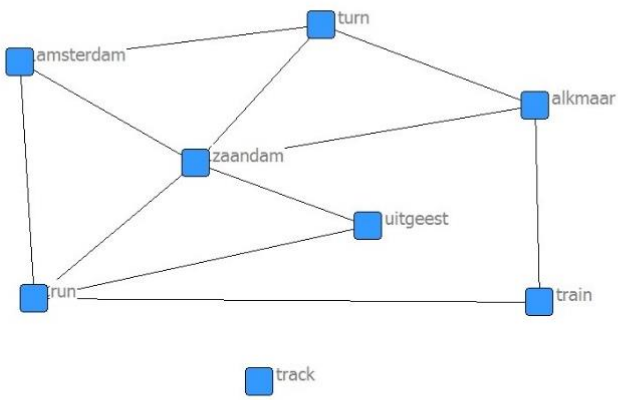


Figure 7.6: Semantic network of key concepts in scenario 2.

The results on the degree centrality metric for the current semantic network are depicted in Table 7.5. In comparison to scenario 1, the key concept in scenario 2 involved the activities around Zaandam station, as is to be expected.

Table 7.5: Degree centrality values for the semantic network in scenario 2.

Concept	**Deg**
zaandam	71.429
run	57.143
turn	42.857
amsterdam	42.857
alkmaar	42.857
uitgeest	28.571
train	28.571
track	0.000

A qualitative assessment regarding clusters in the semantic network in relation to the coordination of operators showed results similar to those in scenario 1, in which the communication initially focused on a gas leak in a tunnel between Zaandam and Amsterdam. Similarly, throughout the scenario, the coordination focused on identifying and ensuring the capacity in the disrupted area by dealing with the portion of rolling stock that was difficult to allocate to a station track or shunt to a yard. Since the disruption affected Zaandam station, which is located between Uitgeest and Amsterdam, the focus was clearly stronger on turning options, unlike in scenario 1. Operators may have often discussed the track options for specific train numbers, which might explain the disconnected node 'track'. The linkage of concepts to parts of the coordination provides insights into the developments of the network situation awareness in scenarios 1 and 2.

7.6.3 Comparisons between networks

The discrepancies between the two communication network structures were also investigated. A paired t-test was used (i.e. a density test in UCINET) to analyze whether there is significant difference between two networks with similar actors.

t-Tests were conducted to investigate the difference between the networks related to the undirected communication flow, the directed communication flow, the directed flow of failed communication attempts and the average length of the conversations in scenarios 1 and 2. No significant results were found, indicating that all four types of communication networks are not significantly different between scenarios. This may indicate that the alternative procedure is not significantly different when it comes to the communication flow, in comparison to the current procedures.

Although it is fairly remarkable that identical concepts in scenarios 1 and 2 were found, there was a discrepancy in the number of relations between nodes. However, no significant difference were found for the conducted t-test, indicating that the introduced disruptions and differences in procedures did not change the overall information exchange between operators.

7.7 Discussion and conclusion

In the current work, a macrocognitive paradigm was taken upon, following group cognition as a theoretical stream. This study focused on investigating similarities and differences in the network cognition between two types of procedures, i.e., the current and proposed way of working during a disruption. Quantitative social network analysis measures were used as an explorative technique to investigate the network cognition in the Dutch railway and passenger traffic control. This was conducted through communication flow and content, which was in line with the group cognition perspective that emphasized on interactions rather than on individual knowledge. Social network theory and analysis were applied to quantify and visualize the communication structures within the railway network.

7.7.1 Findings

7.7.1.1 Communication flow and content network structures

Different communication network structures were explored that were conceptualized through the four communication flow variables. The findings also show that each centrality metric in a network follows different interpretations of centrality and, therefore, different implications are related to the analysis. For example, the Zaanlijn train traffic controller may not be that central in terms of the number of contacts with other operators, but nonetheless serves as an important node as gatekeeper of information between different subgroups in the network. The identification of gatekeepers in decentralized command and control structures in relation to the number of failed communication attempts proves to be an important indicator of the possible inefficiency in operations and of an increased workload. As illustrated from the findings, operators occasionally called the RNC a couple of times before they were successful in reaching the operator. It should be noted that this issue only occurred with operators that were not co-located. The inability to reach an operator, especially the RNC, was not only caused by a busy line, but also due to the fact that an operator ignored the call, being too busy. The load of the RNC also has been reflected by the high centrality values in both scenarios, being the so-called 'spider' in the traffic control network. Given these findings, it would be interesting to conduct a more elaborate workload analysis for this role, in order to investigate other task load next to the communication load during a disruption.

In terms of overall values, predominantly the RNC and TTC Zaanlijn have high centrality values. This might be explained by the phase of the disruption, in which operators from the railway infrastructure manager (ProRail) need to mitigate the situation, especially in the first moments of a disruptions. It would be interesting to investigate if and how network values for different organizations would change when the disruption goes into a next phase, i.e., when traffic control operations run in accordance to an adapted time table and the final phase, i.e., when traffic control operations are scaling up train traffic to run in accordance with the regular time table.

Each conceptualization of communication flow provided unique insights into the communication and collaboration structure in terms of the centrality of operators, as also identified in earlier research (e.g., Houghton et al., 2006). It is notable that efficiency in terms of communication is a structural issue that is independent of the type of disruption mitigation procedure. As such, the synchronization of communication between operators in a decentralized command and control network is a key element for the coordination and optimization of performance (Stanton & Baber, 2006).

Further on, the current study looked into the assessment of semantic networks through communication content in relation to the coordination of the railway traffic and passenger traffic operations to provide insights into different activations of knowledge for network situation awareness. One major assumption for this qualitative assessment was that only one major event was introduced during the scenario. The key information elements were identified, which mainly focus on the train stations (location) and rolling stock, and on the actions, such as running or turning possibilities. It was also identified that the entire coordination revolves around capacity allocation, in which similar information is largely shared across the entire network. The need to share highly detailed information can be explained by its traces in the historical development of the current command and control structure. As such, the current ways of coordination may be seen as a reflection of an organization culture that has been observed to be rather resistant to the change (Steenhuisen, 2009). It is, however, difficult to draw firm conclusions on the basis of a qualitative analysis. Therefore, it is emphasized that research focusing on investigations of network situation awareness should analyze patterns of communication content and flow altogether, in order to be able to relate certain communication flow to communication content to identify network situation awareness. This linking of the communication flow and the content network structure can be performed using the EAST method (e.g., Stanton, 2014).

All in all, with the increased demands on higher infrastructure capacity in the future, developments in traffic management system are considered imminent. The current findings indicated that it takes six calls to inform the entire network about a disruption. Especially in the first moments of a disruption, every second counts to conduct safety measures in the traffic control system and to hold trains at stations that would leave towards the disrupted area. One obvious finding from this study is to reduce the communication overload, which can be realized by providing operators with newly and more specific shared (display) information to reduce the amount of verbal communication. For instance, a communication system could be used that is accessible by all operators, including TTCs and TDs which is currently not the case. Also in the light of improving operational efficiency during the first moments of a disruption, other developments could be in the automation of certain demanding tasks of the TTC, e.g., by letting the system take safety measures to alarm other nearby trains, revoke signals and/ or holding trains towards the disrupted area.

7.7.1.2 Comparison of two procedures

No significant differences were found between the two operating procedures on each of the four communication flow networks and the semantic networks, indicating that the different procedures did not have a significant impact on the way that operators communicated with each other, such as the communication frequency and length, or on the information they shared. The non-significant difference might be explained by the method that projects at ProRail adhere in designing and testing the newly created disruption mitigation procedure: new designs are often simplified and mostly remain proof-of-concepts before they are tested in a simulated environment (Van den Hoogen & Meijer, 2012). As such, independent of the outcome the application of social network analysis metrics provides useful support in testing the difference amongst alternative modes of the railway system as they can quantify the network and support the qualitative assessment of the network graph (e.g., Houghton et al., 2006).

7.7.1.3 The use and design of table-top simulations

The current study showed how table-top simulations can be applied to investigate operational processes. Although table-top design might be faster and cost-friendlier than the development of a human-in-the-loop simulation, careful design choices need to be made, such as the identification of information needs of operators and their routines in operational activities. Some operators can more easily adopt simulation environments that are more abstract, while others prefer to have a simulated system that is fully comparable to their real system. This difference might be related to how individual operators develop their situation awareness, which for some may be more in line with the information processing or distributed cognition perspective (e.g., Endsley, 1988; Stanton, Salmon, & Walker, 2015).

Another design choice that can be seen as a limitation in this study is the use of non-identical scenarios. Table-top simulation designers considered slightly different disruptions, that were developed by SMEs, to avoid learning effects. Operators may have dealt with the disruption in a faster and improved manner. For the current study, careful considerations were made to limit the converging implications of the chosen scenarios.

7.7.2 Limitations of this study

A limitation of the study is that only one composition of a network was assessed. Given possible variations between different team compositions, more research is needed into factors that influence team process behaviors within railway traffic operations.

The few available facilitators were a limiting factor in this study, particularly for the role of train driver. As only one facilitator was available for the role of TD shunting and TD passenger, he was in contact with six operators. This may have resulted in that the number of outgoing calls as TD passenger was lower than preferred, so that the TTCs received less calls from TD passenger. A consequence would be that TTCs' workload would be lower in comparison to an actual real-life disruption.

Although table-top simulation environments have proven their value in providing insights into the team processes and interactions, a limitation is the difficulty of collecting objective data (e.g., performance) through log files. Initial indications regarding the validity of the current table-top simulation are discussed by Lo and Meijer (2013). However, the validity of these isomorphic rule-based simulation environments should be more elaborately assessed in subsequent studies.

7.7.3 Future work

Further studies could also investigate the role of nonverbal communication, which might play a role especially in work environments where there are many operators. For instance, little explicit communication was observed between operators in some co-located rooms, which might be because operators listened to each other's conversations. In relation to nonverbal communications, operators could, for example, signal to one another to confirm that they heard a certain update without explicitly talking to each other at all.

Future work could also analyze the situation awareness of the railway traffic system or a subset thereof in terms of systemic SA, which is operationalized through the distributed situation awareness approach (e.g., Sorensen & Stanton, 2011; Salmon et al., 2009; Stanton et al., 2015). A comparison between the findings would indicate the contribution of information held in non-human components of the system.

Also, further research should focus on obtaining broader insights into the system's characteristics in different scenarios and in the actual work environment using the EAST method or dynamic network analysis (e.g. Schipper, Gerrits, & Koppenjan, 2015; Stanton, 2014). Comparisons between outcomes in an actual work environment and a simulated (table-top) environment could also provide indications in the validity of the used simulated environment.

References

Algemene Rekenkamer. (2012). Aanbesteden door NS-Railinfrabeheer 1995–2000 [Procurement by NS-Railinfrabeheer 1995–2000].

Aydoğan, R., Lo, J.C., Meijer, S.A., & Jonker, C.M. (2014). Modeling Network Controller Decisions Based Upon Situation Awareness through Agent-Based Negotiation. In: Meijer S.A., Smeds R. (eds) Frontiers in Gaming Simulation. ISAGA 2013. *Lecture Notes in Computer Science*, vol 8264. Springer, Cham.

Baber, C., Stanton, N. A., Atkinson, J., McMaster, R., & Houghton, R. J. (2013). Using social network analysis and agent-based modelling to explore information flow using common operational pictures for maritime search and rescue operations. *Ergonomics, 56*(6), 889-905.

Caro, P.W. (1973). Aircraft simulation and pilot training. *Human Factors, 15*, 502-509.

Cooke N. J., & Gorman, J. C. (2006) Assessment of team cognition. In: W. Karwowski (Ed.), *International encyclopedia of ergonomics and human factors (*pp 270–275). Boca Raton, Florida: CRC Press.

Cooke, N. J., & Gorman, J. C. (2009). Interaction-based measures of cognitive systems. *Journal of Cognitive Engineering and Decision Making, 3*(1), 27-46.

Cooke, N. J., Salas, E., Kiekel, P. A., & Bell, B. (2004). Advances in measuring team cognition. In E. Salas & S. M. Fiore (Eds.), *Team cognition: Understanding the factors that drive process and per-formance* (pp. 83–106). Washington, DC: American Psychological Association.

Cooke, N. J., Gorman, J. C., Kiekel, P. A., Foltz, P., & Martin, M. (2005). *Using team communication to understand team cognition in distributed vs. co-located mission environments* (Technical report for ONR Grant N00014-03-1-0580). Las Cruces: New Mexico State University.

Cooke, N. J., Gorman, J. C., Kiekel, P. A. (2008). Communication as team-level cognitive processing. In Letsky, M., Warner, N., Fiore, S., Smith, C. A. P. (Eds.), *Macrocognition in teams: Theories and methodologies* (pp. 51–64). Hants, UK: Ashgate.

Cooke, N. J., Gorman, J. C., Myers, C. W., & Duran, J. L. (2013). Interactive Team Cognition. *Cognitive Science, 37*(2), 255-285.

Dormans, J. (2011). Beyond Iconic Simulation. *Simulation & Gaming, 42*(5), 610-631.

Duke, R. D., & Geurts, J. L. A., 2004. *Policy Games for Strategic Management*. Dutch University Press, Amsterdam.

Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement. *Proceedings of the 32nd Human Factors Society Annual Meeting, 32*, 97-101.

Endsley, M. R. (1995). Towards a Theory of Situation Awareness in Dynamic Systems. *Human Factors, 37*(1), 32-64.

Endsley, M. R., Bolté, B. & Jones, D. G. (2003). *Designing for Situation Awareness*. New York, NY: Taylor & Francis Group.

Farrington-Darby, T., & Wilson, J. R. (2009). Understanding social interactions in complex work: a video ethnography. *Cognition, Technology & Work, 11*(1), 1-15.
Farrington-Darby, T., Wilson, J. R., Norris, B. J., & Clarke, T. (2006). A naturalistic study of railway controllers. *Ergonomics, 49*(12-13), 1370-1394.
Foltz, P. W., & Martin, M. J. (2009). Automated communication analysis of teams. In E. Salas, J. Goodwin, & C. S. Burke (Eds.), *Team effectiveness in complex organizations: Cross-disciplinary perspectives and approaches* (SIOP Organizational Frontiers Series, pp. 411–431). New York: Routledge.
Freeman, J., Weil, S. A., & Hess, K. P. (2006). *Measuring, monitoring, and managing knowledge in command and control organizations.* Technical report. Aptima Inc, Washington, DC
Funke, J. (1991). Solving complex problems: Exploration and control of complex systems. In R. J. Sternberg & P. A. Frensch (Eds.), *Complex problem solving: Principles and mechanisms* (pp. 185–222). Hillsdale, NJ: Lawrence Erlbaum.
Golightly, D., Wilson, J. R., Lowe, E., & Sharples, S. (2010). The role of situation awareness for understanding signaling and control in rail operations. *Theoretical Issues in Ergonomics Science, 11*(1-2), 84-98.
Golightly, D., Sandblad, B., Dadashi, N., Andersson, A. W., Tschirner, S., & Sharples, S. (2013). A socio-technical comparison of rail traffic control between GB and Sweden. In N. Dadashi, A. Scott, J. R. Wilson, & A. Mills (Eds.), *Rail human factors: Supporting reliability, safety and cost reduction* (pp. 367–376). London, UK: Taylor & Francis.
Gorman, J. C., Cooke, N. J., & Winner, J. L. (2006). Measuring team situation awareness in decentralized command and control environments. *Ergonomics, 49*(12-13), 1312-1325.
Gregoriades, A., & Sutcliffe, A. G. (2006). Automated assistance for human factors analysis in complex systems. *Ergonomics, 49*(12-13), 1265-1287.
Hauland, G. (2008). Measuring individual and team situation awareness during planning tasks in training of en route air traffic control. *The International Journal of Aviation Psychology, 18*(3), 290-304.
Haythornthwaite, C. (1996). Social network analysis: An approach and technique for the study of information exchange. *Library & Information Science Research, 18*(4), 323-342.
Houghton, R. J., Baber, C., McMaster, R., Stanton, N. A., Salmon, P., Stewart, R., & Walker, G. (2006). Command and control in emergency services operations: a social network analysis. *Ergonomics, 49*(12-13), 1204-1225.
Kiekel, P. A., Gorman, J. C., & Cooke, N. J. (2004). Measuring speech flow of co-located and distributed command and control teams during a communication channel glitch. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 48*(3), 683-687.
Klein, G. (2008). Naturalistic Decision Making. *Human Factors, 50*(3), 456-460.
Letsky M. P., & Warner, N. W. (2008). Macrocognition in teams. In: M.P. Letsky, N. W. Warner, S. Fiore, C. A. P. Smith (Eds.), *Macrocognition in Teams: Theories and Methodologies* (pp 1–14). Aldershot, Hampshire: Ashgate Publishing Ltd.
Lo, J. C., & Meijer, S. A. (2013). Measuring Group Situation Awareness in a Multi-Actor Gaming Simulation: A Pilot Study of Railway and Passenger Traffic Operators. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 57*(1), 177-181.
Lo, J. C., Van den Hoogen, J. & Meijer, S. A. (2013). Using Gaming Simulation Experiments to Test Railways Innovations: Implications for Validity. In R. Pasupathy, S. H. Kim, A. Tolk, R. Hill & M. E. Kuhl (Eds.), *Proceedings of the 2013 Winter Simulation Conference (WSC)* (pp. 1766-1777).
McMaster, R., Baber, C., & Houghton, R. J. (2005). Investigating alternative network structures for operational command and control. *10th International Command and Control Research and Technology Symposium*.
Meijer, S. A. (2012). Gaming simulations for railways: Lessons learned from modeling six games for the Dutch infrastructure management. In X. Perpinya (Ed.), *Infrastructure Design, Signaling and Security in Railway.* Croatia: IntechOpen.
Meijer, S. (2015). The power of sponges: Comparing high-tech and low-tech gaming for innovation. Simulation & Gaming, 46(5), 512-535.

Meijer, S., Hofstede, G. J., Beers, G., & Omta, S. W. F. (2006). Trust and tracing game: learning about transactions and embeddedness in a trade network. *Production Planning & Control, 17*(6), 569-583.
Millot, P. (2015). Situation Awareness: Is the glass half empty or half full? *Cognition, Technology & Work, 17*(2), 169-177.
Mogford, R. H. (1997). Mental models and situation awareness in air traffic control. *The International Journal of Aviation Psychology, 7*(4), 331-341.
NS. (2014). *Achter de Schermen* [Behind the screens].
Owens, H. (2004) Rail reform strategies: the Australian experience. In: I. Takatoshi, & A. O. Krueger, (Eds.), *Governance, Regulation and Privatization* (pp. 279 - 303). University of Chicago Press, Chicago
Pickup, L., Wilson, J. R., Sharpies, S., Norris, B., Clarke, T., & Young, M. S. (2005). Fundamental examination of mental workload in the rail industry. *Theoretical Issues in Ergonomics Science, 6*(6), 463-482.
Raser, J. C. (1969). *Simulations and Society: An Exploration of Scientific Gaming*. Boston, MA: Allyn & Bacon.
Salas, E., Prince, C., Baker, D. P., & Shrestha, L. (1995). Situation Awareness in Team Performance: Implications for Measurement and Training. *Human Factors, 37*(1), 123-136.
Salmon, P. M., Stanton, N. A., & Jenkins, D. P. (2017). *Distributed Situation Awareness: Theory, Measurement and Application to Teamwork*. Surrey: Ashgate.
Salmon, P. (2011). Coordination during multi-agency emergency response: issues and solutions. *Disaster Prevention and Management: An International Journal, 20*(2), 140-158.
Schipper, D., & Gerrits, L. (2018). Differences and similarities in European railway disruption management practices. *Journal of Rail Transport Planning & Management, 8*(1), 42-55.
Schipper, D., Gerrits, L., & Koppenjan, J. F. (2015). A dynamic network analysis of the information flows during the management of a railway disruption. *European Journal of Transport and Infrastructure Research, 15*(4).
Sorensen, L. J., & Stanton, N. A. (2011). Is SA shared or distributed in team work? An exploratory study in an intelligence analysis task. *International Journal of Industrial Ergonomics, 41*(6), 677-687.
Sorensen, L. J., & Stanton, N. A. (2012) Should we assess distributed situation awareness before, during or after command and control activity*? Journal of Battlefield Technology, (15*)1, 41–48.
Stahl, G. (2006). *Group Cognition: Computer Support for Building Collaborative Knowledge*. Cambridge: MIT Press.
Stanton, N. A. (2014). Representing distributed cognition in complex systems: how a submarine returns to periscope depth. *Ergonomics, 57*(3), 403-418.
Stanton, N. A., & Baber, C. (2006). The ergonomics of command and control. *Ergonomics, 49*(12-13), 1131-1138.
Stanton, N. A., Stewart, R., Harris, D., Houghton, R. J., Baber, C., McMaster, R., . . . Green, D. (2006). Distributed situation awareness in dynamic systems: theoretical development and application of an ergonomics methodology. *Ergonomics, 49*(12-13), 1288-1311.
Stanton, N. A., Salmon, P. M., Walker, G. H., & Jenkins, D. P. (2010). Is situation awareness all in the mind? *Theoretical Issues in Ergonomics Science, 11*(1-2), 29-40.
Stanton, N. A., Salmon, P. M., & Walker, G. H. (2014). Let the reader decide: a paradigm shift for situation awareness in sociotechnical systems. *Journal of Cognitive Engineering and Decision Making, 9*(1), 44-50.
Steenhuisen B (2009) *Competing public values: coping strategies in heavily regulated utility industries.* Dissertation, Delft University of Technology
Tannenbaum, S. I., Mathieu, J. E., Salas, E., & Cohen, D. (2015). Teams are changing: Are research and practice evolving fast enough? *Industrial and Organizational Psychology, 5*(1), 2-24.
Tenney, Y. J., & Pew, R. W. (2006). Situation awareness catches on: what? So what? Now what? *Reviews of Human Factors and Ergonomics, 2*(1), 1-34.
Van den Hoogen J, Meijer SA (2012) Deciding on innovation at a rail- way network operator: a grounded theory approach. *CESUN 2012 third international engineering systems symposium.*
Van den Hoogen, J., & Meijer, S. (2014). Gaming and simulation for railway innovation: A case study of the dutch railway system. *Simulation & Gaming, 46*(5), 489-511.
Top, J. V. D., & Steenhuisen, B. (2009). Understanding ambiguously structured rail traffic control practices. *International Journal of Technology, Policy and Management, 9*(2), 148-161.

Walker, G. H., Stanton, N. A., Baber, C., Wells, L., Gibson, H., Salmon, P., & Jenkins, D. (2010). From ethnography to the EAST method: A tractable approach for representing distributed cognition in Air Traffic Control. *Ergonomics, 53*(2), 184-197.

Walker, G. H., Stanton, N. A., Salmon, P. M., Jenkins, D. P., Rafferty, L., & Ladva, D. (2010). Same or different? Generalising from novices to experts in military command and control studies. *International Journal of Industrial Ergonomics, 40*(5), 473-483.

Weil, S. A., Foster, P., Freeman, J., Carley, K., Diesner, J., Franz, T., Cooke, N. J., Shope, S., Gorman, J. C. (2008). Converging approaches to automated communications-based assessment of team situation awareness. In: M. P. Letsky, N. W. Warner, S. Fiore, C. A. P. Smith (Eds.), *Macrocognition in Teams: Theories and Methodologies* (pp 277–303). Aldershot, Hampshire: Ashgate Publishing Ltd.

Werkwijze Verkeersleider (2012). ProRail VL Veiligheid, Logistiek en Vakmanschapcluster Be- en Bijsturing [ProRail Traffic Management Safety, Logistics and Expert Cluster].

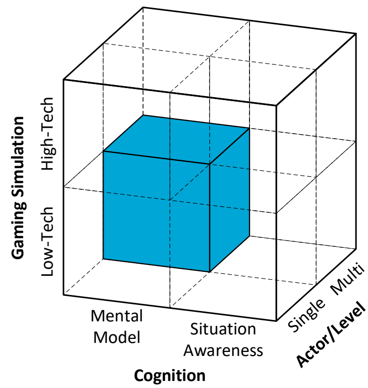
Wildman, J. L., Salas, E., & Scott, C. P. R. (2013). Measuring cognition in teams: A cross-domain review. *Human Factors, 56*(5), 911-941.

Wilson, J. R., & Norris, B. J. (2006). Human factors in support of a successful railway: a review. *Cognition, Technology & Work, 8*(1), 4-14.

Young, M. S., Brookhuis, K. A., Wickens, C. D., & Hancock, P. A. (2015). State of science: mental workload in ergonomics. *Ergonomics, 58*(1), 1-17.

8 Participatory design in large-scale railway infrastructure using gaming simulations: The role of shared mental models

This chapter has been submitted as:
Lo, J. C. & Meijer, S. A. (in review). Participatory design in large-scale railway infrastructure using gaming simulations: The role of shared mental models.



Abstract
This article focuses on participatory design on two abstraction levels: (1) the use of gaming simulation (or game) as a tool in the participatory design process of a complex socio-technical system, and (2) the participatory design of a gaming simulation itself by participants. A case study in the railway domain is presented in the form of a multi-actor table-top game. The OV-SAAL game has been used as part of a participatory design process to test different railway infrastructure configurations, in which insights could be retrieved to decide on the implementation of a solution. Secondly, railway and passenger traffic operators as participants were given room to redesign the game in terms of rules and flow during the sessions, in which we posit that the redesigned gaming simulation was strongly facilitated along the entire session through a development in their shared mental models. Applied game design principles are discussed and their implications are related to the cognition of operators and to the game outcomes.

Keywords: Participatory design; shared mental models; gaming simulation; multi-actor; table-top, railway system

8.1 Introduction

The railway sector adheres the typical characteristics of a socio-technical system (Wilson, Farrington-Darby, Cox, Bye, & Hockey, 2007). It combines technical elements such as computer-based automated train traffic control with social elements such as railway and passenger traffic operators (De Bruijn and Herder, 2009). As part of a long-term governmental and organizational program to match the projected growth in passenger volume of public transportation in the Netherlands, the Dutch railway infrastructure agency ProRail is in need to increase the infrastructure capacity to facilitate more trains on the railway network. Solutions are sought in near and far future innovations which directly impact the task-space of the railway traffic operators. To simulate alternative modes of the railway system, a number of gaming simulation sessions have been created as part of the Railway Gaming Suite program since 2009. Gaming simulation is herein defined as a simulation of a system where human participants take part and gaming methods and design principles are employed, in line with Duke and Geurts (2004). On occasion, the term 'games' will be used as an abbreviation of 'gaming simulations'. Gaming simulations encapsulate a broad spectrum of representations, i.e. high-fidelity simulator environments on the one end of a continuum to low-tech table-top simulation environments on the other end (Meijer, 2015). For the Railway Games, low-tech table-top environments or board games, in which railway traffic operators participated have been heavily used (for an overview see Meijer, 2015; Van den Hoogen, Lo, & Meijer, 2014). The multi-actor table-top games were predominantly low-tech games in the sense that they make use of analogue materials and playful visualizations (such as sponges and pegs) to represent components of the railway system (see Figure 8.1). In general, these railway gaming simulations are often closely designed in collaboration with subject matter experts (SMEs) who often are former railway traffic operators themselves (e.g. Meijer, 2015).

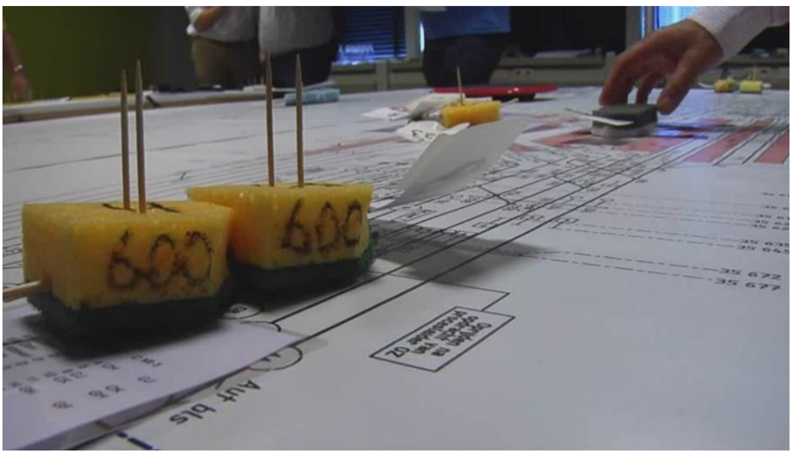


Figure 8.1: An illustration of a low-tech table-top railway game

Putting these games into relation with the participatory design (PD) process of large-scale systems, it can be stated that they are mostly applied in a mature stage of the design process. That is, project teams often perceived their solutions (i.e. new railway concepts or designs) as rather far-developed although in practice these showed to be rather simplified and conceptual (Van den Hoogen & Meijer, 2012). To improve the quality of the identified solutions, these were then tested in low-tech multi-actor environments as they were low-cost and be fairly fast to develop.

Thus, in the current work, the aim is to provide insights in the use of a gaming simulation that has been used as part of a PD process in testing different infrastructure configurations of the railway system, while keeping a certain degree of flexibility for participants to redesign the game during the session itself. As such, participatory design occurs on two abstraction levels: 1. the game as part of a participatory design process of a socio-technical system, 2. participatory design of the game itself through participants' involvement in shaping the game design rules and play throughout the session. As the level of PD design of the gaming simulation prior to the session itself was very limited, it was not included in this study. Moreover, the cognition of participants in this particular gaming simulation design is explored by investigating the development of their mutual understanding as an indicator for the creation of their shared mental model. Additionally, the extent to which an increased understanding between operators (i.e. shared mental model) can be related to the participatory design process of a game during the game session itself is investigated. Herein, the relation between the participatory design of the game during the session itself and the development of a mutual understanding (i.e. shared mental model) of operators is explored, resulting in a set of propositions.

The current case study aims to illustrate how to design and use a hybrid research-design type of game where a balance needs to be found: maintaining as much as possible the scope of using a game as a research tool (in the PD process of a system) while allowing flexibility in the game design (i.e. PD of the gaming simulation during the session) within the boundaries of experimental rigidity. Thus, although the two roles of participatory design in this article emphasize a broad topic, its wide scope also provides a realistic example how scopes are entangled and cannot always be confined to single demarcated topics.

This article is cross-disciplinary of nature and can provide insights in 1. game design considerations from a computer-based work environment such as railway traffic control to an analogue simulation environment, 2. game design trade-offs and its outcomes by the rigid requirements of research games and the flexible, participatory redesigned game play, and 3. the implications of shared mental model development in a participatory redesigned game. In the following section, the focus of participatory design in relation to the PD of a socio-technical system and the PD of a gaming simulation is elaborated. Section 3 provides a literature review on the concepts of shared mental models. In section 4, the design approach to multi-actor table-top gaming simulations is described to provide insights in the design steps taken. Section 5 addresses the applied game design principles for the case study, the OV-SAAL game, and an observational analysis of the designer-led role of the operators in which the development in their shared mental models is analyzed. Finally, the discussion and conclusion is presented in the last section.

8.2 Participatory designs

Although already existing for decades, design research is moving from a user-centered to a co-design, co-creation approach in which the user takes on a role in the design team (Sanders & Stappers, 2008). Participatory design is a field of design and research that encompasses a process in which artifacts, knowledge, work organization or system is created by various stakeholders (Muller, 2003; Spinuzzi, 2005). Depending on the stage in the development cycle and which participants (i.e. end users or designers) are included in the activity, different tools and techniques exist for PD such as questionnaires, interviews, ethnographic fieldwork techniques, but also including more context-oriented tools such as mock-ups, low-tech prototyping and design games (Kensing & Blomberg, 1998; Muller & Kuhn, 1993).

8.2.1 Participatory design of a socio-technical system

In designing a socio-technical system different design phases can be identified such as system design, system implementation and system adaptation and improvement, next to multiple design spaces such as the product, social and institutional space (Carayon, 2006; Reich & Subrahmanian, 2015). By translating these types of spaces to the railway sector, one can relate the product space to physical elements such as infrastructure, timetables and rolling stock and the social space to crew scheduling and roles and responsibilities. The institutional space can be related to for instance disruption mitigation procedures (Lo, Van den Hoogen, & Meijer, 2014). Within different design phases of a socio-technical system computer simulations are often applied to support early stages of the design cycle in order to explore more radical innovations, whereas gaming simulations are more useful in later stages to make predictions about the actual system performance (Sterman, 1987; Van den Hoogen & Meijer, 2015).

Games are often referred to as a third space or an in-between environment in supporting participatory design in which participants can combine different knowledge and insights to create solutions (Muller; 2003). Gaming simulations used in participatory design processes have been applied in various areas, such as the building and urban environment, healthcare, seaport and inland container terminals and maritime (Bekebrede & Meijer, 2009; Bockstael-Blok, Mayer, & Valentin, 2003; Garde, 2013; Habraken and Gross, 1988; Mayer et al., 2013). Most of these games included professionals that use the game environment to increase their knowledge in order to further optimize their organizational or system design. Simulated environments that include the operating layer in a complex system are more commonly used in the human factors and ergonomics discipline. Usually, highly realistic human-in-the-loop simulators are applied in rather mature stages of the design cycle. Table-top environments are applied more rarely (e.g. Stanard, Bearden, & Rothwell, 2013).

8.2.2 Participatory design of a gaming simulation

Next to the application of games as a tool in the participatory design of a system they themselves need to be designed at first. The use of participatory design with interaction designers has gained in popularity over the last years, yet little is applied in the design of games (Khaled, Abeele, & Mechelen, & Vasaloub, 2014). This scarce amount of cases in application is stated to be due to the related increase in complexity in which game designers would require to possess domain and game design knowledge (Khaled & Vasaloub, 2014). To support the interaction in the participatory design between game designers and users, values such as a common goal, a common language, a common understanding of the work context and tools that are simple and represent the work are required (Alenljung & Maurin Söderholm, 2015; Habraken & Gross, 1988; Olsson & Jansson, 2005). Participatory design of a gaming simulation is predominantly done before the actual gaming simulation session. The current article focuses on the participatory design of the game during the gaming simulation session itself through an exploration of the development in the shared mental models.

8.3 The role of shared mental models

Generally speaking about participatory design, the main object of study can be identified as defining users' knowledge in which the challenge is to make the tacit knowledge of participants explicit, as the knowledge of participants is highly valued (Spinuzi, 2005). Gaming simulations can serve as a platform to learn about tacit, systemic knowledge (i.e. mental models) (Hofstede, de Caluwé & Peters, 2010; Salas, Stout, & Cannon-Bowers, 1994). One important reason is that through the establishment of a common language using gaming simulations, a common understanding through developed shared mental models can be established in which a PD process can be most optimal fostered (Brandt, 2006; Dong, Kleinsmann, & Deken, 2013; Frauenberger, Good, Fitzpatrick, & Iversen, 2015). In line with this notion, the extent to which an increased understanding between operators (i.e. shared mental model) leads a participatory design process of a game within the game itself is investigated. The concepts of mental models and situation awareness are focal terms in complex systems (e.g. Sheridan, 2002). They are predominantly approached from an information-processing model which is based on individual cognition (Cooke, Gorman, Myers, & Duran, 2013; Klimoski & Mohammed, 1994). Mental models are often referred to as "mechanisms whereby humans are able to generate descriptions of system purpose and form, explanations of system functioning and observed system states, and predictions of future states" (p.7, Rouse & Morris, 1985). Mental models can be seen as a basis for the development of situation awareness which in simple terms refers to knowing what is happening around oneself (Endsley, 1988; 2000). As such, situation awareness closely relates to operational decision-making and a dynamic environment, while a mental model is a more abstract cognitive construct that is also relevant for strategic decision-making (Endsley, 1995; Schwenk, 1988).

Shared mental models can be conceptualized through shared knowledge structures, i.e. declarative, procedural and strategic mental models, in which declarative knowledge refers to the knowledge of what (e.g. facts and rules) (Lo, Van den Hoogen, & Meijer, 2014; Mohammed, Ferzandi, & Hamilton, 2010; Salas, Stout, & Cannon-Bowers, 1994). In contrast, procedural knowledge refers to the knowledge of how (e.g. refer timing and sequential knowledge) and systemic or strategic knowledge refer to knowledge concept and contingency plans (e.g. knowledge for problem solving). Another conceptualization also exists, such as the differentiations into technology/equipment, job/task, team interaction and team types of mental models (Cannon-Bowers, Salas, & Converse, 1993; Lim & Klein, 2006; Mathieu, Heffner, Goodwin, Salas, & Cannon-Bowers, 2000). The technology/equipment type refers to, for instance, equipment functioning and limitations, the job/task type to task strategies and procedures, the team interaction type to role and responsibilities and the team type to the knowledge of other team member's knowledge to, amongst others, skills and preferences.

Shared mental models can therefore be seen as the degree of knowledge that is overlapping between actors in a team or group. A related term to shared mental models is that of shared memory which is used in the context of the design of an artifact (Konda, Monarch, Sargent, & Subrahmanian, 1992). Shared memory can be distinguished in vertical memory (in-depth knowledge gained by experience) and horizontal memory (specific knowledge relating to one's role or profession). Cross-training is a known strategy for shared mental model development in which an understanding of other team members' roles and responsibilities is fostered by rotating team roles (e.g. Cannon-Bowers, Salas, & Converse, 1993; Espevik, Johnsen, & Eid, 2011). A related term for the development of shared mental models is 'mutual learning' (Luck, 2018).

Additionally, a number of factors are essential in a gaming simulation, such as activating implicit knowledge and learning from each other (Hofstede, de Caluwé, & Peters, 2010). Games facilitate the social interaction process by providing participants a platform to develop a common vocabulary that leads to better communication for experiences and perspectives to be shared (Harteveld, 2013; Weick, 1995). Also the use of a shared and simplified abstracted graphical interface are posited to be a driving factor in the facilitation of a common understanding as participants have the means to relate to their work environment without going to deep into details and discussions (Meijer, Kracht, Luipen, & Schaafsma, 2009).

8.4 Low-tech table-top gaming simulation design

Building on the experiences of previous gaming simulations, a number of steps in the translation from computer-based work environments of railway traffic operators to analogue table-top gaming simulation or board games were followed (Lo & Meijer, 2019). An important approach herein considered is the use of iconic simulation, in which rules of the reference system are induced to the simulated system (Dormans, 2011).

1. Determining the *purpose* of the gaming simulation: does the problem statement provided by the organization need to be tested or explored? Does it require the active involvement of participants and their creative capacity on the design of (non-)physical elements of the system, or is it the purpose for participants to gain knowledge that is required for their task? Identifying a type of gaming simulation supports the design process of the gaming simulation due to boundaries that are implied for its validity (Peters, Vissers, & Heijne, 1998). This phase is mostly determined by project managers from the organization and game designers.
2. Determining *components* of the reference system that are to be simulated, which can be translated through the identification and selection of physical elements of the simulated space (i.e. infrastructure for the railway system) and social elements, i.e. railway and passenger traffic and network operators. Project managers, subject matter experts and game designers are key actors in this phase of the game design.
3. Determining *scenario(s)*, i.e. type and strength of disruption(s) for the generalizability of the outcome of the gaming simulation in the case of an experimental, an explorative or a training focus. For gaming simulations where new (non-)physical elements are attributed to the system, e.g. identifying where to place new railway tracks or developing a new disruption mitigation procedure, scenarios may not be needed, but possibly desired. Especially SMEs are crucial in this phase to assess the consequences of different scenarios. Occasionally, computer simulations could support in gaining new insights in the impact of different types of disruptions.
4. Determining the *information needs* of operators, i.e. which information and in what format needs to be presented when to participants. Similar to the application of scenarios in the gaming simulation, information needs are especially significant in gaming simulations where task-related decisions take place, on the contrary to gaming simulations where more creative processes for newly created solutions are addressed. So far, the information needs have been determined in a few cases through SMEs and in later stages through a goal-directed task analysis (GDTA) (e.g. Endsley, Bolté, & Jones, 2003). The GDTA is a variation of a cognitive

task analysis that is created together with SMEs and maps the goals, related decisions and required information for those decisions. This phase has been experienced as most challenging in the design of a gaming simulation. Even with the support of SMEs during the design phase, participants may not accept the form in which relevant information has crystalized (e.g. Meijer, 2009). This unacceptance may be especially apparent when a combination of low-tech representation is used for the infrastructure, but rather high-tech visualization is presented for the information. A resemblance may be drawn with the phenomenon of the uncanny valley where rejection in the artefact becomes higher with close to real artefacts. This is often observed with humanoid robots.

For an elaborate description on the design aspects and representation of these analogue games, see Meijer (2015). Toups, Kerne, Hamilton and Shahzad (2011) describe a similar type of gaming simulation, which they call zero-fidelity simulation. In these games, human- and information-centric elements of the reference system are abstracted. However different game design aspects are highlighted (Toups, Kerne, & Hamilton, 2009).

8.5 Case study: OV-SAAL

In line with the long-term program to increase the capacity of the railway infrastructure in the Netherlands, a problem statement was presented involving a 1 billion Euro infrastructure investment for public transportation (in Dutch: openbaar vervoer (OV)) on the corridor Schiphol-Amsterdam-Almere-Lelystad (SAAL) (see Figure 8.2). The question raised here was which one of the four defined infrastructural expansions would be most robust against different types of disruptions and therefore most favorable to invest in. Additionally, the organization sought for feedback from expert participants with regards to other possible infrastructural improvements that could assess and optimize each solution even more.

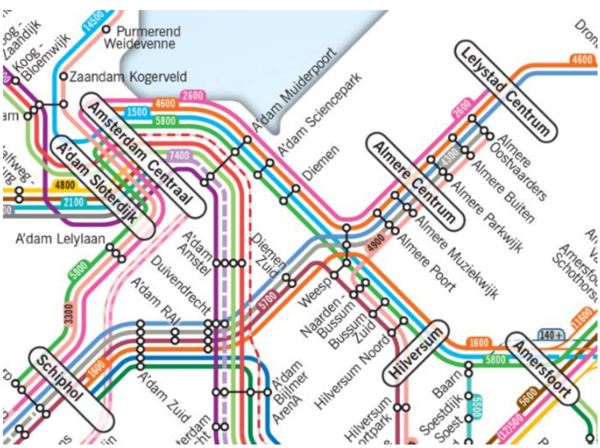


Figure 8.2: Area of the train traffic lines on the SAAL corridor in 2013

8.5.1 Method

In the OV-SAAL game, the game designers are faced with a challenge. While the main purpose of the game invokes certain design principles of a research gaming simulation, the game also inhibits the characteristics of design gaming simulation. In the latter case, flexibility within the gaming simulation was provided to the participants to adapt for instance the game rules and flow during the session itself. Thus, it was not known beforehand how the course of the game would develop during the session itself. Given the central role of the participants in the gaming simulation itself, care should be therefore taken in creating materials and rules that are manageable (Duke & Geurts, 2004). The following subsections describe the planned setup of the game environment and session prior to the gaming simulation itself.

8.5.1.1 Setup of the gaming simulation environment

Table 8.1 provides an overview of the OV-SAAL gaming simulation characteristics based on a framework for gaming simulation design (Meijer, 2009; Lo & Meijer, 2013).

Given the availability of different configurations on a specific part of the railway system and the desire to investigate these options, the gaming simulation serves the purpose of an experimental setup: what is the robustness of train traffic control in the different options?

The relevant infrastructure to be included in the simulated environment was already determined by the problem statement itself, focusing on the SAAL corridor. This area is controlled by one railway traffic control center responsible for the infrastructure capacity from the ProRail side and a passenger traffic control center that is responsible for the rolling stock (i.e. trains) and personnel from the NS side (in Dutch: Nederlandse Spoorwegen). Thus, relevant roles for this area were included. To account for the consequences of the disruptions on other regions, two other roles were included that assessed their impact.

For the scenarios it was decided to determine four types of disruptions to more elaborately explore the outcome of the different expansions with or without the inclusion of a European standardized signaling system (i.e. European Rail Traffic Management System - ERTMS) – thus in terms of design creating 4 types of infrastructure configurations x 4 types of disruptions x 2 signaling possibilities conditions to be evaluated. As the number of sessions was limited to one day it would be unmanageable to test all conditions. Therefore, instead of a continuous time model, a step-wise time model of 15 minutes is adopted to shorten the time length of each round. This step-wise time model can be related to decision rounds, in which participants needed to indicate at the end of each round what decisions or actions they would take after 15 minutes.

Table 8.1: Gaming simulation design choices for the OV-SAAL game.

Core Aspect	**Description**
Purpose	Exploring the impact of different infrastructural expansions
Scenarios	1. No infrastructural expansion, 2. Four additional tracks at Almere station, 3. Additional haul tracks at Weesp station, 4. Four additional tracks between Duivendrecht and Weesp station, 5. Implementation of European Rail Traffic Management System (ERTMS) in all four infrastructural layouts
Simulated world	Railway infrastructure on two corridors: Amsterdam Central Station - Lelystad and Amsterdam Zuid - Hilversum, co-location of operators occurred by seating arrangements (each table was a control center). Current timetable.
# of participants	8
Roles (# of participants)	Train traffic controller (TTC) (2), regional network controller (RNC) (2), national network controller (NNC) (1), regional passenger traffic monitor (RPTM) (2), national passenger traffic controller (NPTC) (1)
Type of role (# of participants)	Similar to own (5), prior experience in role (3)
Objectives	Determining own decisions for the next 15 minutes given the status of the system at paused moment
Constraints (# of roles)	Separation of train traffic regions: one regional train (2) and passenger traffic center (2) each versus other remaining regional train traffic center (2), exclusion of roles outside the defined infrastructure area, exclusion of train driver
Load	Four types of disruptions; 1. Local train delay (+5 min), 2. Freight train delay (+10 min), 3. Corridor train (intercity) delay (+10 min), 4. Disruption as chosen by participants themselves
Situation (external influencing factors)	Presence of observers and video cameras. At the end of the day results were discussed with invited stakeholders
Time model	Step-wise (per time periods of 15 min)

Operators received the information necessary to make task-related decisions including additional sheets to make notes to support their decision-making. Senior operators were invited to the gaming simulation as it was expected that the session might be cognitively challenging for operators. Consequently, it was expected that a well-developed mental model due to more work experiences would facilitate the process.

8.5.1.2 Setup of the gaming simulation session

The session was planned as follows: before the actual start of the scenarios all participants were asked to fill in a questionnaire on for instance their current function and work experience. Two video cameras were used to record all scenarios.

Each scenario would start at a similar time with a certain infrastructure layout and a certain disruption. Operators took place at their own control center (i.e. table), where they were asked to formulate their decisions for the next 15 minutes in a form, e.g. train 823 (train number) should be cancelled (decision) as it was expected that it would cause delays for other trains (consequences). Operators also were asked what they thought of the planned decisions of other operators. Participants were free to move around to inspect the infrastructure layout and discuss with other roles in the meantime. After their decisions were made, all participants gathered to discuss their individual decisions and agreed on one decision that would be implemented to instigate the next round (see Figure 8.3). Facilitators would implement the changes in the train schedule and positions of trains. This iteration continued for another five times, covering a total of about 1.5 hours of train traffic flow and until all infrastructural layouts were assessed or the game time had run out. At the end of the session a debriefing with all participants was held. About 5 hours of game time including breaks was available to test the different infrastructural options.

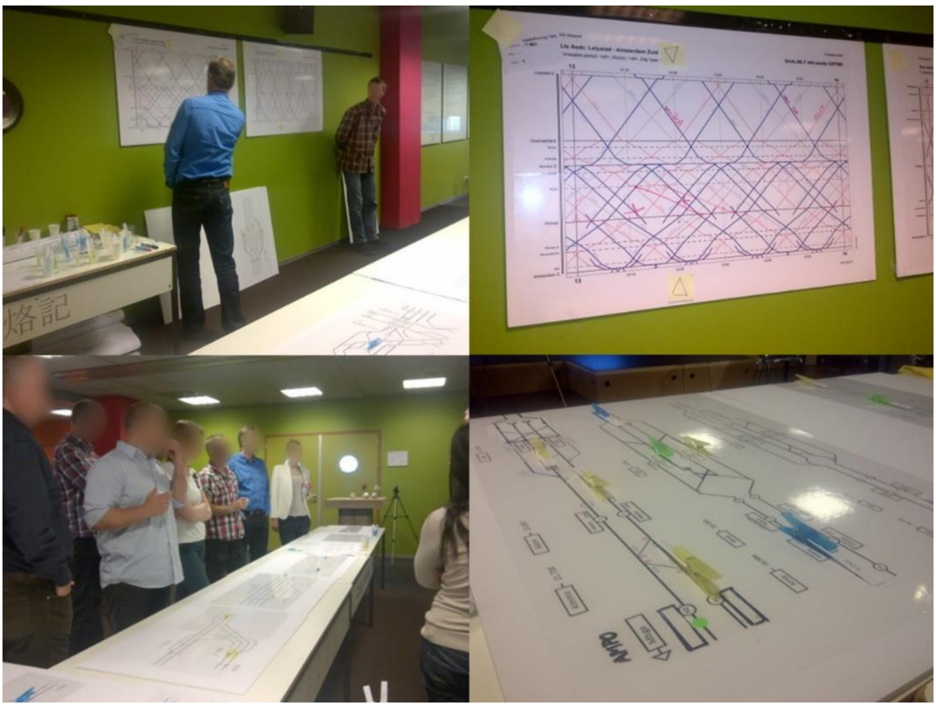


Figure 8.3: Top and bottom left: plenary session after each made decision where facilitators updated the status of the railway traffic flow (left top side) and participants discussed their decisions (left bottom side). Top right: implemented decisions in the traffic schedule on the corridor Amsterdam Centraal - Lelystad. Bottom right: railway infrastructure layout with pegs as trains, carrying train numbers.

8.5.1.3 Observation of shared mental model development

The process of shared mental model development amongst participants was observed during and after the gaming simulation by two researchers. During the game session, one researcher took on a role as facilitator and the other researcher took the role of an observer during the gaming simulation session. An indicator for shared mental model development was the level of congruency in opinions and decisions between operators in the discussions during the plenary sessions. Another indicator was the time needed to take a decision in each round. Video recordings were used to deduct the lenght of each round. A higher congruency between participants in their communication and faster decisions were assumed to be indicators of better developed shared mental models (e.g. Espevik, Johnsen, & Eid, 2011).

During each round, participants were also asked to fill in a form in which they formulated their individual preferred decision, before presenting and discussing their decision in the plenary session at the end of each round. The level of congruency in these decisions were also used as indicator for the level of mutual understanding between operators.

At the end of the session, participants received a questionnaire with Likert scale items on their perception of a mutual understanding between operators as an indicator for their shared mental model. Additionally, questions were included on their experience of the game (e.g. satisfaction, difficulty, clarity and atmosphere) and their overall satisfaction of the decisions throughout the session.

8.5.2 Outcomes of PD in the gaming simulation

In the process towards the analysis of infrastructural robustness, developments of shared mental model of operators were identified in parallel with the progress of the session and the increased role that participants had in determining the flow of the session. As such, three propositions were developed: as the session unfolded they (1) settled a common ground of the simulated environment (shared mental model development of the gaming simulation) on which they could build (2) a shared understanding on their tasks (shared mental model development of railway and passenger traffic operations), in order to (3) develop a shared understanding of the mechanisms in railway and passenger traffic control (shared mental model development of the solution space). Firstly, results from the questionnaires are presented.

8.5.2.1 Experience of the gaming simulation

The average work experience of the participants in the railway domain was 19.7 years (SD = 9.72). All eight operators indicated that they experienced the overall gaming simulation session as good in terms of satisfaction with how the game unfolded; M = 4.0, SD = .76. They did not think that the pace of the game flow was too slow; M = 2.3, SD =.71. The goal of the game was clear; M = 4.0, SD = .60. On the difficulty of the game they agreed slightly; M = 3.6, SD = .52.

Overall, they also thought that the atmosphere in the group was very good; *M* = 4.5, *SD* = .54.

With regards to participants' perception of their shared mental models, operators indicated in the post-questionnaire to have a good mutual understanding amongst each other; *M* = 4.0, *SD* = .54. They also indicated to be satisfied with their decisions made during the session *M* = 4.1, *SD* = .35.

8.5.2.2 Proposition 1: Shared mental model development of the gaming simulation environment occurs in the first stage

As a table-top environment was used to simulate the operational system of railway traffic controllers, operators received deviating tooling than they were accustomed to. The infrastructural representation was relevant for train traffic controllers (see Figure 8.3, bottom row), while the other operators were accustomed to a graphical time-location interface (see Figure 8.3, top row). As such, it can be expected that operators individually needed time to adjust to the different (representation of) their tooling to be able to make operational decisions.

After the first round, individual decisions were discussed, and the agreed collective decision was registered on the graphical time-location interface that represented the state of the train traffic flow. The tooling and collective decision-making process was new as operator's individual decision-making usually occur via their own systems. However, participants did not object nor comment on the rather abstract representation of their tooling and that the decision-making in the first round went smoothly. Also, there were no indications in terms of issues in their decision-making that were caused by an inability to understand the presented information. It can be argued that the use of a single graphical interface seemed to establish a common ground for operators to discuss and mutually agree on taking a decision.

8.5.2.3 Proposition 2: Shared mental model development of railway and passenger traffic operations occurs in the second stage

In the first round the disruption was directly introduced. It took about 10 minutes for operators to make an individual decision and another 4 minutes to agree on the implementation in the train traffic flow. During the second round already a change in the collaboration between operators could be observed: co-located operators were increasingly working together on how they would decide in the next round. For instance, a train traffic controller and regional network controller would discuss which train to prioritize over another and which train to hold.

In the second round it took approximately 7 minutes for operators to formulate their decisions and 4 minutes to agree on the decision to be implemented. Here the game leader decided to introduce some adaptations to the game flow by allowing space to participants to determine what the location of trains would be in order to increase the speed of this phase. At this stage operators already showed a high level of interaction. Operators did not indicate their individual preferred decision on the form they were asked to use, but instead approached their colleagues to discuss the different solutions and collectively decide. As such, the discussion regarding differences between operator's preferred decisions in the plenary session at the end of each round was no longer relevant. It is notable that operators commented that the high agreeableness between co-workers is rather uncommon in comparison to their daily experiences. The current case may provide an indication for the advantages of co-location, in which operators are able to quickly approach each other to discuss issues.

Additionally, it was questioned whether the current disruptions were interesting enough, which resulted in an early introduction of disruptions that were selected by operators themselves. This could be seen as an indication that operators reached a next level, namely having developed together a shared mental model of the solution space.

8.5.2.4 Proposition 3: Shared mental model development of the solution space occurs in the third stage

Through the development of operators' individual and shared mental model on the gaming simulation environment and on the goals and needs of their co-workers, it could be observed that operators reached a next stage in their shared mental model development of the railway system. In parallel, participants increasingly determined the course of the gaming simulation.

Firstly, it was observed that the length of the subsequent round of the individual decisions did not directly decrease after the introduced disruption (a switch malfunction) that was proposed by the operators, in comparison to the pre-selected disruption scenarios. This decrease could be explained by the increased complexity of the proposed disruption. However, in following rounds participants no longer thought it was necessary to formulate individual or co-located decisions as they could propose decisions on the spot in the plenary rounds, resulting in much faster decisions and solutions to the train traffic problem at hand.

Secondly, next to an increased speed in testing the outcome of certain disruptions, operators identified disruptions that would impact a certain infrastructure layout more heavily. Additionally, operators developed a typology of disruptions that have different impact on the railway system and which they differentiated in: light disruptions (e.g. small delays), infrastructural malfunctions (e.g. switch malfunctions) and structural delays (e.g. similar but strong delays on a number of trains). Based on this observation it appeared that operators developed a certain level of abstraction given the current subset of the railway system. They did not make use of a timetable planning anymore to evaluate or test the impact of the delay or disruption. Also, participants could more easily identify weaknesses within different infrastructural layouts, which resulted in a list of possibilities and constraints that could be used by the project management group in their consideration which infrastructural configuration to advice.

8.5.3 Outcomes of the gaming simulation in the PD process of a railway system

A list with implications of each solution was drafted on whiteboards during the game session based on the input from operators and at the end of the session re-evaluated. Based on these collected insights, the options (now extended to nine variations in total) were further assessed in a matrix with respect to 10 criteria, such as infrastructural implications (i.e. freight paths, integration with long-distance (corridor) train paths) and investment costs. This resulted in an advice report to the Dutch ministry of infrastructure and environment who made the final decision-making on the solution. Thereafter, the chosen solution was communicated by the respective ministry towards the Dutch parliament to inform them about the decision.

Although the findings provided insights in the impact of each option, the gaming simulation was not perceived as a success. The findings from the game session were qualitative of nature and additional input was provided by operators on the conceptual design of the options. As such, the purpose of the gaming simulation did not reflect that of an experimental setup. Rather, the gaming simulation could be perceived as a design and explorative manner of investigating the options. This type of gaming simulation was perceived as undesirable in a late stage of a decision-making process.

8.6 Discussion and conclusion

The focus of this article was twofold: to investigate the design and use of a gaming simulation as part of a participatory design process in which their goal was to test different infrastructural layouts with railway and passenger traffic operators. It was specifically designed to facilitate a large number of conditions (variations in infrastructure layout in combination with different types of disruptions), therefore using a step-wise time model and a single graphical interface that was shared by all operators. The usefulness of gaming simulation proofed itself by the integration of the results across multiple decision-making layers which resulted in a decision letter to the Dutch parliament. Especially qualitative assessments made by operators about the (dis-)advantages related to each solution were used as key arguments in this decision-making. Furthermore, the participatory design approach that is taken on within the gaming simulation itself was investigated, in which operators had a designer-led role in determining the flow of the game.

This study did not measure actual shared mental models. Rather, the process of shared mental model development was investigated, whereas three shared mental model types that can be related to the redesign of the gaming simulation itself were identified. Based on the findings, three propositions were formulated, in which (1) participants developed a shared understanding regarding the game environment, they then (2) developed a shared understanding of each other's goals and they (3) developed a shared understanding on the characteristics (strengths and weaknesses) within this specific subset of the railway system and related implications for different types of disruptions. Parallels can be drawn with the different types of shared mental models, in which a shared mental model of the game environment can be related to the technology/equipment type, the shared mental model of railway and passenger traffic operations can be related to the team interaction and team type and the shared mental model of the solution space can be related to the job/task type. Consequently, participants became faster in making decisions and were able to make more general implications about the impact of different types of disruptions as the rounds in the gaming simulation moved forward. Thus, operators indirectly redesigned the game flow in a participatory manner. Similarities in shared mental model development can be drawn with how a design project unfold over time; the initial phase is centered on a shared understanding between all involved parties, followed by addressing design issues in the project (Lundmark, 2017). All in all, it is posited that an understanding of shared mental model development contributes in identifying change points in the design of the gaming simulation: with a certain established level of mutual understanding, a change in the game design aspects by participants can be expected; the game rules could be altered in terms of length of rounds and structure, but also a change in the use of the visualizations can be expected.

A number of conclusions may be drawn from our analysis. First of all, the use of a single and more abstract graphical interface may have been a determining factor in building the three different shared mental model components, which was crucial for the acceptance and the design-led approach of operators in the gaming simulation. This is line with the expectations from a previous study, which stated that a richer graphical interface, e.g. with geographical information, can lead to a detailed level of discussion and higher resistance of the gaming simulation environment (Meijer, Kracht, Luipen, & Schaafsma, 2009). As such, participants were able to draw on their tacit knowledge which was also facilitated by the dialogues between operators in building a common language (Brandt, Messeter, & Binder, 2008; Wood et al., 2014). Operators who normally work in a dynamic environment and focus on different detailed aspects of the railway system were forced to think in a more abstract but similar graphical representation, facilitating a mutual understanding and therefore higher agreeableness (Toups, Kerne, Hamilton, & Shahzad, 2011). As operators already have a strong knowledge base of the railway system, the current game design directed them into thinking on a meta-level of the railway system (i.e. different types of disruptions and their implications) and making their tacit knowledge explicit (Wood et al., 2014). Thus, from a top-down approach in testing different disruptions and infrastructure layouts, the gaming simulation turned into a bottom-up approach where operators were able to identify weaknesses in the infrastructure and related disruptions that could amplify these weaknesses.

Future work should look into the implications for gaming simulation validity in these hybrid research-design types of games. Also, group dynamics in the development of shared mental models of participants could be investigated, e.g. which group composition may lead to worse or better results. Additionally, it would also be interesting to test the quality of the derived typology of railway disruptions in subsequent gaming simulations.

References

Alenljung B., & Maurin Söderholm H. (2015). Designing Simulation-Based Training for Prehospital Emergency Care: Participation from a Participant Perspective. In: Kurosu M. (Ed.) *Human-Computer Interaction: Design and Evaluation. HCI 2015*. Lecture Notes in Computer Science, 9169. Springer, Cham.

Bekebrede, G., & Meijer, S.A. (2009). Understanding Complex Infrastructure Systems: The Case of SimPort-MV2. *Infrastructure Systems and Services: Developing 21st Century Infrastructure Networks (INFRA)*, 1-6.

Bockstael-Blok, W., Mayer, I., & Valentin, E. (2003). Supporting the Design of an Inland Container Terminal through Visualization, Simulation and Gaming. *Proceedings of the 36th Annual Hawaii International Conference on System Sciences*.

Brandt, E. (2006). Designing Exploratory Design Games: A Framework for Participation in Participatory Design? *Proceedings of the ninth conference on Participatory design: Expanding boundaries in design - Volume 1*. ACM, Trento, Italy, 57-66.

Brandt, E., Messeter, J., & Binder, T. (2008). Formatting Design Dialogues – Games and Participation. *CoDesign, 4*(1), 51-64.

Carayon, P. (2006). Human Factors of Complex Sociotechnical Systems. *Applied Ergonomics, 37*(4), 525-535.

Cannon-Bowers, J.A., Salas, E., & Converse, S. (1993). Shared Mental Models in Expert Team Decision Making. In: Castellan, N.J. (eds) *Individual and Group Decision Making: Current Issues*. Lawrence Erlbaum Associates, Hillsdale, NJ, 221-246.

Cooke, N. J., Gorman, J. C., Myers, C. W., & Duran, J. L. (2013). Interactive Team Cognition. *Cognitive Science, 37*(2), 255-285.

De Bruijn, H. D., & Herder, P. M. (2009). System and Actor Perspectives on Sociotechnical Systems. *Systems, Man and Cybernetics, Part A: Systems and Humans, 39*(5), 981-992.

Dong, A., Kleinsmann, M. S., & Deken, F. (2013). Investigating Design Cognition in the Construction and Enactment of Team Mental Models. *Design Studies, 34*(1), 1-33.

Dormans, J. (2011). Beyond Iconic Simulation. *Simulation & Gaming, 42*(5), 610-631.

Duke, R. D., & Geurts, J. L. A., 2004. *Policy Games for Strategic Management*. Dutch University Press, Amsterdam.

Endsley, M. R. (1988). Design and Evaluation for Situation Awareness Enhancement. *Proceedings of the Human Factors Society Annual Meeting, 32*(2), 97-101.

Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors, 37*(1), 32-64.

Endsley, M.R. (2000). Theoretical Underpinnings of Situation Awareness: A Critical Review. In: Endsley, M.R., & Garland, D. J. (Ed.) *Situation Awareness Analysis and Measurement*. Lawrence Erlbaum Associates, Mahwah, NJ.

Endsley, M.R., Bolté, B., & Jones, D.G. (2003). *Designing for Situation Awareness*. Taylor & Francis Group, New York, NY.

Espevik, R., Johnsen, B. H., & Eid, J. (2011). Outcomes of Shared Mental Models of Team Members in Cross Training and High-Intensity Simulations. *Journal of Cognitive Engineering and Decision Making, 5*(4), 352-377.

Frauenberger, C., Good, J., Fitzpatrick, G., & Iversen, O. S. (2015). In Pursuit of Rigour and Accountability in Participatory Design. *International Journal of Human-Computer Studies 74*, 93-106.

Garde, J.A. (2013). *Everyone Has a Part to Play: Games and Participatory Design in Healthcare*. Universiteit Twente, Enschede.

Habraken, N.J. & Gross, M.D. (1988). Concept Design Games. *Design Studies 9*(3), 150-158.

Harteveld, C. (2013). *Making Sense of Virtual Risks: A Quasi-experimental Investigation into Game-based Training*. IOS Press, Amsterdam.

Hofstede, G. J., de Caluwé, L., & Peters, V. (2010). Why Simulation Games Work-In Search of the Active Substance: A Synthesis. *Simulation & Gaming, 41*(6), 824-843.

Kensing, F., & Blomberg, J. (1998). Participatory Design: Issues and Concerns. *Computer Supported Cooperative Work (CSCW), 7*(3), 167-185.

Khaled, R., & Vasalou, A. (2014). Bridging Serious Games and Participatory Design. *International Journal of Child-Computer Interaction, 2*(2), 93-100.

Khaled, R., Abeele, V. V., Mechelen, M. V., & Vasalou, A. (2014). Participatory Design for Serious Game Design: Truth and Lies. *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play*, Toronto, Ontario, Canada, 457-460.

Klimoski, R., & Mohammed, S. (1994). Team Mental Model: Construct or Metaphor? *Journal of Management, 20*(2), 403-437.

Konda, S., Monarch, I., Sargent, P., & Subrahmanian, E. (1992). Shared Memory in Design: A Unifying Theme for Research and Practice. *Research in Engineering Design, 4*(1), 23-42.

Lo, J. C., & Meijer, S. A. (2019). Assessing Network Cognition in the Dutch Railway System: Insights into Network Situation Awareness and Workload using Social Network Analysis. *Cognition, Technology & Work*.

Lo, J. C., & Meijer, S. A. (2013). Measuring Group Situation Awareness in a Multiactor Gaming Simulation: A Pilot Study of Railway and Passenger Traffic Operators. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 57*(1), 177-181.

Lo, J. C., Van den Hoogen, J., Meijer, S. A. (2014). Testing Changes in the Railway System through Gaming Simulation: How Different Types of Innovations Affect Operators' Mental Models. *Advances in Human Aspects of Transportation: Part III, 9*, 291-302.

Lim, B.-C., & Klein, K. J. (2006). Team Mental Models and Team Performance: A Field Study of the Effects of Team Mental Model Similarity and Accuracy. *Journal of Organizational Behavior, 27*(4), 403-418.

Luck, R. (2018). What is it that Makes Participation in Design Participatory Design? *Design Studies, 59,* 1-8.

Lundmark, S. (2018). Design Project Failures: Outcomes and Gains of Participation in Design. *Design Studies, 59*, 77-94.

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology, 85*(2), 273-283.

Mayer, I., Zhou, Q., Lo, J., Abspoel, L., Keijser, X., Olsen, E., . . . Kannen, A. (2013). Integrated, Ecosystem-Based Marine Spatial Planning: Design and Results of a Game-based, Quasi-experiment. *Ocean & Coastal Management, 82*, 7-26.

Meijer, S.A. (2009). *The Organisation of Transactions: Studying Supply Networks Using Gaming Simulation*. Academic Publishers, Wageningen.

Meijer, S. (2015). The power of sponges: Comparing high-tech and low-tech gaming for innovation. *Simulation & Gaming, 46*(5), 512-535.

Meijer, S. A., Van der Kracht, P., Van Luipen, J. J. W., & Schaafsma, A. A. M. (2009). Studying a control concept for high-frequency train transport. *Infrastructure Systems and Services: Developing 21st Century Infrastructure Networks (INFRA),* 1–6.

Mohammed, S., Ferzandi, L., & Hamilton, K. (2010). Metaphor No More: A 15-Year Review of the Team Mental Model Construct. *Journal of Management, 36*(4), 876-910.

Muller, M. J. (2003). Participatory Design: The Third Space in HCI. In: Sears, A, Jacko, J.A. (Ed.) *Human-Computer Interaction: Development Process*, 165-186.

Muller, M. J., & Kuhn, S. (1993). Participatory Design. *Commun. ACM 36*, 24-28.

Olsson, E., & Jansson, A. (2005). Participatory Design with Train Drivers - A Process Analysis. *Interact. Comput., 17*(2), 147-166.

Peters, V., Vissers, G., & Heijne, G. (1998). The Validity of Games. *Simulation & Gaming, 29*(1), 20-30.

Reich, Y., & Subrahmanian, E. (2015). Designing PSI: An Introduction to the PSI Framework. *Proceedings of the 20th International Conference on Engineering Design (ICED15),* Milan, Italy.

Rouse, W., & Morris, N. (1985). *On Looking into the Black Box: Prospects and Limits in the Search for Mental Models*. Technical Report 85-2, Center for Man-Machine Systems Research.

Salas, E., Stout, R. J., & Cannon-Bowers, J. A. (1994). The Role of Shared Mental Models in Developing Shared Situational Awareness. In: Gilson, R. D., Garland, D. J., & Koonce, J. M. (Ed.) *Situational Awareness in Complex Systems*. Embry-Riddle Aeronautical University Press, Daytona Beach, Florida, 297-304.

Sanders, E. B. N., & Stappers, P. J. (2008). Co-creation and the new landscapes of design. *CoDesign, 4*(1), 5-18.

Schwenk, C. R., (1988). The Cognitive Perspective on Strategic Decision Making. *Journal of Management Studies 25*(1), 41-55.

Sheridan, T. B. (2002). *Humans and Automation: System Design and Research Issues*. John Wiley & Sons, Inc, Santa Monica, CA.

Spinuzzi, C. (2005). The Methodology of Participatory Design. *Technical Communication 52*(2), 163-174.

Stanard, T., Bearden, G., & Rothwell, C. (2013). A Cognitive Task Analysis to Elicit Preliminary Requirements for an Automated UAV Verification & Planning System. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 57*(1), 210-214.

Sterman, J. D. (1987). Testing Behavioral Simulation Models by Direct Experiment. *Management Science, 33*(12), 1572-1592.

Toups, Z. O., Kerne, A., & Hamilton, W. (2009). Game Design Principles for Engaging Cooperative Play: Core Mechanics and Interfaces for Non-Mimetic Simulation of Fire Emergency Response. *Proceedings of the 2009 ACM SIGGRAPH Symposium on Video Games*, New Orleans, Louisiana, 71-78.

Toups, Z. O., Kerne, A., Hamilton, W. A., & Shahzad, N. (2011). Zero-fidelity Simulation of Fire Emergency Response: Improving Team Coordination Learning. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Vancouver, BC, Canada, 1959-1968.

Van den Hoogen, J., & Meijer, S. A. (2012). Deciding on Innovation at a Railway Network Operator: A Grounded Theory Approach*. CESUN 2012: 3rd International Engineering Systems Symposium*.

Van den Hoogen, J., & Meijer, S. (2015). Gaming and Simulation for Railway Innovation: A Case Study of the Dutch Railway System. *Simulation & Gaming, 46*(5), 489-511.

Van den Hoogen, J., Lo, J. C., & Meijer, S. A. (2014). Debriefing in Gaming Simulation for Research: Opening the Black Box of the Non-Trivial Machine to Assess Validity and Reliability. *Proceedings of the 2014 Winter Simulation Conference*. IEEE Press, Savannah, Georgia, 3505-3516.

Weick, K. E. (1995). *Sensemaking in Organizations*. Sage, Thousand Oaks, CA.

Wilson, J. R., Farrington-Darby, T., Cox, G., Bye, R., & Hockey, G. R. J. (2007). The Railway as a Socio-Technical System: Human Factors at the Heart of Successful Rail Engineering. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 221*(1), 101-115.

Wood, G., van der Horst, D., Day, R., Bakaoukas, A. G., Petridis, P., Liu, S., . . . Pisithpunth, C. (2014). Serious games for energy social science research. *Technology Analysis & Strategic Management, 26*(10), 1212-1227.

Section III

Applications of human factors research and gaming simulations: Intelligent agents, participatory system design and hybrid gaming simulation

This section consists of several chapters that focus on human factors and gaming simulations in several application areas. These application areas are themed on the fields of (1) multi-agents systems, i.e. intelligent agents, (2) participatory system design and (3) hybrid gaming simulations.

Chapter 9 focuses on the modelling of operator's cognition, specifically situation awareness, that leads to the development of an intelligent agent. The issue of scarce operator resources available for gaming simulations can be solved by using intelligent agents, in which agents can make decisions on behalf of an operator.

Chapter 10 connects the previous chapters on operator's mental models and situation awareness to railway system design. Strategic decision-makers recognize the need to test the impact of operational changes using gaming simulation before they are implemented. The focus in this chapter is on the requirements of cognitive components that extract certain operator knowledge, the value of this operator knowledge in relation to the organization's questions and implications for the gaming simulation design.

Chapter 11 focuses on the validity of so-called hybrid gaming simulations, i.e. gaming simulations that are multi-purposed. In the currently applied games, the existence of hybrid gaming simulations is caused by the presence of an organizational research focus and in parallel an academic, human factors research focus. Implications for validity are discussed.

9 Modeling network controller decisions based upon situation awareness through agent-based negotiation

This chapter has been published as:
Aydoğan, R., Lo, J.C., Meijer, S.A., & Jonker, C.M. (2014). Modeling Network Controller Decisions Based Upon Situation Awareness through Agent-Based Negotiation. In: Meijer S.A., Smeds R. (eds) Frontiers in Gaming Simulation. ISAGA 2013. *Lecture Notes in Computer Science*, vol 8264. Springer, Cham.

Abstract
This chapter focuses on the cognitive modeling of one of the railway traffic operating roles, specifically that of the national network controller. With the growing demand to conduct more gaming simulations more operators are needed as participants in gaming simulations, resulting in higher pressure for the organization in terms of personel and financial resources. As a solution to reduce the pressure on these resources, the use of intelligent software agents is proposed to replace participant roles that do not have a key role in game scenarios. As the negotiation and decision-making is a key task of network operators, a cognitive model is developed in which a situation awareness model (SAM) is introduced in relation to evaluating offers in complex and dynamic systems.

Keywords: Cognitive modelling; negotiation; situation awareness; intelligent agents; network control and management

9.1 Introduction

The Dutch railway system is one of the most heavy utilized railway infrastructures in the world comprising over 7000 kilometers of track to transport over 100.000 tons cargo and 1.2 million passengers per day (Meijer, 2012; Over ProRail, 2012). Management of such systems is highly complex since it requires to take into account both social and technical aspects. Those aspects involve issues related to the infrastructure, such as track conditions and operational issues related to the trains and traffic control. Handling the aforementioned issues becomes even more challenging as they are influenced by a variety of factors such as environmental and technical factors (Al-Ibrahim, 2010). For instance, some problems may occur on the tracks under hard weather conditions; consequently, it may cause (primary) delays. Due to interdependencies between the trains, delays affect other trains' schedules (secondary delays), which is especially a crucial issue in the highly interconnected and dense train traffic.

When focusing on the social aspects of the railway system, in particular train traffic control, the complexity in such situations might also evolve when a group decision making process occurs, e.g. national and regional controllers often need to collaborate with each other when failures, incidents and/or emergencies in the train traffic network occur. In such cases, the interest of individuals from different divisions and in separated locations may conflict, since they perceive the problem from different perspectives – different information, knowledge and understanding of processes ('mental models') – of the railway domain. A decision might be favorable for one individual but unacceptable for the others. To reach a joint agreement they may negotiate on their decisions. Finding a mutually agreed decision would be time consuming under some circumstances. Especially in train traffic control, it is crucial to make fast decisions before conflicts in the system escalate (Meijer, Van der Kracht, Van Luipen, & Schaafsma, 2009). A concept that has been often used for operational decision-making is 'situation awareness'. In simple terms, situation awareness is about knowing what is going on. A high level of situation awareness would be a predictor of a good decision. In a negotiation setting, a high level of situation awareness can be seen as a predictor of generating an offer.

The Dutch railway network infrastructure organization uses different decision support tools (e.g. case studies, computer simulations and gaming simulations) to improve decision-making mechanisms in terms of speed and quality. Gaming simulations are used by the Dutch railway infrastructure organization to understand the underlying problems, to explore alternative models of the organization, to experience exceptional situations, and to improve the decision-making process (e.g. Meijer, Mayer, Van Luipen, & Weitenberg, 2011). It is expected that the amount of gaming simulation sessions will increase from 5 to 50 per year. Since gaming simulations require the presence of individuals from different divisions, alignment of participation presence is not only problematic, but also costly in terms of time and resources. Employing intelligent software agents to act on behalf of human participants might save the latter two.

This study presents an approach in which intelligent software agents negotiate as virtual represents of stakeholders in the operational traffic control process in gaming simulations. A key contribution is the way in which the situation awareness of human participants is modeled in the agent-based system for a negotiation setting. Section 2 provides a background on situation awareness and the Dutch railway traffic control system. Section 3 presents a model of situation awareness used for the evaluation of offers in a negotiation setting, followed by Section 4, which elaborates on the negotiation model. Finally, theoretical and practical implications are given, as well as remarks and recommendations for future work.

9.2 Background

9.2.1 Situation Awareness

A common concept of cognition used in complex and demanding activities, is that of situation awareness. In transport related research it is a common concept in studying driver and pilot behavior (Lee, Choi, Choi, & Ujimoto, 2007). Situation awareness is argued to be an epiphenomenon of cognition and forms critical input for decision-making (Croft, Banbury, Butler & Berry, 2004; Endsley, 1995). One of the most widely used definitions of situation awareness is: "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" (Endsley, 1988, p.97). Situation awareness is constructed through mental models or schemas, which can be seen as representations of a domain (e.g. traffic control) in terms of knowledge and processes (Gentner, 2001). This theoretical definition takes upon an information processing approach within the psychological domain, whereas long-term memory structures, such as mental models and schemas are significant parts of obtaining high situation awareness (Endsley, Bolté, & Jones, 2003).

Situation awareness and mental models have been studied in a variety of research fields. For example, Jonker, Van Riemsdijk, Vermeulen, and Den Helder (2010) investigated shared mental models to increase the performance of teams using agents. Hoogendoorn and Van Lambalgen (2011) focused on modeling situation awareness through mental models. McCarley, Wickens, Goh, and Horrey (2002) presented a situation awareness model based on visual attention and memory decay. This model showed more alignment with plausible cognitive processes than other computational models in a review by Rousseau, Tremblay, and Breton (2004). However, the model mainly focuses on the role of attention and working memory to achieve situation awareness. So and Sonenberg (2004) proposed a computational model of situation awareness in which the main focus was to evaluate and enhance the capability to predict foreseeable situations and act to these situations in a proactive manner. Their computational model of situation awareness is based on rule-based knowledge and forward chaining reasoning. Studies using models of situation awareness in train traffic control and/or the railway domain are fairly limited. There are some studies focusing on a situation model of the train driver's performance (McLeon, Walker, & Moray, 2005), and a model of the train driver's information processing (McGuinness & Foy, 2000). In this study, a situation awareness model of train network controllers is studied. Section 3 presents the situation awareness model, which will be used for evaluating offers in a negotiation setting.

9.2.2 Railway Traffic Control Operations

The operations of train traffic control can be divided into three traffic control roles, namely as train traffic controller, regional network controller and national network controller. Thirteen regional control centers in the Netherlands focus on an assigned area of the train traffic network. The coverage of the regional control center depends on the size of the stations and complexity of the railway conditions. The train traffic controller (TTC) is responsible for ready and correct availability of safe, distributed infrastructure capacity in a sub-region of a regional control center. In ideal conditions, the TTC solely needs to monitor the status and progress of the timetable. Allocation software (in Dutch ARI) automatically runs the predefined train allocation plan. The regional network controller (RNC) (also called a de-central network controller; Meijer, Van der Kracht, Van Luipen, & Schaafsma, 2009) overlooks the status of the train activities in all sub-regions and coordinates between the train traffic controllers. As trains travel between multiple sub-regions, the RNC is responsible for the optimization and management of the train traffic at a regional control center. Similarly, but on a national level, the national network controller (NNC) (also called central network controller; Meijer, Van der Kracht, Van Luipen, & Schaafsma, 2009) coordinates the activities between regional network controllers in case of failures, incidents and emergencies in the railway network and focuses on trains that need to cross the borders of multiple regional control centers. Additionally, the NNC interacts with the train service providers on the availability of personnel and timetable alterations for cross-regional and international passenger trains and on timetable alterations of freight trains. As mentioned before, this study focuses on the negotiation that may occur between the regional network controller and the national network controller.

9.3 Modeling Situation Awareness in Negotiations

As situation awareness is an important cognitive indicator for decision-making, this construct has similar theoretical implications for the decision-making process in a negotiation setting. This section focuses on how situation awareness is used in a negotiation setting and accordingly introduces a situation awareness model for dynamic evaluation of the offers during a negotiation.

9.3.1 Situation Awareness Model (SAM)

In order to have the software agent for the NNC act coherently and to able to generate and respond to different proposals, the needs/preferences of the NNC on possible actions in certain situations need to be understood. Since the experts' preferences on actions in traffic network control can change with respect to time, situation and environmental factors, a dynamic and elaborate evaluation of offers is required. Furthermore, expected consequences of the possible actions play a key role in train traffic and network controllers' decision making.

This chapter proposes a situation awareness model in terms of possible actions for a given situation and ranking of those actions, based on the desirability of the projected expected consequences. Consequently, dynamic evaluation of offers can be made according to the preferences induced from situation awareness model. The proposed model of situation awareness is based on the theoretical approach proposed by McGuiness and Foy (2000), which extends Endsley's three-level situation awareness definition by adding a fourth level. According to this approach, SA consists of four levels: 1. the perception of the elements in current situation, 2. the comprehension of these elements in the current situation, 3. the projection of these elements in future states and 4. the evaluation of a subset of available projected actions. In this study, the evaluation of the projected future states in a given situation is achieved by ranking the projected future states. In the model, the ability to evaluate the outcomes includes the attitudes, motivation and goals of an operator to yield the final level of situation awareness, prior to decision-making. Figure 9.1 shows the four-level SA consisting of the following modules:

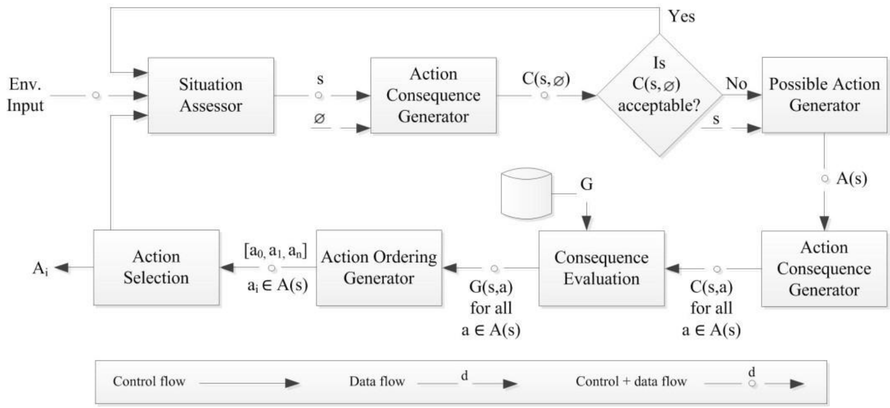


Figure 9.1: Situation Awareness Model (SAM).

- ***Situation Assessor*** *(SAss)* takes the environmental inputs (e.g. a malfunction of a signal), and outputs the current situation s where S is a set of all possible situations and s is a variable over S. For example, the situation "switch 1021 broke down between Schiphol and the junction to Amsterdam Lelylaan and Amsterdam Zuid" may represent the current situation. An environmental input can be anything that may cause a change of the current situation. For instance, "an accident in Amsterdam Central" will change the current situation. The current module refers to the first two SA levels (perception and comprehension).
- ***Possible Action Generator*** *(PAG)* considers the current situation, s and generates all applicable actions in that situation, A(s) where A(s) = {a ϵ A | app (a, s)}. Here, A is a set of all possible actions; a is a variable over A and app(a, s) denotes that action a is applicable in situation s. For example, rerouting a train between Amsterdam Central and Leiden

Central station might be a possible action to be taken in order to deal with the current situation. PAG includes the comprehension of the situation, which relates to the second SA level (comprehension).

- ***Action Consequence Generator*** *(ACG)* generates expected consequences (future projected states) of each applicable action in the current situation s. For all a ϵ A(s), it outputs C(s, a) = {c ϵ C | pcons (s, a, c)} where C is the set of all possible consequences, where a consequence is described by a subset of S (C= $\wp$(S)). Let c be a variable over C, and let pcons (s, a, c) denote that c is a possible consequence of action a in situation s. The module ACG focuses on the third situation awareness level (projection).
- ***Consequence Evaluator*** *(CE)* evaluates all possible consequences generated by *ACG,* C(s, a) and outputs which goals in G are satisfied by them. For all a ϵ A(s) it outputs G (s, a) = { c ϵ C(s, a), g ϵ G | satisfies (c, g)} where G is a set of goals and satisfies(c, g) denotes that consequence c satisfies goal g. The *Consequence Evaluator* relates to the fourth situation awareness level (evaluation).
- ***Action Order Generator*** *(AOG*) orders all applicable actions in the current situation, A(s) with respect to how well they satisfy the goals in G (from most preferred action to least preferred one) and returns an ordered sequence of actions O(s), where $O(s)_i$ refers to the i^{th} element of the list, $i \in$ N, counting from 0. The first element of the list is the most preferred, and the last element is the least preferred applicable action and for all i and j, where i < j, $O(s)_i$ is preferred over $O(s)_j$. There are several ways to order the applicable actions. In our model, the actions are ordered according to either the goal-based *lexicographic* criterion or the goal-based *cardinality* criterion defined in Visser, Aydoğan, Hindirks, & Jonker (2012). When goals in G have priorities such as g is more important than g’ where g, g’ ϵ G, the lexicographic criterion is taken into account. That is, the preference on actions is determined by the satisfied goals with the highest priority. For example, if the expected consequence of action a satisfies goal g whose priority is higher than the priority of g’ that is satisfied by the expected consequence of action a’, then the actions are ordered as [a, a’] where a is preferred over a’. When more than one goal is satisfied by taking an action a, the goal whose priority is the highest in comparison is considered. In case that the lexicographic criterion assigns all goals the same priority, then the ordering of actions is done by adopting the cardinality criterion, where the number of goals satisfied by each consequence is taken into account. If the number of goals satisfied by taking action a is greater than the number of goals satisfied by action a’, then a is preferred over a’. *AOG* is related to the fourth situation awareness level (evaluation).

SAM starts by interpreting the current situation through environmental inputs (SAss module). For optimization purposes, it checks whether taking no action is acceptable. To do this, SAM projects the expected consequences of not acting in the current situation and checks acceptability using a knowledge base. For example, a future in which trains are delayed by more than five minutes is not acceptable. In case doing nothing is not acceptable, SAM searches all possible admissible actions in the current situation (PAG module) and projects their expected consequences (ACG module). Afterwards, it evaluates the consequences with respect to which goals they satisfy (CE module) and accordingly orders the actions based on the satisfaction of goals (AOG module). The ordered actions will be used in decision making process by the negotiating agent.

9.3.2 Operationalization of SAM for a Specific Situation

Building the situation awareness model described above requires significant operational data for each specific region and railway traffic network. The following procedure has been developed and used to acquire the data:

- Use the automatic logging data that is obligatory for most rail traffic controllers to collect the environmental input, situation assessments and transformations, possible alternative actions, expected consequences of actions, and the decisions made by the controller;
- Analyze and select the collected data;
- Perform a Goal-Directed Task Analysis (GDTA). This technique is a variation of a cognitive task analysis (see e.g. Endsley, Bolté, & Jones, 2003) and is specifically designed to uncover the operators' dynamic information needs (i.e. situation awareness requirements) by identifying the operators' major goals, sub goals, related decisions and necessary information for the decision;
- Align the set of goals G from the Goal-Directed Task Analysis (GDTA) to the SAM model;
- Interview (some) human actors to collect lacking in-depth information about alternative actions, expected consequences, priorities over goals, relations between goals and consequences, and preferences over actions in relation to expected consequences and goals.

9.3.3 Test Case Results

To test the feasibility of implementing a SAM for a specific situation, the aforementioned information was acquired for a test case on the busy Amsterdam-Utrecht- Amersfoort triangle. For a clear understanding, the example is simplified to one situation for one particular area of the infrastructure. To do so, a manual version of retrieving information was undergone with the assistance of one of the National Network Controllers working in The Netherlands. Only the related part of SAM for the current situation denoted by s is described as follows in Figure 9.2.

- s: the current situation after taking actions such as rerouting and turning to deal with the problem of the congested train traffic at Schiphol station during morning peak hours
- A(s) is the set of alternative applicable actions, $\{a_1, a_2, a_3\}$ where
  - a_1 is to cancel a train series at Utrecht Central
  - a_2 is to cancel a train series at Amersfoort
  - a_3 is to cancel a train series at Maarssen
- C is the set of possible consequences, $\{c_1, c_2, c_3, \ldots\}$ where
  - c_1: the track space of the cancelled train is occupied, limiting the infrastructure capacity of the station where the action has been taken.
  - c_2: similar loss of infrastructure capacity of a station as with c_1. However, the infrastructure capacity on a more critical area (Utrecht Central) is relieved with an extension for a longer term.
  - c_3: similar loss of infrastructure capacity on a station as with c_1. However, the infrastructure capacity on a more critical area (Utrecht Central) is relieved, but canceling train series at Maarssen is not allowed anymore.
- $G = \{g_1, g_2, \ldots\}$ is the set of goals used to rank the actions. For this example, all goals have equal priorities. Specific goals are
  - g_1 = relieve the disrupted area (satisfied by c_1, c_2, c_3)
  - g_2 = minimize secondary delays (satisfied by c_1, c_2, c_3)
  - g_3 = include buffers in infrastructure capacity (satisfied by c_1, c_2)
  - g_4 = keep the infrastructure capacity of stations high (satisfied by c_1, c_2)
  - g_5 = reduce the infrastructure capacity load of major corridors (satisfied by c_2)

Action Consequence Generator:
- $C(s_3, a_1) = c_1$; $C(s_3, a_2) = c_2$; $C(s_3, a_3) = c_3$

Consequence Evaluator:
- $G(s_3, a_1) = \{g_1, g_2, g_3, g_4\}$; $G(s_3, a_2) = \{g_1, g_2, g_3, g_4, g_5\}$; $G(s_3, a_3) = \{g_1, g_2\}$

Action Order Generator:
- Ordered action list = $[a_2, a_1, a_3]$

Figure 9.2: An illustration of a test case.

The applicable actions are ordered with respect to how many goals are satisfied by the expected consequences since it is assumed that all goals have the same priority for this scenario. The expected consequence of action a_2 satisfies five goals where those of a_1 and a_3 satisfy four and two goals respectively. The most preferred action is a_2. The ordered list is identical to the applicable action list ordered by the human NNC according to his preferences.

9.4 Negotiating Software Agents in Gaming Simulations

The focus in this study is on the multiparty negotiation in which the software agent representing NNC denoted as A_{NNC} negotiates with a number of human RNC actors on which action needs to be taken. The number of RNC human actors may vary according to the situation.

The proposed negotiation protocol is inspired from the single text negotiation protocol (Klein, Faratin, Sayama, & Bar-Yam, 2003), where the mediator agent generates offers and asks the negotiating parties for their approval or disapproval of the offers. The negotiation outcome is determined through the votes of the parties during the negotiation. In that protocol, the mediator's role is to help the negotiating agents reach a consensus. To do this, the mediator generates the offers based on the negotiating agents' responses, regardless of its preferences. That is, it generates the first offer randomly, and for further offers it modifies the most recently accepted offer by all agents, by exchanging one value with another value randomly in the offer.

9.5 Discussion and Conclusion

This study introduces a human-agent negotiation approach in which a software agent negotiates with multiple RNC human agents on behalf of the NNC on the actions to be taken in a certain situation, during gaming simulations on operational innovation in railway traffic control. The dynamic evaluation of actions to be offered by the NNC is modeled through a situation awareness model (SAM). The model is a straightforward approach to deal with the complicated decision-making process. Ordering applicable actions in certain situations involves a reasoning task on consequences and goals. Furthermore, this approach enables the software agent to understand the preferences of the human NNC and to explain why one action is preferred over another.

Using the input of a human NNC, a particular situation was constructed in the situation awareness model. This shows that collecting the data required is feasible and yields meaningful results. In future studies, the accuracy of the model has to be evaluated with more situations that also involve more alternative actions. After construction of the model for a particular situation, the ordering of actions gained from the model can be compared to the real ordering of actions of multiple human NNCs. In the case of an inconsistency, the underlying reasons may help to understand the relation between consequences and goals, update the goal hierarchy if necessary, and find possible hidden factors that affect the preferences. Furthermore, in the current approach, direct transfer is used from actions to consequences based upon interviews with the NNC. Over time, the consequences of actions in gaming simulations can be observed and a probabilistic model can be built by associating probabilities to the consequences.

Another extension to the current approach is to apply the model of situation awareness for the RNCs in addition to the NNC, when one or more RNC human agents are not be available in gaming simulations. Moreover, shared situation awareness can be used in order to improve the negotiation outcome.

Finally, next steps are to investigate the influence of using software agents in low- tech multi-actor gaming simulations on their performance, and to obtain insights about the applicability of the software agents in this setting.

References

Al-Ibrahim, A. (2010). *Dynamic Delay Management at Railways: A Semi-Markovian Decision Approach*. Amsterdam: Thela Thesis.

Croft, D. G., Banbury, S. P., Butler, L. T., Berry, D. C. (2004). The role of awareness in situation awareness. In S. Banbury & S. Tremblay (Eds.), *A Cognitive Approach to Situation Awareness: Theory and Application* (pp. 82-103). Bodmin, Cornwall: MPG Books Ltd.

Endsley, M. R. (1988). Design and Evaluation for Situation Awareness Enhancement. *Proceedings of the Human Factors Society Annual Meeting, 32*(2), 97-101.

Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors, 37*(1), 32-64.

Endsley, M. R., Bolté, B. & Jones, D. G. (2003). *Designing for Situation Awareness*. New York, NY: Taylor & Francis Group.

Gentner, D. (2002). Mental models, Psychology of. In N. J. Smelser & P. B. Bastes (Eds.), *International Encyclopedia of the Social and Behavioral Sciences* (pp. 9683-9687). Amsterdam: Elsevier Science.

Hamilton, W.I., & Clarke, T. (2005). Driver performance modelling and its practical application to railway safety. *Applied Ergonomics, 36*(6), 661-670.

Hoogendoorn, M., Van Lambalgen, R., & Treur, J. (2011) Modeling situation awareness in human-like agents using mental models. *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI'11)*.

Jonker, C. M., Van Riemsdijk, M. B., Vermeulen, B., & Den Helder, F. (2010). Shared mental models: A conceptual analysis. *Proceedings of International Conference on Autonomous Agents and Multiagent Systems (AAMAS).*

Klein, M., Faratin, P., Sayama, H., & Bar-Yam, Y. (2003). Protocols for negotiating complex contracts. *IEEE Intelligent Systems, 18*(6), 32-38.

Lee, Y. H., Choi, Y. C., Choi, S. H., & Ujimoto, K. V. (2007). Analysis of survey data on situation awareness of helicopter pilots: The case of helicopter accidents in Korea. *Proceedings of TRB Annual Meeting 2007*, TRB 07-0346.

McGuinness, B., & Foy, L. (2000). A subjective measure of SA: The crew awareness rating scale (CARS). *Proceedings of the First Human Performance, Situation Awareness, and Automation Conference*.

McLeod, R. W., Walker, G. H., & Moray, N. (2005). Analysing and modelling train driver performance. *Applied Ergonomics, 36*(6), 671-680.

Meijer, S. A. (2012). Gaming simulations for railways: Lessons learned from modeling six games for the Dutch infrastructure management. In Perpinya, X. (Ed.), *Infrastructure Design, Signaling and Security in Railway*. Croatia: IntechOpen.

Meijer, S. A., Mayer, I. S., van Luipen, J., & Weitenberg, N. (2011). Gaming rail cargo management: Exploring and validating alternative modes of organization. *Simulation & Gaming, 43*(1), 85-101.

Meijer, S. A., Van der Kracht, P., Van Luipen, J. J. W., & Schaafsma, A. A. M. (2009). Studying a control concept for high-frequency train transport. *Infrastructure Systems and Services: Developing 21st Century Infrastructure Networks* (pp. 1–6). Piscataway, NJ: IEEE.

McCarley, J. S., Wickens, C. D., Goh, J., & Horrey, W. J. (2002). A computational model of attention/situation awareness. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 46*(17), 1669-1673.

ProRail. (2012). *Over Prorail* [About ProRail]. Retrieved from http://www.prorail.nl/Over%20ProRail/Pages/default.aspx

Rousseau, R., Tremblay, S., & Breton, R. (2004). Defining and modeling situation awareness: A critical review. In S. Banbury & S. Tremblay (Eds.), A cognitive approach to situation awareness: Theory and application. Bodmin, Cornwall: MPG Books Ltd.

So, R., & Sonenberg, L. (2004). Situation awareness in intelligent agents: Foundations for a theory of proactive agent behavior. *Proceedings of the IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT '04)*.

Visser W., Aydoğan, R., Hindriks, K. V., & Jonker, C. M. (2012). A framework for qualitative multi-criteria preferences. *Proceedings of the 4th International Conference on Agents and Artificial Intelligence (ICAART 2012)*.

10 The value of operator knowledge in participatory system design

Abstract
With every decision to change physical infrastructure, e.g. by removing level crossings or adding more tracks, strategic decision-makers shape the future state of the railways. However, when it comes to its details and their implications, knowledge of operators is especially key in finetuning the system's design and sustaining its performance. This chapter focuses on the role of gaming simulation in the context of a socio-technical system design. It also focuses on the role and knowledge of the operator. Extracting operator knowledge is not trivial when for instance validity comes into play. A multi-disciplinary framework is introduced that connects the implications of extracting specific operator knowledge (1) in relation to cognitive components such as mental models and/or situation awareness, and (2) in relation to a system design process. The role of validity and selection of participants are discussed, as well as the implications for gaming simulation design and strategic decision-making.

Keywords: Complex systems; design; decision-making; participatory design; command and control operations

10.1 Introduction

The design of a socio-technical system is often seen as complex due to their economic, social and physical elements (De Bruijn & Herder, 2009; Herder, Bouwmans, Dijkema, Stikkelman & Weijnen, 2008). As also described in previous chapters, one of the core challenges for the railway sector in the Netherlands is the management of the overall need to increase railway infrastructure capacity. Solutions can be found in process innovations, such as those in the timetable (e.g. higher frequency of trains) and/or in the organization (e.g. an alternative allocation of geographical responsibility for operators), but also changes in the infrastructure (e.g. fewer switches).

The landscape of changing and shaping the railway system is influenced by multiple actors on different decision-making levels, i.e. in the strategic, tactical and operational layer (Fountas, Wulfsohn, Blackmore, Jacobsen, & Pedersen, 2006). Strategic decision-makers focus on problem identification, goal formulation, generation of alternatives, and evaluation or selection (Schwenk, 1984), often from a management or project team role, while tactical decision-makers focus on estimates of future projections (Dexer, Ledolter, & Wachtel, 2005), such as employees with a timetable planning role. Finally, operational decision-makers are responsible for the daily operational activities. These operators hold the knowledge and skills in daily practice to influence parts of the system. Railway traffic operators work in a safety-critical and time-demanding environment, often accompanied with high uncertainties. As such, operators develop specific skill sets enabling them to perform in such conditions. The foundation for this lies in well-developed mental models from which situation awareness can be obtained (for an elaborate description of these concepts see Chapter 3). Knowledge from these operators can support strategic and tactical decision-makers to investigate the designs of future railway configurations. Using this knowledge, they may finetune their design(s) before a final one is put into operation. In doing so, the final design will likely have a higher level of acceptance and performance.

Gaming simulation has so far been used as a tool within the Dutch railway infrastructure manager ProRail, to address questions that the organization wishes to investigate and to support in the design of a changed operational environment. Gaming simulation has also been used to teach/train railway traffic operators in respect to the implementation of these changes, i.e. to prepare them for a disruptive change in the railway system. In essence, the underlying goal with games as a tool has been to reduce uncertainty in complex system-level changes (Meijer, 2015).

One constant factor is that gaming simulation always involves the knowledge of operators in a certain manner, whether it is to create new knowledge through training or to extract operator knowledge for designing or testing. The process of obtaining operator knowledge is not trivial however, as knowledge can be derived from different cognitive components, such as from operator's mental models (e.g. general, abstract knowledge during conversations) or from operator's situation awareness (e.g. applied knowledge in an environmental context such as a real-time human-in-the-loop simulation). A certain level of validity can be related to the obtained knowledge from a mental model or situation awareness. Another factor is the prerequisite of triggering operator knowledge in certain project stages of the design of a changed operational environment.

In sum, the current chapter builds on a multi-disciplinary approach: (participatory) system design, decision-making, human factors/cognitive engineering and gaming simulation with railway traffic control as the domain. The aim is to advance the system design field by proposing a multi-disciplinary framework that involves users, i.e. railway traffic operators, throughout the different stages in a system design process. This involvement can be facilitated through gaming simulation as a tool, in which knowledge from operators can be extracted that is based on their mental models or situation awareness. This knowledge from operational decision-makers can be used as input in the development of the system design and across different project design stages, while the overall design process can be steered by strategic decision-makers who focus on the needs of the design process and can manage what and when operator knowledge is needed. Game designers face the challenge to create a valid simulation environment that reflects the required conditions to extract operator knowledge. The current framework focuses on the system design process and its different stages and brings together the value of operational knowledge from operational decision-makers in different stages of a system design process and the process management of strategic decision-makers across a system design proces. The framework also touches upon the topic of validity when extracting operator knowledge by inducing a certain cognitive state through the facilitation of cognitive components such as mental models and/or situation awareness, which varies for each stage of the system design process.

10.2 Perspectives on changing the operational environment

This section addresses the different theoretical approaches of system design through the disciplines of human factors/cognitive engineering, participatory design, railway system design and gaming simulation.

10.2.1 Human factors in engineering design

There are several views in the human factors field that approach the system design process. More traditional views entail a linear process, in which user inputs are profound in the design and evaluation stages and the focus is on fitting the person to the system (Czaja & Nair, 2012; Eason, 1991). Other system design approaches are for instance the socio-technical systems approach, participatory ergonomics and user-centered design focuses on active user participation (Czaja & Nair, 2012).

A model that outlines different stages in the design process and includes human factors components for each stage is depicted in Figure 10.1. This Human Factors Engineering Program Review model is used by the U.S. Nuclear Regularatory Commission (Joe, 2017). The model identifies the steps of (1) planning and analysis, such as conducting task analyses, and controlling staffing and qualification, (2) design of e.g. interfaces, procedure development or training program, (3) verification and validation to test solutions and (4) implementation and operation that focus on the design implementation and the monitoring of human performance.

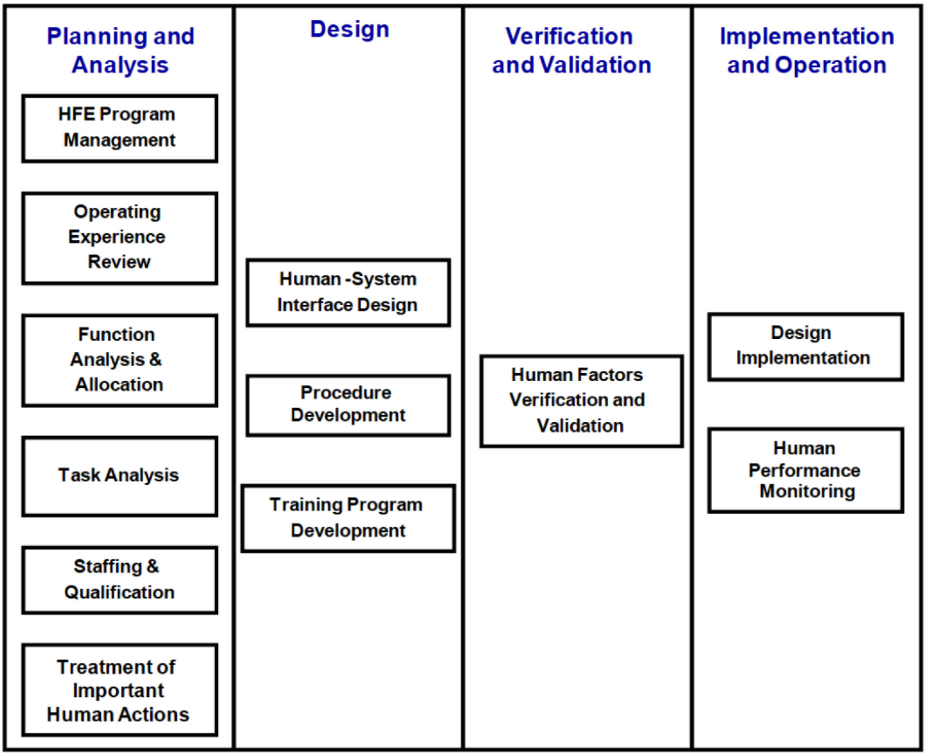


Figure 10.1: A human factors framework in engineering design (Joe, 2017).

10.2.2 Participatory design

Participatory Design (PD) is an approach that was developed in the 1970s and has been a maturing field of research and practice since the 1990s (Kensing & Blomberg, 1998; Spinuzzi, 2005). With PD, user involvement is seen as a key component in the development of a (change) process or product (Johnson, 1998). This stems from an underlaying value that people that are affected by a decision should be entitled to influence the process (Schuler & Namioka, 1993).

Spinuzzi (2005) argues that PD is not only about the design but also about research as it has its own paradigm, methodology and research design. Hence, the object of study is describing and defining users' knowledge. Spinuzzi distinguishes three basic research design stages that are iterative: (1) initial exploration of work, (2) discovery process and (3) prototyping. In the 'initial exploration of work' phase, the designers familiarize themselves with the technologies, procedures and users. In the 'discovery process' stage, designers and users together uncover the goals and values of the project. In the 'prototyping' phase, the prototype is iteratively developed by the designers and users.

The notion of PD to actively involve users, as they can co-interpret, co-design and co-analyze data, is noteable (Spinuzzi, 2005). Through this approach, the validity of analyzing and applying user's knowledge is specifically addressed as observers may not fully relate to the interpretation of user's knowledge; therefore creating ethnographical research of their own (Forsyth, 1999).

Eliciting tacit knowledge from users is a core goal in participatory design (Spinuzzi, 2005). Chapter 1 described the relation between knowledge and mental models through the equipment-task-team interaction-team model and the declarative, procedural and strategic model. Chapter 4 addressed and illustrated the implications of tacit, implicit knowledge versus explicit knowledge. Although tacit knowledge is very difficult to grasp on a practical level, its value to enable it as a creative source should be recognized (Krogh, Ikujiro & Kazuo, 2009). Additionally, Krogh, Ikujiro and Kazuo (2009) posit that knowledge creation requires the necessary context, such as a situation or people.

10.2.3 Railway system design

The design process in railway operations consists of a number of steps, in which various and different subject matter experts can be involved. A model that describes the design process in railway operations is the cascade model (see Figure 10.2, Van den Top, 2010).

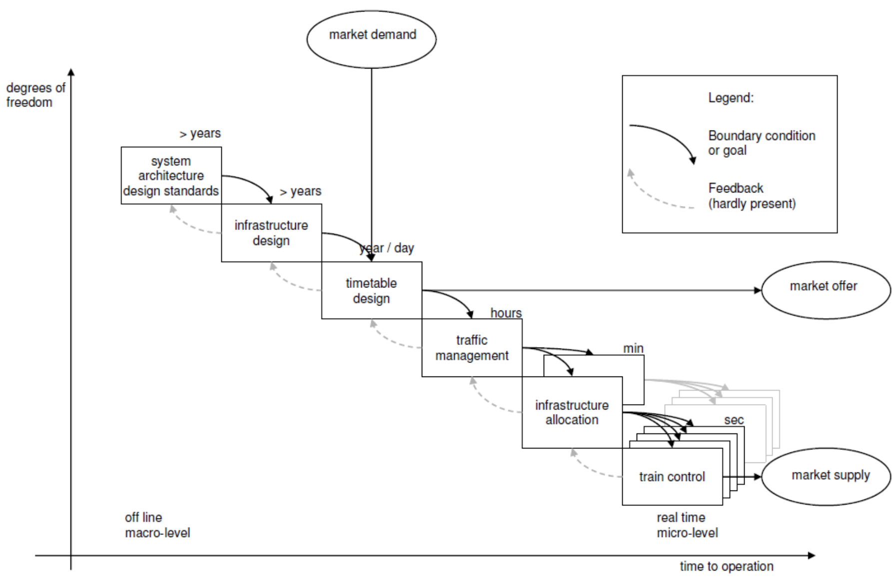


Figure 10.2: Cascade model (Van den Top, 2010).

This model captures the process ranging from a generic, abstract architecture of the railway system to real-time operation of train control, in which multiple stakeholders in the railway system (e.g. transport service providers and railway infrastructure manager) are involved.

The model also includes the parameter regarding the amount of degrees of freedom that the stakeholders have. The degrees of freedom of stakeholders decrease when the design process matures. The design process starts with a system architecture design standard that is started years ahead and ends with its implementation in a real-time train and train traffic operations. As such, effects of the design in earlier phases in the time process may have a better fit with computer simulation tools, while (multi-actor) gaming simulation may we wished for when testing the effects in later stages of the design process.

10.2.4 Railway system elements

In order to design gaming simulations and particularly to understand how the different knowledge types can be extracted in the current case with railway traffic operators, the railway systems needs to be unraveled. Table 10.1 presents an overview of three theoretical approaches on elements that delineate the railway system (Lo, Van den Hoogen, & Meijer, 2014). Following the socio-technical framework, system design can either focus on the system itself, i.e. the hardware, on the rules and institutions, i.e. the software, and on human actors, i.e. the liveware (Geels, 2004). Similar notions are found in the P-S-I framework (Reich & Subrahmanian, 2015), where design is portrayed as taking place on the level of a product (what is being designed), the level of the social sphere (who designs) and the level of institutions (which rules govern the design process). If this is translated to the design content of a railway system instead of to the design process, the P-S-I framework would frame design of a system as either changing the product space (P), the social space (S) or the institutional space (I). Both frameworks however are relatively abstract and allow many different systems to be studied. Goverde (2005) typified the dimensions that characterize the railway system. In this model the design process is of a hierarchical nature, where hardware elements such as tracks, signaling and timetables are designed first and subsequently serve as input for downstream design efforts on for instance crew scheduling and traffic control procedures.

Table 10.1: Three conceptions of the focus of intended design changes and illustrations specific to the railway system.

Geels (2004)	Reich and Subrahmanian (2015)	Goverde (2005) slightly adapted	Example
Socio-technical system	Product Space (P)	Railway network (adapted to include traffic control (1) hardware-part)	Connections, doubling tracks, power supply, signaling, switches, control panels, information systems
		Line planning	Frequencies, service patterns and connections
		Timetable	Detailed planning, clock-face planning
		Rolling stock circulation	Composition of trains, length of service
Humans	Social Space (S)	Crew scheduling	Changing crew schedule
		Traffic control (2) liveware-part	Roles, responsibilities, knowledge
Institutions	Institutional Space (I)	Traffic control (3) software-part	Procedures for handling disruptions, rules, coordination mechanisms

10.2.5 Gaming simulation

Game simulations exist in different types of representations (e.g. analog and/or digital), varying numbers of participants (e.g. a single actor, a team or even a network of actors) and in different designs depending on its purpose. Chapter 3 discussed multiple types of gaming simulation, in which they can be applied for teaching, research, design and policy (Grogan & Meijer, 2017).

The value of gaming simulations is often easily recognized by an organization through five criteria: complexity, communication, creativity, consensus and commitment, as described by Duke and Geurts (2004). These five criteria can be identified in gaming simulation with participants, in which *complexity* is made more explicit and therefore also more tangible, *creativity*-stimulating power is introduced because simulations are often experienced as enjoyable and productive, and through *communication* between participants a better understanding is achieved which leads to more *consensus* and a higher *commitment*.

However, depending on the strategic course that an organization holds, gaming simulations can be employed with a different emphasis. When the strategic course of the organization is governed by a systems perspective, the decision-making process is marked with a rational characteristic: decisions are taken in a sequence of successive phases (De Bruijn & Herder, 2009). On the other hand, the actor perspective takes the different interests and perceptions of many actors into account. As such, the decision-making process is interactive and facilitates each actor's need for the level of specificity of the problem, the required information and the accuracy of the information. Russ (2010) discusses the application of gaming simulations in change implementations and distinguishes two types of approaches that can be linked to the systems and actor perspectives on governing socio-technical systems. The programmatic view on gaming simulation follows the system perspective, in which there is a top-down manner of communicating about the planned changes by management and gaming simulation is used to convince employees of the necessity of these changes. The participatory view on gaming simulation follows the actor perspective, in which dialogical communication is stimulated between employees so that they together can shape the change program and find more consensus in the decision-making.

10.3 Inducing a cognitive state to extract knowledge

At any given time, operators can be queried to collect their knowledge on any specific subject. However, if validity of this collected knowledge is an important requirement, the conditions in which the queries take place are of importance. One of the conditions is to facilitate a certain cognitive state, i.e. a setting in which cognitive components such as mental models and situation awareness can be triggered.

Mental models can also be seen as a form of knowledge. They are also described as knowledge structures, organized knowledge, a mental represententation of knowledge (Cannon-Bowers, Salas and Converse, 1993; Klimoski & Mohammed, 1994). Classifications of different types of knowledge are for instance the Declarative, Procedural, Strategic (DPS) model (Rouse, Cannon-Bowers & Salas, 1992) and the Equipment/technology, Task/job, Team-interaction and Team (ETTT) model (e.g. Cannon-Bowers, Salas & Converse, 1993). Following the DPS model, knowledge can be distinguished into three types: declarative, procedural and strategic. To briefly recapitulate from Chapter 1: declarative models refer to knowledge of facts, rules and relationships (knowledge of what). Procedural models refer to the timing and sequential type of knowledge (knowledge of how). Strategic models refer to knowledge that forms the basis for problem solving (knowledge of the concept and contingency plans). In line with the ETTT model, different knowledge types are: technology/equipment (knowledge about technology and tooling), job/task (understanding of e.g. strategies and work prcedures), team-interaction (awareness of e.g. role interdependencies, responsibilities) and team types of knowledge (understanding of teammembers' skills or preferences) (Cannon-Bowers, Salas and Converse, 1993). The technology/equipment model is related to the technology and equipment that is used to execute tasks in a team. As mental models form the fundament for knowledge, they also can be extracted in an abstract context, i.e. in an environment without time constraints and/or environment cues. The focus on mental models is therefore particularly interesting when new knowledge needs to be developed. An example of this are design games, which may require creative, open-minded and out-of-the-box thinking.

Situation awareness focuses on the perception, understanding, and prediction of elements in the environment that the operator is encountering (Endsley, 1988). Situation awareness builds on developed mental models, in which these knowledge structures are relevant in a certain situation or environment in which the operator is acting. For instance, a train traffic operator can perceive that the status of a train route turned red. He/she understands that the red status implies that the automatic route setting functionality has been switched off. The operator predicts that not acting timely on the red status means that the train will not proceed with its route, which will cause a delay for this train and possibly also for other trains. When an operator is able to foresee different implications of an event, he or see has reached a high situation awareness. Facilitating for situation awareness ensures that more detailed and accurate knowledge can be elicited and extracted in relation to the situation at hand.

In the current work, it is posited that mental models are always active in an operator's cognitive state as they are fundamental knowledge structures. For instance, an interview with a railway traffic operator about railway traffic control tasks at an air traffic controller's workspace setting does not necessarily impact the extracted knowledge, because the knowledge is drawn from declarative, procedural and/or strategic knowledge. Situation awareness becomes relevant in circumstances when environmental cues are present, e.g. in the operational environment itself. A cognitive state where a representative situation awareness can be developed is when the operator is immersed in an environment with high levels of cues that are similar to the actual work environment. In these conditions operator knowledge is more likely to be detailed and closer to the situation at hand. Therefore, a higher validity of this knowledge can be expected to be obtained in relation to the actual/real work environment.

10.4 The role of operator knowledge in different stages of a design process

Infrastructural projects often tend to be complex and large in scale, however these projects, like every project, undergo similar stages in the design process. Every project stage or phase in the design processes of a socio-technical system triggers the need to answer a specific question and requires specific knowledge that is needed from the operational decision-maker or user (Spinuzzi, 2005). The source of knowledge can be abstract, such as theory-based knowledge that is triggered in the form of the operator's mental model. On the other hand, applied knowledge that is part of a scenario with the actual situation awareness and performance of the user, entails more specific and representative knowledge, and can also be applied to collect data and derive output that can be used by strategic and tactical decision-makers. As different stages may require different types of knowledge, this section introduces a framework to identify which type of knowledge is needed in different parts of the design process.

10.4.1 A framework for connecting operator's cognitive state to different stages in the design process

The current framework connects existing approaches, frameworks and perspectives, Congregrating to the following framework as depicted in Figure 10.3. The framework also features the assignment of relevant cognitive components to each gaming simulation purpose to trigger a certain cognitive state in operators.

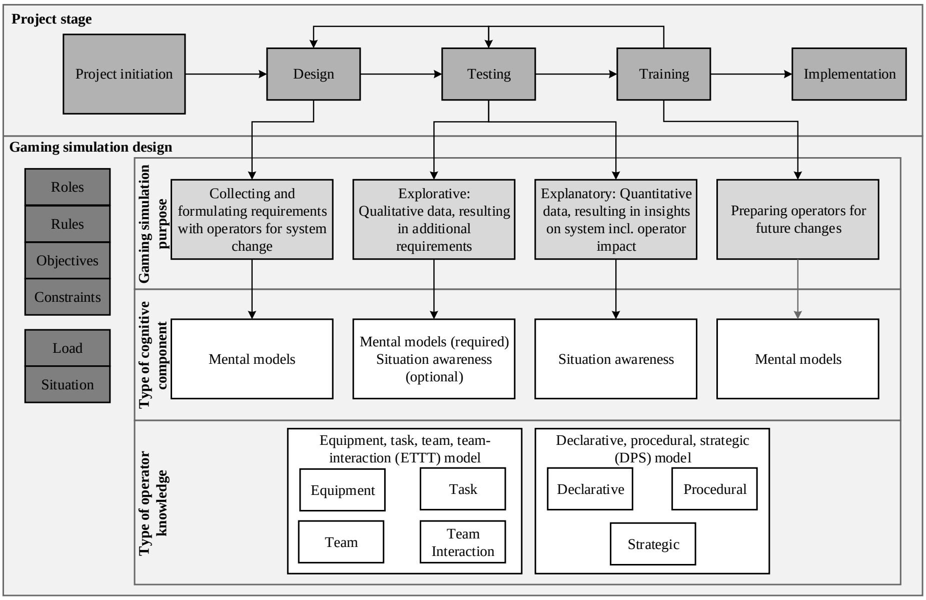


Figure 10.3: Components in a participatory design process using gaming simulation and the related required operator's cognitive state.

The following components can be identified in the framework:

- *Project stage.* Often participatory design frameworks strongly focus on the user when designing and testing a component of the solution. However, the project that instigates a change in the system does not end with only a final design of the solution; it will also have to implement the change in the organization. As such, the framework in Figure 10.3 depicts project stages that include (1) the project initiation that provides the business case for the system (re)design, (2) the design of the concept together with e.g. its problem statement, requirements specification, (3) testing of the concept which is strongly aligned with human factors verification and validation activities, (4) training of users when the design is to be implemented and (5) the project implementation itself. Iterations between stages is possible as tested designs may need adaptation and return to the design cycle, and when training of users may be needed before conducting thorough exploranatory testing games.

- *Gaming simulation design*. In addition to the fact that the requirements for validity of the gaming simulation are predefined by their purpose and required type of cognitive component, the gaming simulation design is also influenced by its precise scope, such as the roles and rules, but also the available time and resources (Meijer, 2008). Restrictions in time and financial resources may especially influence the representation of the gaming simulation, e.g. analog versus digital and the amount of game sessions that relates to generalization of the results. Careful design considerations should be made to find a trade-off in the gaming simulation representation needed, as the design could influence the type and level of accurateness of the required cognitive component. For instance, by leaving out environmental cues in the gaming simulation design, operators are unlikely to develop their situation awareness and remain in a cognitive state where knowledge from their mental models is used. Also, by varying the level of detail of the environmental cues, different levels of situation awareness may be invoked, which may differ (higher or lower) from those that operators can usually develop in real-life.
- *Gaming simulation purpose.* Relating to the project stages, strategic decision-makers have different questions and urgencies for the use of gaming simulations. In the design stage of a project, relatively open questions may be formulated, such as 'at which location in the infrastructure should a switch be removed?', opposed to more specific questions posed in a testing stage of the design process within a project, such as 'what is the impact of the removal of switch number 118?'. Additionally, testing/research games, can be distinguished into those with an explorative or explanatory purpose. Whereas explorative testing games focus on hypothesis generation, explanatory testing games focus on hypothesis testing (Lo, Van den Hoogen & Meijer, 2013). The latter type of game requires rigid experimental conditions as explanatory research games investigate causal relations. Towards the end of a project, the needs of strategic decision-makers shift to project implementation. There the need is more focused on a successful implementation of the design, which can only be achieved if the operational decision-making layer is well-informed and prepared. As such, training with gaming simulation can be used to prepare operators.
- *Type of cognitive component.* A third element in the framework is the type of cognitive component that is required to induce a certain cognitive state. As discussed in Chapter 3, gaming simulation purposes can be put in relation to certain basic operator cognitive components, i.e. mental models and situation awareness.

- In a design game, the focus is to cooperatively make sense of the design at hand together with researchers and users (Spinuzzi, 2005). Out-of-the-box ideas and solutions may be wished for. Therefore, a creative state of mind should be triggered, in which a focus on an operator's mental models can be used to obtain operator knowledge, as an open-minded view of both researchers and users are expected. Facilitating a design game for situation awareness would not be needed because extracting knowledge on a detailed contextual level is not yet required.
- Explorative and explanatory research games employ different demands on the level of knowledge specificity that is elicited from operators. As described in Chapter 3, research games can facilitate both mental models and situation awareness. In an explorative gaming simulation, concepts can be roughly tested with a lower level of detail and there is less emphasis on the generalization of the results as these results could lead to hypothesis generation. In this case, obtaining knowledge from well-developed mental models is required while operator's situation awareness does not need to reach a highly representative state. On the other hand, explanatory gaming simulation adheres to principles of hypothesis testing with high validity requirements. This ideally requires a full one-on-one comparison between the operational and the simulated environment of operators, in order to simulate their actual work speed and workload and to investigate causal relations. As such, explanatory games require operators to obtain a high representation level of their situation awareness, in which knowledge can be extracted or derived from. This knowledge is rich in environmental and situational details and therefore most likely to be more accurate in its description, in comparison to knowledge directly derived from mental models. Data from operational system loggings, such as train punctuality can be used in these cases. For both research games, the debriefing after an in-game session provides a setting to discuss operator insights gained during the in-game session and the possibility to collect additional data.
- The relation between the type of cognitive component and training games can vary when it comes to testing far-future states of the railway system. A training game that aims at learning and training of users focuses on mental model development. However, with a full transfer of learning to the actual operational environment, the situation awareness of operators should be comparable to that in the actual operational environment. As such, human performance monitoring measurements can be conducted (Joe, 2017).

It should be noted that the quality of operator's mental models and situation awareness can vary from low to high, depending on the level of experience of operators and the gaming simulation design. For instance, if required information for operators is missing because of the gaming simulation design, a low situation awareness may be expected. Likewise, an inexperienced operator may be still developing his/her mental model and may therefore occasionally develop a lower situation awareness.

- *Type of operator knowledge.* A fourth component is the type of operator knowledge. Operator knowledge should be recognized and leveraged to allow as input in the design process (Spinuzzi, 2005). Knowledge is operationalized in terms of the declarative-procedural-strategic (DPS) model and/or the equipment-task-team team-interaction (ETTT) model. Although challenging, extracting and measuring (tacit) operator mental models or knowledge has been a frequently discussed research topic. There is a plethora of existing measurement techniques. Examples are paired comparison ratings, concept mapping, card sorting, questionnaires and scenarios (e.g. Ellis, 2006; Cooke, Kiekel & Helm, 2001; Mohammed, Ferzandi & Hamilton, 2010; Smith-Jentsch, Cannon-Bowers, Tannenbaum, & Salas, 2008). The challenge posed in the current study therefore does not focus on the measurement techniques, but on the validity of the extracted knowledge.

10.5 Case study

Chapters 4 to 8 introduced and discussed multiple studies with professional railway operators as participants. These studies had a predominant primary game purpose to conduct exploratory or explanatory tests of the concepts, which have been deemed relevant for strategic decision-making (e.g. Meijer, 2015). Two case studies are selected to illustrate how these specific innovations make use of operator knowledge and the gaming simulation design required to create a certain cognitive state for operators.

Table 10.2 depicts an overview of their characteristics in terms of the initial research questions as formulated by strategic decision-makers, the related purpose of the game, the required cognitive component, the type of knowledge in line with the ETTT or DPS model and input obtained from operators that have been used for reports. For each knowledge type, a reference is also made with the P-S-I framework, in which the specific knowledge element is defined as falling under the product space (P), social space (S) or institutional space (I).

Table 10.2: Characteristics of the gaming simulation studies.

Gaming simula-tion	Organizational question	Initial purpose of the game	Required cognitive component	Knowledge following ETTT model	Knowledge following DPS model
PRL game Utrecht 2015	What is the impact of the different workspace designs on railway traffic operations?	Research - explanatory	Situation awareness	*Equipment*: Railway infrastructure (P) *Task*: Geographical area of responsibility (I) *Team*: Roles (S)	*Declarative*: Railway infrastructure (P), *Strategic*: Impact of the changed geographical area for own and other train traffic operators
OV SAAL game	What are the implications of different predefined long-term infrastructural investments?	Research - explanatory	Mental model due to game design restrictions	*Equipment*: Railway track (P)	*Declarative*: Railway tracks (P), *Strategic*: Impact of the changed availability of switches in the scenarios

10.5.1 PRL game Utrecht 2015

10.5.1.1 Project stage

The request to test different geographical workspace designs using gaming simulation (also discussed in Chapter 6) was initiated by the management of one of the regional control centers. The project team from the regional control center developed three different workstation configurations that were deemed viable solutions. As such they were in a testing stage of their project. Two gaming simulation sessions were held, in which four operators participated per session. The use of multiple gaming simulation sessions was employed to retrieve input from multiple operators and to approach a higher number of participants for statistical purposes.

10.5.1.2 Gaming simulation design

The simulation environment was designed to reflect the actual operational environment as much as possible, in terms of roles, operational procedures, tasks, and communication. A detailed overview of the gaming simulation design characteristics can be found in Chapter 6. With stricter validity restrictions in the research design, the different designs were more comparable between each other and the measurements during the gaming simulation would more generalizable in relation to the actual operational environment.

10.5.1.3 Gaming simulation purpose

Considering the availability of the three designs and the focus to test the impact of these designs, the type of gaming simulation that was chosen for this question was a research game. At the time of the gaming simulation design, the type of gaming was chosen as an explanatory gaming simulation due to the organizational need to select one out of three geographical workspace designs and extra research relating to team situation awareness and operator's mental workload. However, in retrospect, because the gaming simulation outcome also resulted in identifying advantages, bottlenecks and suggestions for improvement of each of the designs, it rather was an exploratory game.

10.5.1.4 Type of cognitive component

It was the aim to achieve a similar cognitive state and processes of the operator in relation to an actual operational environment. Hence, the simulation environment also facilitated for the situation awareness of operators, while well-developed mental models were supported through a briefing of each of the new geographical workspace designs.

10.5.1.5 Type of operator knowledge

Operators were impacted by the change in terms of a changed availability of infrastructure, e.g. switches, additional timetable in new geographical area. This impacted multiple elements under the ETTT model: the different geographical area of responsibilities (related to the equipment type), the role distribution of areas for three operators (team type) and the area of focus (task type). In terms of the DPS model, knowledge was relevant in terms of declarative knowledge regarding multiple elements in the railway infrastructure and strategic knowledge regarding the impact of the changed geographical area for own and other train traffic operators.

Three facilitators from the gaming simulation team conducted observations during the scenarios, led the debriefing and collected additional data through questionnaires. Input from the debriefing was used as the main source of data for the report; i.e. the approach was to trigger and activate specific knowledge during the gaming simulation, while qualitatively extracting the knowledge in the debriefing. To extract this knowledge, the facilitators worked together, each with specific expertise that was complementary to the other. One facilitator focused on the technical setup of the simulation environment, another on the development of the research setup and materials and one facilitator led the sessions and the debriefing. A semi-structured format of questions and approach was applied in the debriefing, where there was room for additional questions by all facilitators. Examples of questions to extract knowledge from train traffic controllers were: what do you identify as bottlenecks with the simulated geographical workspace design option? What are the advantages of this specific geographical workspace design? Operators indicated that their experience with the different geographical workspace designs and scenarios in the gaming simulation sessions supported in their reflection during the debriefing.

10.5.2 OV-SAAL game

10.5.2.1 Project stage

Chapter 8 described the OV-SAAL game and its use to support the decision-making of local municipalities and to advise the Dutch government in which large railway infrastructure to invest. A project team developed four types of high potential infrastructural configurations, for instance the expansion of two to four tracks at station Almere. However, it was yet uncertain for the decision-makers what the robustness of each new infrastructural configuration was under different disrupted conditions. As computer simulations were limited in their capabilities to provide performance insights in such disrupted conditions, decision-makers therefore were dependent on expert operators' insights. Additionally, decision-makers were interested in acquiring feedback on the four proposed infrastructural designs from multiple expert operator roles. In total eight operators were included in the session.

10.5.2.2 Gaming simulation design

Each round was characterized with decisions of operators, in which the train traffic state was updated with related decision changes, such as cancellation of trains and the train traffic flow. In these rounds, occasionally in-depth discussions were formed that deviated from the proposed gaming simulation design. Herein, operators also provided input for a new design element in a specific infrastructure case, for instance by adding a switch that could be of use during disruptions. Occasionally this also led to a quick test of proposed designs. As such, insights from operators in terms of the operational consequences due to the availability of railway tracks were not only obtained during the debriefing after each scenario, but also after each round in a scenario. Although a lot of insights were collected, the representiveness of operator's knowledge was not extensively tested to provide results that could be generalized to a high predictive operational state. A detailed overview of the gaming simulation design characteristics is provided in Table 8.1 in Chapter 8.

10.5.2.3 Gaming simulation purpose

Given the goal of testing the impact of the four designs and ultimately the selection of one design, the purpose of the gaming simulation was a research type of game with an explanatory character. However, the dynamics during the gaming simulation itself shifted to an additional focus of introducing new designs. As such, the gaming simulation was a design and exploratory research type of game.

10.5.2.4 Type of cognitive component

In relation to the gaming simulation purpose, the required operator knowledge should be facilitated by operator's situation awareness. However, as only one testing day was available to conduct the four proposed infrastructural designs with operators, the game was designed with step-wise rounds instead of real-time operations. Given this game design choice, a full, representative degree of situation awareness of operators could not be achieved. As such, the gaming simulation focused on extracted knowledge from operator's mental models.

10.5.2.5 Type of operator knowledge

Knowledge of operators specifically addressed the railway tracks in multiple areas. This type of knowledge can be seen as equipment in the ETTT model, as declarative in the DPS model as well as strategic with regards to the impact of the changed availability of railway tracks.

In order to extract knowledge from operators, the following approach was applied: in the gaming simulation operators were able to observe the status of the train traffic flow and discuss with other operators about their decisions after each round. After all four designs were tested and discussed in the gaming simulation sessions, an overall debriefing was held where the advantages and bottlenecks of each design were summarized. These points were substantiated by examples from the gaming simulation session and occasionally by quickly testing certain theoretical assumptions.

10.5.3 Discussion and conclusion

This chapter presented a framework that combines multiple disciplines, namely that of (participatory) system design, human factors/cognitive engineering and gaming simulation in the system design of the railways. The framework advocates the involvement of users throughout the system design process, so that operational knowledge can be captured and iteratively processed in feedback loops, predominantly between the system design stages of designing and explorative testing. Opposed to a classic programmatic top-down system design, a participatory approach therefore can lead to a higher acceptance and a robuster end-state of the system.

The framework also provided insights in which level of abstraction operator knowledge or cognitive component should be used throughout the system design. It can be noted that in earlier explorative stages of a design process, abstract knowledge from operators such as their mental models is suitable to be elicited. Analog and abstracted gaming simulation design can be applied in this stage. In stages of the design process where hypothesis testing is needed, a high validity of the gaming simulation is required, quantitative outcomes are more likely wished for and more representative levels of situation awareness and mental workload are required to be elicited. Extracting operator knowledge in the latter condition would lead to a level of abstraction that holds context-rich, detailed and representative knowledge. However, there are a number of factors that could influence an operator's situation awareness and mental model, which are operator's experience and the gaming simulation design. For instance, due to design choices made in the OV-SAAL game it was not possible for operators to gain a representative degree of situation awareness as the gaming simulation did not run real-time, but in rounds of 15 minutes and the graphical interface in the gaming simulation was new to some operators. A high level of operator experience was important in both case studies, where operators were required to quickly adapt to the new situation and update their mental models. However, there may be factors other than operator experience that influence the adaptation speed, such as motivation and attitude. The previous gaming simulations showed that operators need to have an open mindset towards gaming simulation as a tool and to be interested in being part of the design process. Especially in a design game, in-depth experience and out-of-the-box thinking may be important success factors. All in all, the selection of participants may particularly be of importance in the success of the gaming simulation in design and exploratory games.

10.5.4 Application of the framework by strategic decision-makers

Strategic decision-makers most likely do not have detailed knowledge on operational dynamics and their impact. However, strategic decision-makers shape the design process by allocating its scope, time and resources. The framework introduced in this chapter can support strategic decision-makers to identify the necessity of operator's knowledge depending on the stage of the design process. Iteration of project stages in a participatory system design process are common, moreover even wished for. As such, introducing a gaming simulation for the first time in a final testing stage, close to implementation, will not be well-received by users as little can be done with their input.

Further on, the framework can support strategic decision-makers in the formulation of their research questions and the research at hand by taking the role of validity and accuracy into account when extracting operator knowledge in different stages of the system design. What do strategic decision-makers expect of the results? To what extent should these results be generalizable? What are the exact research questions? These questions seem straightforward, but often are not well-defined and formulated by strategic decision-makers. It is important that strategic decision-makers are aware of the differences between system design stages and of the implications of the recommended validity levels for each design stage. Only if the scope is properly identified, the gaming simulation can be accordingly designed and required knowledge from operators can be accordingly addressed. The case studies showed that both the PRL game Utrecht 2015 and OV SAAL game did not have a well-defined purpose and balanced scope in terms of game design and resources.

Additionally, the framework also can support in prioritization of resources. For instance, managers may select operators that are open and out-of-the-box thinkers in a design stage of the process, while operators that can easily adapt to new changes are highly preferred in a hypothesis testing stage.

10.5.5 Application of the framework by game designers

Game designers are dependent on the scope and resources that strategic decision-makers provide. However, game designers should also advocate clear requirements of the gaming simulation design. The framework provides a guideline to game designers on the conditions that need to be created in the gaming simulation environment. Different degrees of situation awareness can be achieved through the validity of the gaming simulation design, i.e. in close-to-real gaming simulations the degree of obtained situation awareness is more comparable to that in practice. As a rule of thumb one can state that every deviation from the actual work environment may lead to a lower situation awareness in the designed simulation environment. It should be noted that the actual work environment may have design flaws and does not optimally facilitate for situation awareness, though in terms of generalizability the created condition in the simulation environment reflects the actual work environment. Another point is that the conditions to create the desired cognitive state may not be reached due to the gaming simulation design, as was illustrated in the OV-SAAL game through the design of step-wise rounds instead of real-time operations.

10.5.6 Conclusion

In conclusion, the current chapter described a framework that aimed to provide structure in the use of gaming simulation in support of a certain cognitive state of operators. Gaming simulation could have a purpose to design, test or train operators. However, the case studies also showed that a gaming simulation purpose can change and/or can have different or multiple purposes. For instance, the OV-SAAL game was used for testing, but designing elements could also be identified. Chapter 11 discusses these hybrid gaming simulations and their implications on validity.

References

Cannon-Bowers, J.A., Salas, E., & Converse, S. (1993). Shared Mental Models in Expert Team Decision Making. In: Castellan, N.J. (eds) *Individual and Group Decision Making: Current Issues*. Lawrence Erlbaum Associates, Hillsdale, NJ, 221-246.

Czaja, S. J., Nair, S. N. (2012). Human Factors Engineering and Systems Design. In: G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics*, pp. 38-56. New York: John Wiley & Sons, Inc.

Cooke, N. J., Kiekel, P. A., & Helm, E. E. (2001). Measuring team knowledge during skill acquisition of a complex task. *International Journal of Cognitive Ergonomics, 5*(3), 297-315.

De Bruijn, H., & Herder, P. M. (2009). System and actor perspectives on sociotechnical systems. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 39*(5), 981-992.

Dexter, F., Ledolter, J., & Wachtel, R. E. (2005). Tactical decision making for selective expansion of operating room resources incorporating financial criteria and uncertainty in subspecialties' future workloads. *Anesthesia & Analgesia, 100*(5), 1425-1432.

Duke, R. D., & Geurts, J. L. A. (2004). *Policy Games for Strategic Management*. Dutch University Press: Amsterdam.

Eason, K.D. (1991). Ergonomic perspective on advances in human-computer interaction. *Ergonomics, 34*(6), 721-741.

Ellis, A. P. J. (2006). System breakdown: The role of mental models and transactive memory in the relationship between acute stress and team performance. *Academy of Management Journal, 49*(3), 576-589.

Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement. *Proceedings of the 32nd Human Factors Society Annual Meeting*, 32, 97-101.

Forsythe, D. E. (1999). It's just a matter of common sense: Ethnography as invisible work. *Computer Supported Cooperative Work (CSCW) 8*(1): 127-145.

Fountas, S., Wulfsohn, D., Blackmore, B. S., Jacobsen, H. L., & Pedersen, S. M. (2006). A model of decision-making and information flows for information-intensive agriculture. *Agricultural Systems, 87*(2), 192-210.

Geels, F. W. (2004). From sectoral systems of innovation to socio-technical systems: Insights about dynamics and change from sociology and institutional theory. *Research Policy, 33*(6), 897-920.

Goverde, R. M. (2005). *Punctuality of Railway Operations and Timetable Stability Analysis.* PhD Thesis. Delft: Netherlands TRAIL Research School.

Grogan, P. T., & Meijer, S. A. (2017). Gaming Methods in Engineering Systems Research. *Systems Engineering, 20*(6), 542-552.

Herder, P. M., Bouwmans, I., Dijkema, G. P., Stikkelman, R. M., & Weijnen, M. P. (2008). Designing infrastructures using a complex systems perspective. *Journal of Design Research, 7*(1), 17-34.

Joe, J.C. (2017). A human factors meta model for U.S. nuclear power plant control room modernization. *Proceedings of the 10th International Topical Meeting on Nuclear Plant Instrumentation, Control, and Human-Machine Interface Technologies (NPIC & HMIT 2017)*, 1833-1842, San Francisco, CA.

Johnson, R. R. (1998). *User-Centered Technology: A Rhetorical Theory for Computers and other Mundane Artifacts*. New York, NY: SUNY Press.

Kensing, F., & Blomberg, J. (1998). Participatory Design: Issues and Concerns. *Computer Supported Cooperative Work (CSCW), 7*(3), 167-185.

Klimoski, R., & Mohammed, S. (1994). Team mental model: Construct or metaphor? *Journal of Management, 20*, 403–437.

Krogh, G. V., Nonaka, I., & Ichijo, K. (2000). *Enabling Knowledge Creation: New Tools for Unlocking the Mysteries of Tacit Understanding*. Oxford University Press, Inc.

Lo, J. C., Van den Hoogen, J. & Meijer, S. A. (2013). Using Gaming Simulation Experiments to Test Railways Innovations: Implications for Validity. In R. Pasupathy, S. H. Kim, A. Tolk, R. Hill & M. E. Kuhl (Eds.), *Proceedings of the 2013 Winter Simulation Conference (WSC)* (pp. 1766-1777).

Lo, J. C., Van den Hoogen, J. & Meijer, S. A. (2014). Testing Changes in the Railway System through Gaming Simulation: How Different Types of Innovations Affect Operators' Mental Models. In T. Ahram & T. Marek (Eds.), *Proceedings of the 5th International Conference on Applied Human Factors and Ergonomics (AHFE)* (pp. 8054-8065).

Meijer, S. A. (2008). *The Organisation of Transactions: Studying Supply Networks using Gaming Simulation.* Wageningen: Academic Publishers.

Meijer, S. (2015). The power of sponges: Comparing high-tech and low-tech gaming for innovation. *Simulation & Gaming, 46*(5), 512-535.

Mohammed, S., Ferzandi, L., & Hamilton, K. (2010). Metaphor no more: A 15-year review of the team mental model construct. *Journal of Management, 36*(4), 876-910.

Reich, Y. & Subrahmanian, E. (2015). Designing PSI: an introduction to the PSI framework. *Proceedings of the 20th International Conference on Engineering Design (ICED15),* Milan, Italy.

Rouse, W. B., Cannon-Bowers, J. A., & Salas, E. (1992). The role of mental models in team performance in complex systems. *IEEE Transactions on Systems, Man, and Cybernetics, 22*(6), 1296-1308.

Russ, T. L. (2010). Programmatic and participatory: Two frameworks for classifying experiential change implementation methods. *Simulation & Gaming, 41*(5), 767-786.

Schuler, D., & Namioka, A. (1993). *Participatory design: Principles and practices.* CRC Press.

Spinuzzi, C. (2005). The methodology of participatory design. *Technical Communication 52*(2): 163-174

Schwenk, C. R. (1984). Cognitive simplification processes in strategic decision-making. *Strategic Management Journal, 5*(2), 111-128.

Smith-Jentsch, K. A., Cannon-Bowers, J. A., Tannenbaum, S. I., & Salas, E. (2008). Guided team self-correction: Impacts on team mental models, processes, and effectiveness. *Small Group Research, 39*(3), 303-327.

Van den Top, J. (2010). *Modelling Risk Control Measures in Railways: Analysing how designers and operators organise safe rail traffic*. PhD Thesis. Delft: Netherlands TRAIL Research School.

11 Validity revisited: Hybrid forms of gaming simulations

Abstract
Where Chapters 4 to 8 mainly focus on human factors research, the opportunity for these studies has been largely dependent on the organizational need to use gaming simulations, also described in Chapter 10. Each of these gaming simulations had a research question formulated by practitioners in combination with additional research questions formulated by researchers. As such, these cases exemplify so-called multi-purpose game types or hybrid gaming simulation. This chapter describes these hybrid gaming simulations through an assessment of the studies in Chapters 4 to 8. Following the assessment, the fine design of multi-purpose games and its validity requirements are discussed through a synthesis of previous frameworks.

Keywords: Hybrid gaming simulation; multi-purpose gaming types; research gaming simulation; validity; debriefing

11.1 Introduction

In a gaming simulation with one research question, it is straightforward that its main purpose is to address and answer this question. Its game design can therefore be strictly finetuned. In practice however, the research question and purpose may not be clearly and strictly defined by the organization, i.e. by strategic decision-makers. Therefore the question may change during a gaming simulation session or the gaming simulation purpose may be seen as unsuitable by its participants. For instance, if participants find mistakes in the tested conceptual design in an explanatory research type of game this could lead to the gaming simulation inhibiting characteristics of an exploratory research game, in which operators discuss and provide input on how the conceptual design could be further improved.

One of the characteristics of gaming simulations is that they can quickly span borders, transcending into multi-purposed applications due to their open nature (Grogan & Meijer, 2017). During a gaming simulation session, the borders between different types of gaming simulations may become less defined, i.e. the gaming simulation type may switch back and forth between a design gaming simulation and an exploratory research gaming simulation (e.g. see Chapter 8). This also exemplifies that gaming simulations can be multi-purposed, which can also be seen as a hybrid gaming simulation. In these types of gaming simulation, a single gaming simulation session can focus on for instance design as well as research/testing. In Chapter 10 the purposes of the gaming simulations were described from an organizational perspective; these games had a focus on testing the impact of a changed process or infrastructure with a different subsets of the traffic operating system. However, next to research gaming simulations, the organization also requested the use of training games to prepare operators for future infrastructural or timetable changes. Additionally, a similarity between all studies throughout Chapters 4 to 8 is that in parallel to the request by the organization to conduct a training or research gaming simulation, a second line of studies has been done to investigate operator's cognition through their mental models and situation awareness. This second study could have different requirements for the gaming simulation design and different implications for validity. This chapter builds on the existence and use of hybrid simulations and focuses on the role and ensurance of validity in these types of games.

The current chapter provides a discussion based on empirical observations. In the following section, an assessment of gaming simulation and validity will be provided, followed by an exploration of the characteristics of the conducted hybrid gaming simulations in Chapter 4 to 8. The implications of validity for hybrid gaming simulations will be discussed.

11.2 Hybrid gaming simulations: Five case studies

Following the five gaming simulations in Chapters 4 to 8, the gaming simulations are assessed with characteristics from the framework on validity, research design and execution from Chapter 2. These characteristics are:

1. research configuration
    a. testing: measurement of the variable on different time points in the study, i.e. pre-test and/or post-test
    b. conditions: division of participants in different groups in the study
    c. sampling: selection of the participants in the study by non-probability (convenience) or probability (random) sampling
2. measurement instruments: tools or techniques that are used to collect data, such as questionnaires, observations
3. research environment: the environment in which the study is collecting data, i.e. a laboratory, simulator or field setting

The research questions (both organizational as well as human factors), gaming simulation purpose and the main measured variable are also included as characteristics of the context in which gaming simulation was applied. Table 11.1 provides an overview of these five games. A description is also provided per gaming simulation in relation to the transfer from one gaming simulation type to the other.

Table 11.1: Characteristics of the gaming simulation studies described in Chapters 4 – 8.

Chapter	Organizational question	Organizational purpose gaming simulation	Human factors question	Research purpose gaming simulation	Analyzed cognitive construct	Research configuration			Measurement instrument	Represen-tation
						Testing	Conditions	Sampling		
4	No research question, but request to prepare train traffic operators for the changed infrastructure.	Training	To what extent do levels in train traffic operator's goals and strategic mental models vary?	Explanatory research game	Mental models	Post-test	Two scenarios, three conditions	Entire population of relevant train traffic control area	Questionnaire, simulator logs, video recording, debriefing	High-tech
5	Study 1: no research question, but request to prepare train traffic operators for the changed timetable.	Training	To what extent is the situation awareness of train traffic operators explicit?	Explanatory research game	Situation awareness	Study 1: pre-test and post-test	Study 1: two scenarios	Study 1: convenience	Questionnaire, simulator logs, video recording, observation	High-tech
	Study 2: no research question, but request to prepare train traffic operators for the changed infrastructure.					Study 2: post-test	Study 2: two scenarios, three conditions	Study 2: entire population of relevant train traffic control area	Questionnaire, simulator logs, video recording, debriefing	

Chapter	Organizational question	Organizational purpose gaming simulation	Human factors question	Research purpose gaming simulation	Analyzed cognitive construct	Research configuration			Measurement instrument	Represen-tation
						Testing	Conditions	Sampling		
6	What is the impact of the different workspace design on railway traffic operations?	Explanatory research game	How does situation awareness develop on a team level within railway traffic control?	Explanatory research game	Situation awareness	Pre-test and post-test	Four scenarios, two groups	Convenience	Questionnaire, video recording, debriefing	High-tech
7	What is the impact of the alternative disruption mitigation procedure?	Explanatory research game	How does situation awareness develop on a network level within railway traffic control?	Explanatory research game	Situation awareness	Pre-test and post-test	Two scenarios, one group	Convenience	Questionnaire, video recording, observation, debriefing	Low-tech
8	What is the impact of different infrastructural changes?	Explanatory research game	How do shared mental models develop between railway traffic control operators?	Explanatory research game	Mental models	Post-test	Five scenarios, one group	Convenience	Questionnaire, video recording, observation debriefing	Low-tech

11.2.1 Gaming simulation Chapter 4

11.2.1.1 Research question and gaming simulation purpose

An organizational need was expressed to prepare train traffic controllers (TTCs) for the new infrastructural implementation (i.e. less switches) through training with a human-in-the-loop simulator. From a human factors perspective the predominant goal was to investigate TTC's mental models. Hence, the purpose of the gaming simulation session was twofold from both an organizational and research perspective: to conduct a training with TTCs while performing a research study.

11.2.1.2 Research configuration

All TTCs that were certified to operate the two applicable workstations were trained with the changed infrastructure. Participants conducted two new scenarios, in which no baseline measurement (pre-test) was conducted.

Study 1 involved eleven operators of approximately twenty-five TTCs that were certified for the relevant workstation. Participants conducted two scenarios, in which scenario 1 was the current infrastructure and timetable and scenario 2 was the new one. Hence there was a pre- and post-test. Participants were asked to participate in the training and study on a voluntary basis.

In study 2, all TTCs that were certified to operate the two applicable workstations were trained with the changed infrastructure. Participants completed two new scenarios, in which no baseline measurement (pre-test) was conducted.

11.2.1.3 Measurement instruments

Questionnaires, simulator logs and observations were used to collect data on the work experience, perceived competences, motivation, perceived situation awareness, observed situation awareness, situation awarenss probes, performance, simulator validity, mental workload, mental model development and learning effects. The study presented predominantly statistical findings.

11.2.1.4 Research environment

A human-in-the-loop simulator was used to conduct the study. Similar to the study described in Chapter 4, questionnaires were used to evaluate the validity of the simulator for the task at hand.

11.2.1.5 Transfer between gaming simulation purpose

Similar requirements as described in 2.1.5 hold for the use of a hybrid simulation in which training and research are combined. Instructions and familiarization with the new infrastructure was provided to facilitate mental model development. Similarly, a self-rating item was included in the post-game questionnaire to evaluate learning effects and mental model development.

11.2.2 Gaming simulation Chapter 6

11.2.2.1 Research question and gaming simulation purpose

Both the organizational and human factors research questions required an explanatory research type of gaming simulation. The organizational research question focused on the investigation of the impact of different TTC geographical areas configurations, while the human factors research question focused on measurements of team situation awareness.

11.2.2.2 Research configuration

The scope of the investigation was on a team level. Two teams with participants, often senior TTCs, were selected by the regional control center's management staff to participate in the study on a voluntary basis. The teams included two train traffic controllers, a planner and in a passive, observing role a regional network controller. Four scenarios were presented, in which one represented the current infrastructure conditions. Hence, the research configuration included a pre- and post-test.

11.2.2.3 Measurement instruments

Several simulator technical obstacles were encountered during the gaming simulation session, which impacted the flow during the gaming simulation session. Therefore, data from questionnaires, observations and qualitative input from the debriefing after each scenario was used. Data was collected on simulator validity, geographical workspace design, team situation awareness, trust and cohesion.

11.2.2.4 Research environment

A human-in-the-loop simulator for TTCs (as described in Chapters 4 and 5) was used to conduct the investigation. The validity of the simulator was tested with a similar questionnaire as the studies in Chapters 4 and 5.

11.2.2.5 Transfer between gaming simulation purpose

Chapter 6 also addresses the topic of two research purposes in a single study. Although the purposes are similar, the paradigms of the organizational and human factors research approach need to be aligned. As they both follow the behavioral paradigm, similar research approaches, designs and measurement techniques are used.

Due to the technical obstructions in the gaming simulation session, the focus of the data shifted towards a qualitative oriented approach. Therefore the gaming simulation type shifted towards an exploratory research game type.

11.2.3 Gaming simulation Chapter 7

11.2.3.1 Research question and gaming simulation purpose

The organizational and human factors research questions in this study both required a research type of gaming simulation. The organizational research question focused on the investigation of the impact of a newly designed disruption mitigation procedure, while the human factors research question investigated the communication exchange between operators as an indicator of network situation awareness and network workload.

11.2.3.2 Research configuration

A larger subset of the railway system was simulated, involving traffic operations from a train operating company in addition to railway traffic operations. A total of twelve operators participated in this study, who were recruited by their management team. The gaming simulation was held once with this group of participants.

11.2.3.3 Measurement instruments

Data was collected through video recordings, observations, questionnaires and the debriefing. For the analysis of network situation awareness, communication between operators in the video recording was transcribed.

11.2.3.4 Research environment

A tabletop/analog board gaming simulation was used to simulate a larger part of the railway subsystem. The gaming simulation was designed like this to facilitate a high validity in terms of the structural validity (all necessary information for operators to make decisions was provided), process validity (operators were separated by co-location and facilitators role-played supported the simulation of the train traffic control and train traffic management system. The validity of the gaming simulation was checked through self-rating items in a questionnaire and discussed during the debriefing.

11.2.3.5 Transfer between gaming simulation purpose

Both the organizational and human factors research questions follow a similar paradign in terms of research approach, design and measurement techniques. However, in answering the organizational research question based on the collected data from the observations and the debriefing, the project team summarized their findings as inconclusive: more questions were raised and the actual verification or falsification of the hypothesis could not be conducted. As such, the game would have a better fit as an exploratory game.

11.2.4 Gaming simulation Chapter 8

11.2.4.1 Research question and gaming simulation purpose

Both the organizational and human factors research questions initially required a research type of gaming simulation. The organizational research question initially focused on the impact of different infrastructural changes. The human factors research question focused on the investigation of shared mental models between railway traffic operators.

11.2.4.2 Research configuration

Railway and passenger (train operating company) traffic operators participated in this gaming simulation. A total of eight operators participated in the study. They were all recruited by their management. The gaming simulation was held one time with one group of participants.

11.2.4.3 Measurement instruments

Video recordings, observations, questionnaires and the debriefing after each round were used to collect data to answer the organizational and human factors research questions.

11.2.4.4 Research environment

A tabletop/analog board gaming simulation was used. Operators shared one foam-board based interface where the time-distance graph of trains was depicted in combination with more detailed information, such as the timetable details of platform numbers and infrastructural paths. Taking upon this representation level ensured that operator knowledge about the train traffic flow was triggered. Operator decisions were taken in rounds instead of a continuous time flow.

11.2.4.5 Transfer between gaming simulation purpose

During the gaming simulation session, the gameplay inhibited characteristics of a design gaming simulation. Railway traffic operators finetuned the infrastructural design concept, while also testing this finetuned concept. At the end of the gaming simulation session, the game had the characteristics of both a design and exploratory research type of gaming simulation. The shift of gameplay during the session also impacted the study on shared mental models. Due to the openness of the game, the study on shared mental models shifted to a more qualitative investigation.

11.3 Designing hybrid gaming simulations

The described gaming simulations in the previous section had a double research question in each study, namely an organizational and a human factors research question. In assessing these gaming simulations, it can be observed that gaming simulations with different purposes are integrated in a single gaming simulation session. Challenges in the design and use of hybrid gaming simulations can be observed when it comes to requirements for validity, limitiations in research configurations, selection of suitable gaming simulation purpose, openness of the gaming simulation and the balance between multiple research questions.

11.3.1 Requirements for validity

The discussed hybrid gaming simulations combined training and research, and gaming simulations that combine design and research. It is noticeable that it was only possible to conduct these hybrid gaming simulations due to a certain degree of maturity of training and designing. In other words, the speed of learning in the training gaming simulation was relatively fast and the amount of fundamental design considerations was limited.

As discussed in Chapters 3 and 10, the three different types of games (design, intervention, policy) have less demands on validity, but research games need to comply to specific level of experimental rigidity. Therefore, hybrid gaming simulations need to facilitate the highest validity standards when one of the purposes includes an explanatory research gaming simulation. For instance, in the case of the railway games used for training and research, validity issues for research games were minimized by focusing on the developed mental model of the operators with respect to the game environment as well as to the changed infrastructure, timetable or procedures. Succesful learning in a training gaming simulation relates to achieving a well-developed mental model, which is necessary for a research gaming simulation. However, if a significant amount of learning is needed, mental models of operators are further developed during the gaming simulation, therefore limiting the valid measurement of actual levels of operator's mental models and situation awareness.

It should be noted that in the research gaming simulations where future concepts (infrastructure, timetables or organization processes) were tested, training also occurred beforehand: operators were prepared on the changed railway system state by a detailed briefing on the changes. Similar to the hybrid training and research gaming simulation, it is expected that the introduced system changes can be rapidly learned.

11.3.2 Limitations in research configurations

In ensuring validity requirements in explanatory research gaming simulation careful design considerations need to be made regarding the research configuration, which are defined by the highest validity standards. These components focus on for instance the selection of participants (sampling), number of conditions, testing at different time points and number of confounding variables as discussed in Chapter 2.

In the case of the railway gaming simulations, most studies did not have a randomized sample, limited number of conditions and had a relatively small sample size for studies that focused on the individual level. In studies that investigated on a team and network level the sample size of the number of groups was one to two. Additionally, the selection of participants was not randomized in the sense that often senior operators were recruited by management. That is, operators who were more often involved in, and open to, organizational matters were predominantly participants in the gaming simulation sessions. The specific sample of participants may have a certain influence on the outcome of a gaming simulation, therefore limiting the generalization of the outcomes (i.e. external validity).

11.3.3 Selection of suitable gaming simulation purpose

Another observation from the assessment is that the maturity of the tested design concept is often underestimated by the organization. For instance, the gaming simulation in Chapter 8 initially inhibited the characteristics of an explanatory research gaming simulation. At the end of the session, characteristics of a design and exploratory research gaming simulation could be observed, in which the human factors research question was also reformulated to an explorative approach. Due to the change of the gaming simulation purpose and its characteristics during the gaming simulation session, the investigation of quantitative, explanatory perspective was adapted to a qualitative, exploratory perspective.

11.3.4 Openness of the gaming simulation

The current gaming simulations had multiple gaming simulation purposes in a single research design. Gaming simulation tends to be open to changes in the research design, which may be especially challenging for explanatory research gaming simulation where research design protocols are important. The assessment shows that gaming simulations on a team and network level are especially useful for design and exploratory research. For hybrid gaming simulations it is desirable to lead the transfer between different gaming simulation types in a controlled manner within the gaming simulation session. It should be remarked that not all combinations of transfers have occured in a predicted and controller manner, as can be observed in the gaming simulation in Chapter 8. A reason why the current gaming simulations tend to be of the exploratory research type of games is that in order to answer the organizational research questions, there is a need to explore the impact in a broader sense than was possible with the tested scenarios.

11.3.5 Balancing between organization and research questions

In working with an organization and its professionals, it is more feasible to conduct an academic based study in conjuction with a research question that has a practical relevance for the organization. Chapter 6 discussed the challenges and opportunities that come along with conducting two research goals in one study, in which the existence of hybrid gaming simulations is a necessity.

11.4 Synthesis of validity

In consolidating the previous frameworks on validity types (in Chapter 2), research design (in Chapter 2), the four gaming simulation types (in Chapter 3) to the meta-framework in accordance to Meijer (2008) and Grogan and Meijer (2017), the following meta-framework is introduced (see Figure 11.1).

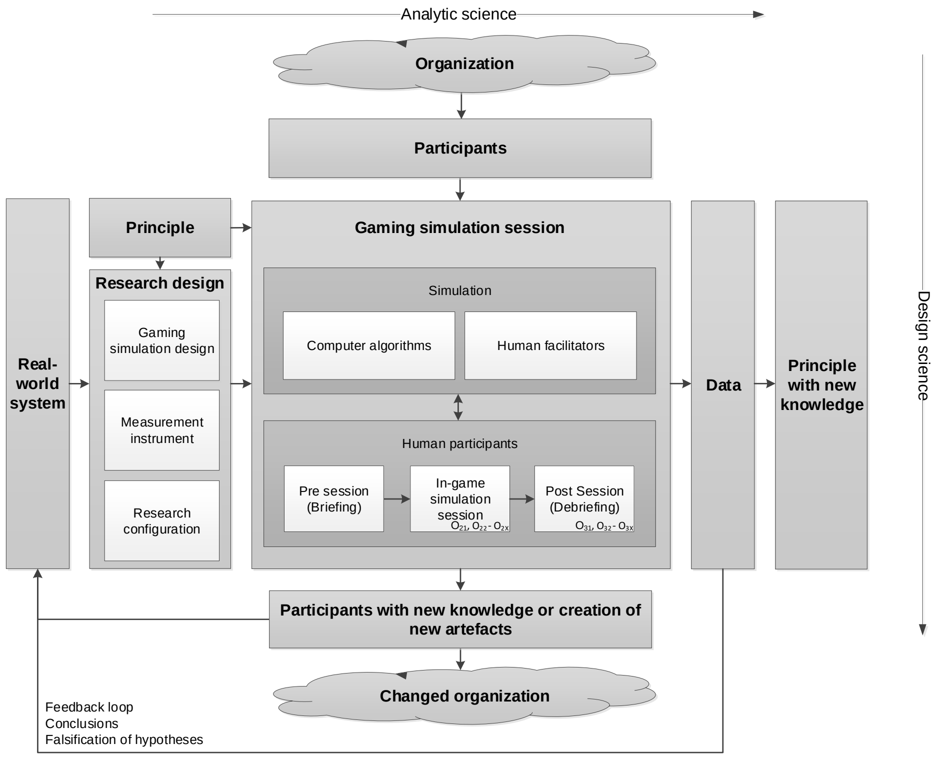


Figure 11.1: Meta-framework for design and analytical science complemented with research design and protocols based on Meijer (2008) and Grogan and Meijer (2017).

11.4.1 Framework components

Next to existing components, such as gaming simulation design, gaming simulation session, participants and data, the framework is complemented with the following new components:

- Research design
- The inclusion and distinction of computer simulation and human participants within a gaming simulation session
- Role of the principle in the framework

11.4.1.1 Principle and participant

In addition to the participant who takes upon a key role in the gaming simulation session, the role of the principle is made explicit in the framework. The principle can be a researcher or a project team (member) who sets the requirements for the gaming simulation and research design. In research games the principle treats the gaming simulation as an experimental setting and is focused on the output of the data from the gaming simulation session. In design games, the principle uses the gaming simulation session to develop and collect ideas and input for a system design change.

11.4.1.2 Gaming simulation design

The position of the gaming simulation design remains unchanged. Depending on the research question and the purpose of the gaming simulation, different design requirements are applicable. In the design, scenario related factors, such as load and situation are also included.

11.4.1.3 Research design

Especially for explanatory research gaming simulations it is important that the research design in terms of its research configuration/experimental design, measurement instrument and research environment (i.e. gaming simulation in the current investigations) are optimally designed to counter validity threats. The implications of these validity threats are described in Chapter 2. Exploratory research gaming simulations are more lenient towards validity threats, in which the outcomes tend to a qualitative orientation.

Similar rules can be held for design, intervention and training gaming simulations, in which these games can be more open for deviation of the intended research design if it contributes to the overall goal of the gaming simulation. In these game types the outcome is directly derived from the gaming simulation session, on the contrary to an analysis after the gaming simulation session with research games.

11.4.1.4 Gaming simulation session

Although the research design is defined in a prior stage, it also extends into the gaming simulation session. For instance, in terms of the gaming simulation design, the representation may vary from an analog, tabletop environment to a digital, human-in-the-loop simulation environment. The control system, such as the traffic management system, may be facilitated or controlled in different ways. For instance, in a tabletop environment, human facilitators may act as the traffic management system, in which they are responsible for the visualization of the train traffic flow by manually moving artefacts that represent trains. In a human-in-the-loop simulator, the simulation of the traffic management system is facilitated by a computer algorithm. While human facilitators will most likely be well-informed and/or the software containing the algorithm well-tested prior to the gaming simulation session, issues can develop during the gaming simulation session that raise validity threats. Similar challenges hold for other research design components, such as the measurement instruments and research configuration.

Further on different data extraction points can be identified throughout the gaming simulation session. The most common occasions to deploy different measurement instruments would be in the briefing, in-game and debriefing. It is also possible to collect data before and after the in-game simulation.

11.4.1.5 Gaming simulation types

Research gaming simulations are located on the horizontal analytical axis where the principle gains knowledge by analyzing the data that comes out of the gaming simulation session. The difference between exploratory and explanatory gaming simulation resides in the amount of validity threats for ecological, external, test and internal validity, in which explanatory gaming simulations require minimalization of all validity types.

Training and invention gaming simulations can be pinpointed to the vertical design science axis, where participants gain or create knowledge. A design gaming simulation follows the vertical axis, in which the principle has the purpose to create an artefact or policy together with, or by, participants.

11.4.2 Implications for hybrid gaming simulations

In hybrid gaming simulations, gaming simulation purposes of the horizontal and vertical axis are most likely to converge. The main challenge is to balance between the gaming simulation purpose and validity requirements. The validity of research gaming simulations in hybrid types of games need be controlled in two stages: the research design and the gaming simulation session. The research design determines the gaming simulation as simulation environment together with its research configuration, measurement tools and protocols. However, during the gaming simulation session the prepared research configuration, tooling and protocols may be implemented differently due to the inherent openness of gaming simulation, especially when used in an organizational setting. The execution of the designed product and protocols are therefore of importance to maintain the expected validity in hybrid gaming simulation with a research focus.

It should also be noted that hybrid gaming simulations are not only marked by their multi-purpose characteristics during the in-game phase in a gaming simulation session. For instance, in the research games that tested future infrastructural designs, training occurred in the briefing phase through a detailed explanation of the changes.

11.5 Discussion and conclusion

This chapter investigated the fine line of the multiple purposes within one gaming simulation based on a number of case studies. The existence of hybrid gaming simulations in organizations can be explained by (1) multiple research questions in a single gaming simulation by multiple stakeholders, (2) testing of future system state without a separate training gaming simulation and (3) the openness of gaming simulation.

In designing for hybrid gaming simulations, the most challenging issue is to ensure its validity requirements. In bringing together the purpose and validity requirements, the following three criteria have been identified. Firstly, the gaming simulation purpose with the strictest validity requirements determines the overall validity requirement of the hybrid gaming simulation. Explanatory research gaming simulations hold highest demands in validity requirements as they follow rigid experimental rules.

Different gaming simulation types should not be perceived as purely categorical, e.g. solely a training game or solely a research game. They can also be differentiated within this category. For instance, a training gaming simulation in which an operator is prepared for the operational impact of the removal of one switch has a different impact than when ninety switches are removed. The successful use of a hybrid gaming simulation can only be guaranteed when a balance can be found in the characteristics of the gaming simulation purposes. For instance, a combination of a training and research game will most likely succeed with the training game where one switch is removed; as such the validity requirement of little to no mental model development is fulfilled.

Finally, the openness of gaming simulations can influence the emergence of a hybrid game. Hybrid gaming simulations can be designed prior to the gaming simulation session. However, hybrid gaming simulations may also occur within the gaming simulation when facilitators do not exert enough control in terms of the research protocol. The uncontrolled emergence of hybrid gaming simulations should be avoided when research gaming simulation are involved, however, they may be wished for in the case of design gaming simulations.

References

Grogan, P. T., & Meijer, S. A. (2017). Gaming Methods in Engineering Systems Research. *Systems Engineering, 20*(6), 542-552.

Meijer, S. A. (2008). *The Organisation of Transactions: Studying Supply Networks using Gaming Simulation*. Wageningen: Academic Publishers.

12 Discussion & Conclusion

Process or technological changes that take place in a complex socio-technical system often have the characteristic of a downward whirl that affect the operational environment in the end. This dissertation focuses on the role of the human operator in gaming simulation and as part of a system (re)design process. This research applies a wide spectrum of disciplines, fields and topics, putting it into perspective of gaming simulation and human factors. The chapters touched upon theoretical implications of validity in gaming simulation and psychological research (Chapter 2), theoretical implications of different gaming simulation types and human factors concepts such as situation awareness and mental models (Chapter 3), human factors research (Chapters 4-8) and applications of gaming simulation and human factors, varying from developing a cognitive model based on situation awareness (Chapter 9), applications for participatory system design and strategic decision-making (Chapter 10) and validity implications of designing hybrid gaming simulations (Chapter 11). Prior to addressing the research questions, a brief summary of the human factors research findings and synthesis of the previous chapters is provided by focusing on the research studies and the introduced frameworks.

12.1 Human factors findings in Dutch railway traffic operations

The human factors investigations in Chapters 4 to 8 provided insights in the characterstics and dynamics of Dutch railway traffic operations for the individual train traffic controller (Chapters 4 and 5), on a team level within a regional control center (Chapter 6) and on a network level between regional control centers, the national control center and passenger traffic control centers (Chapter 7 and 8). A summary of the findings from these studies is provided in Table 12.1.

Table 12.1: Human factors findings in Chapters 4 to 8.

Chapter		**Findings**
4	Individual markers of resilience in train traffic control: the role of operator's goals and strategic mental models and implications for variation, expertise, and performance.	14% of the train traffic controllers perceived the primary organizational goal (i.e. arrival punctuality) as their primary individual goal. Departure punctuality (36%) and platform consistency (18%) are on average perceived as more important goals, indicating an incongruence between organizational and individual goals, and a gap between the work that is expected and the work that is done.
		Train traffic controllers applied up to five different completion strategies in a more disrupted train traffic condition, indicating a relative strong diversity in strategic mental models.
		The relative strong variations in goals and strategic mental models indicate weak resilience at the individual level as the behavior of operators becomes more unpredictable.
		In a more complex state of the traffic system, there is an incongruence between train traffic controllers' self-reported performance indicators and objective performance, possibly indicating goal competition.
5	Explicit or implicit situation awareness? Measuring the situation awareness of train traffic controllers.	As opposed to the widely known three-level model of situation awareness by Endsley (1988), indications of intuitive, unconscious processes, i.e. implicit situation awareness have been found for train traffic controllers as opposed to explicit, reasoned and conscious processes.
		A novel set of analyses is used to identify implicit situation awareness. Each analysis looked into the Situation Awareness Global Assessment Technique (SAGAT) score, which reflects explicit knowledge. Indications of implicit situation awareness were found based on: • fairly low absolute values (e.g. 45%) of the SAGAT scores with relatively high performance (i.e. average punctuality of 88%) and correlations between SAGAT scores and multiple performance indicators • a negative relation between work experience and SAGAT scores: more work experience is related to a lower SAGAT score • deviations between the three SA levels (level-1 SA scores were lower than level-2 SA scores)

6	Balancing organizational and academic research: investigating train traffic controller's geographical workspace design and team situation awareness using gaming simulations.	Both the shared displays (due to train traffic controller's responsibility for a similar geographical area) and communication between train traffic controllers are important components in the development of team situation awareness within a control center.
		Trust in colleagues' skills and to be able to count on them are identified by operators as especially important indicators of successful team collaboration.
7	Assessing network cognition in the Dutch railway system: insights into network situation awareness and workload using social network analysis.	In a disruption, the train traffic controllers in the disrupted area serve as an important node and 'gatekeeper' of information between different subgroups in the network.
		In disrupted train traffic conditions, the regional network controller is a 'spider' in the traffic control network, due to his/her high SNA centrality values, in terms of (1) degree centrality (i.e. most frequent contact with different operators in the network), (2) closeness centrality (i.e. efficiently obtaining information by the 'closeness' to other operators in the network, and (3) betweenness centrality (i.e. passing on information in the network).
		In terms of communication content, the entire coordination revolves around capacity allocation, in which similar information is largely shared across the entire network.
		In informing operators about a disruption, it takes six calls to inform the entire network about a disruption.
8	Participatory design in large-scale railway infrastructure using gaming simulations: a case study of shared mental models	Shared mental model development drives the participatory design of the game, in which three stages can be identified: (1) Shared mental model development of the gaming simulation environment occurs in the first stage: operators developed a shared understanding regarding the game environment. (2) Shared mental model development of railway and passenger traffic operations occurs in the second stage: operators developed a shared understanding of each other's goals. (3) Shared mental model development of the solution space occurs in the third stage: operators developed a shared understanding on the characteristics (strengths and weaknesses) of (part of) the railway system and related implications for different types of disruptions.
		A common graphical interface is essential in establishing mutual understanding.

12.2 Synthesis of research studies

The studies throughout Chapters 4 to 8 focused on (1) different *unit of analysis* within the railway traffic control system (i.e. individual, team and network), (2) different *cognitive constructs* (i.e. mental models and situation awareness) and (3) different *types of gaming simulation* environments (i.e. low-tech tabletop environments vs. simulator environments). Table 12.2 summarizes the characteristics of the studies in the five chapters, in which a differentiation between the organizational and research purposes is made. Additionally, the overview lists the implemented gaming simulation type on the contrary to the initial formulated or designed gaming simulation type.

Table 12.2: Characteristics of the studies in Chapters 4 to 8.

Chapter	Represen-tation	Unit of analysis	Analyzed cognitive construct	Implemented organizational purpose gaming simulation	Implemented research purpose gaming simulation
4	High-tech	Individual	Mental models	Training + research	Explanatory research
5	High-tech	Individual	Situation awareness	Training + research	Explanatory research
	High-tech	Individual	Situation awareness	Training + research	Explanatory research
6	High-tech	Team	Situation awareness	Exploratory research	Exploratory research
7	Low-tech	Network	Situation awareness	Exploratory research game	Explanatory research
8	Low-tech	Network	Mental models	Design + exploratory research	Exploratory research

In assessing the dimensions of gaming simulation representations, cognitive construct and unit of analysis (see Table 12.2 and Figure 12.1), it can be observed that gaming simulations in the area of high-tech representations on a network level and low-tech gaming simulations on an individual level are absent. The absence of a high-tech gaming simulation on a network level can be explained due to the ongoing development of this simulation environment.

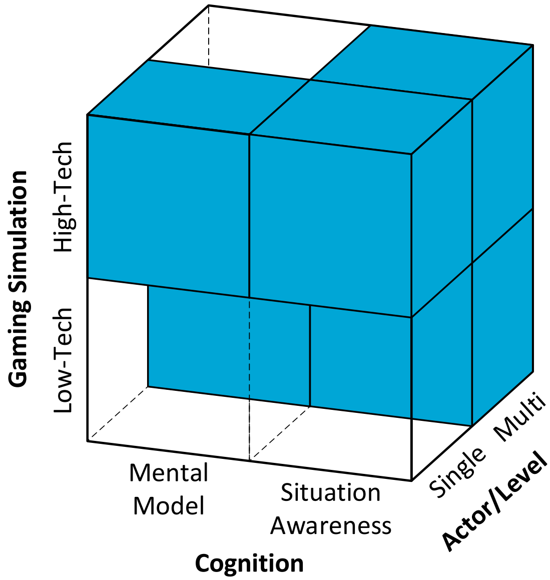


Figure 12.1: Three dimensions for identifying gaming simulation configurations and the studied dimensions in this work (in blue).

However, it can be noted that there was no specific desire to conduct single-actor gaming simulations in a low-tech environment, as the studies were driven by an organizational need. The application of low-tech representations on an individual level is widely applied the field of interface design e.g. Warfel, 2009). Given the current focus of the organization, there was no need such gaming simulation environments.

In terms of organizational purposes, it can be observed that the organizational need focuses on training gaming simulations on an individual scale and more research-oriented gaming simulations on a multi-actor scale.

12.3 Synthesis of frameworks

The theoretical frameworks in this work focus on connecting different concepts and as a guideline for (the design of) a gaming simulation study. The theoretical frameworks in Chapters 2 and 3 provided the fundamental basis to apply and extend the frameworks to bridge the literature gap between gaming simulation and human factors research. The frameworks in the following chapters were introduced:

- A theoretical framework of different validity types in gaming simulation was discussed in Chapter 2 based on literature from computational and social sciences. These validity types occur throughout different gaming simulation stages (design of the game, pre, in-game and post game).

- A theoretical framework on cognitive structures in relation to different gaming simulation types was introduced in Chapter 3; mental models and situation awareness were linked to research, design, training and policy/intervention gaming simulations.
- Building on the theoretical framework in Chapter 3, the framework in Chapter 10 focused on the integration of gaming simulation types and cognition (mental models and situation awareness) throughout the participatory design process. The framework aimed to address implications for validity when extracting operator knowledge in different gaming simulation types and throughout different stages of a participatory design process. This framework is especially relevant for strategic decision-makers and gaming simulation designers.
- Building on the theoretical framework in Chaper 2, the framework in Chapter 11 focused on validity in multi-purpose, hybrid gaming simulations. The meta-framework for different gaming simulations in Chapter 11 integrates validity requirements for the different gaming simulation types. This framework has a specific focus on hybrid research gaming simulations. It is particularly useful for gaming simulation designers that need to design hybrid gaming simulations.

Additionally, Chapter 9 introduces a cognitive model rather than a framework to model a human operator. The concept of situation awareness is used as a theoretical foundation for the model due to its importance as an indicator for good decision-making. The situation awareness model (SAM) is used to model the dynamic evaluation of offers during a negotiation between operators.

12.4 Research questions

In answering the research questions, the terms of mental models and situation awareness are used in a singular form to represent multiple levels (individual, team or network level) for simplification.

1. How are cognitive concepts such as mental models and situation awareness of train traffic and network operators relevant for gaming simulations and vice versa?

Mental models and situation awareness are fundamental concepts in the cognition of human operators. These cognitive concepts can also particularly be of value in the context of participatory system (re)design, in which gaming simulations are used to design and test with operators or to train operators. The relevance of these cognitive concepts for gaming simulations can be approached from three different perspectives:

1. *Measurement studies* can be conducted on the development and/or characteristics of operator's mental models and situation awareness. The research gaming simulations in Chapters 4 to 8 provided insights in (a) individual mental models and situation awareness of train traffic

controllers, (b) team situation awareness development at a regional control center and (c) shared mental model development and situation awareness of railway traffic operators on a network level.

(a) In the study on individual mental models, Chapter 4 investigated the diversity in train traffic controller's goals and strategic mental models. The findings provided indications of a relative strong diversity in primary operator goals and strategic mental models. In terms of resilience, this level of variation between operators could be seen as an indication of weak resilience.

    Chapter 5 measured situation awareness of train traffic controllers and found indications of implicit situation awareness. This was explored through a set of three analyses, which indicated the presence of intuitive, unconscious processes, i.e. implicit situation awareness as opposed to conscious, reasoned process with train traffic controllers.

(b) The study on team situation awareness in Chapter 6 explored team collaboration and situation awareness development at a regional control center. An unexpected finding was that team identity was not held on the regional control center level, but was rather shared by operators who worked on overlapping geographical areas of infrastructure. This overlapping information was also presented on shared displays between train traffic controllers. The qualitative findings indicated that the development of team situation awareness was facilitated by both the shared displays, communication as well as coordination between train traffic controllers.

(c) On a network level, the cognition of multiple operators was also explored. Chapter 7 focused on workload and situation awareness on a network level, in which findings on operator's communication flow and content using social network analysis pointed out so-called 'spiders' and 'gatekeepers' in the network. The identification of these 'spiders' and 'gatekeepers' can be particularly useful in assessing the efficiency and workload of the network, particularly in relation to the number of failed communication attempts.

    Chapter 8 provided insights in the process of shared mental model development on a network level. Findings from this study led to three propositions on the development of shared mental models between operators in a gaming simulation: firstly, operators developed a shared understanding regarding the game environment, then they developed a shared understanding of each other's goals. Finally, operators developed a shared understanding of the characteristics (strengths and weaknesses) of (part of) the railway system and their implications for different types of disruptions.

All in all, these studies contain valueable knowledge on (the development of) operator's mental model and situation awareness. These findings are crucial to optimize or (re)design the railway traffic control system, for instance in terms of operator's interface, procedures or training program.

2. When *extracting operator's knowledge* in a system design process, mental models and situation awareness can be used as a reference for the validity and level of detail of this knowledge. Knowledge of operators is valuable as it holds experience on the system's characteristics and dynamics. In a participatory (re)design process of the system, this knowledge can be used to contribute to finetune the system's design and, indirectly, its performance. The validity and level of detail of this knowledge depends on the stage within the system's design and can be steered by the gaming simulation type and its design.

   The developed framework in Chapter 10 uses the different system (re)design project stages as a foundation. For each stage in the system design, different gaming simulation types are used. Depending on which gaming simulation type is used, either mental model or situation awareness should be facilitated in order to obtain valid, detailed and desired operator knowledge. For instance, situation awareness is only required when context-rich, detailed and representative knowledge from operators is needed.
3. Finally, the theoretical notion of situation awareness is conceptually relevant in relation to good operator decision-making and in relation to the *development of cognitive models*. These cognitive models can be used in the artificial intelligence field to develop intelligent software agents that can support or replace human operators. As the success of gaming simulations is dependent on the input from and involvement of operators, the unavailability of operators in a gaming simulation is a limiting factor in its execution and success.

   Chapter 9 presents the use of intelligent agents to replace human participants as a solution when resources are pressured. Before decision-making and action can be taken by an operator, situation awareness is required. A cognitive model based on situation awareness theories is developed for the role of the national network controller with a specific focus on the operator's situation awareness during a negotiation.

In sum, these perspectives each illustrate the relevance of mental models and situation awareness for gaming simulation and using gaming simulation. Gaming simulation vice versa. Mental models and situation awareness are relevant for gaming simulation in that they are key requirements in the design and use of gaming simulation to extract operator knowledge in a participatory system design. Additionally, situation awareness based cognitive models can aid in the use of gaming simulation when human resources are limited.

Gaming simulations are relevant for mental models and situation awareness, awareness, being the research tool inwith which mental models and situation awareness can be established and measured. The empirical findings in the measurements of mental models and situation awareness in Chapter 4 to 8 are examples of the actual work as done by operators. These insights can directly be used to improve the railway traffic control system.

Vice versa, mental models and situation awareness are relevant for gaming simulation in that they are key requirements in the design and use of gaming simulation to extract operator knowledge in a participatory system design. Additionally, situation awareness based cognitive models can aid in the use of gaming simulation when human resources are limited.

It should be remarked that the first two perspectives can co-exist in parallel: when mental models or situation awareness are established, they can be measured as well as used to extract knowledge from.

2. Which situation awareness theories can be used in the Dutch railway traffic control domain?

Three studies in this dissertation investigated the situation awareness (SA) of railway traffic operators on an individual level (Chapter 5), team level (Chapter 6) and network level (Chapter 7), using different SA theories:

1. The widely used three-level model by Endsley (1988) was explored in Chapter 5. As this model uses the *classical information-processing paradigm*, conscious, active processing of information in the working memory is recognized in establishing situation awareness. Using the Situation Awareness Global Assessment Technique (SAGAT) method as its matching measurement technique, the individual situation awareness of train traffic controllers was investigated. Although the three-level model and SAGAT method are most frequently applied, the findings showed limited support for this theory and method. A set of analyses showed indications for the presence of implicit situation awareness, implying that operators are less likely to actively process their perception, interpretation and prediction of changes (i.e. situation awareness) and therefore are less likely to make this process explicit. The finding that operators process information in an unconscious but fast manner is not uncommon and has been identified in the naturalistic decision-making field, especially with experienced operators. The findings indicate limitations of the classical information-processing paradigm and therefore advocate the use of other cognitive

engineering/human factors paradigms to investigate situation awareness on a team and network level.

2. Opposed to the classical information-processing paradigm, the paradigm of *macrocognition* posits that team and network cognition is beyond the sum of individuals and should be measured on a team or network level. According to this paradigm, communication and coordination are important indicators of team or network situation awareness. Due to limitations in the research study in Chapter 6, team situation awareness at a regional control center could not be studied in accordance to the group cognition theory by Gorman, Cooke and Winner (2006). However, indications on the role of communication and coordination were revealed through observations and data from the debriefing. The qualitative findings showed the development of team situation awareness through the information exchange on operator’s actions and the use of shared displays. The relevance of shared displays is also in line with classical information-processing theories, such as the three-level model on a team level. All in all, these findings indicated that multiple situation awareness theories can exist in parallel.
3. Group cognition theory as a stream within *macrocognition* was used for the study on network situation awareness (Chapter 7). Communication and coordination on a network level were assessed through social network analysis. This method revealed valuable insights in situation awareness and workload in a network of railway traffic operators. The findings indicated that similar information was largely shared across a network of railway traffic controllers. Specifically, one operator was occasionally overloaded with tasks, because of which phone calls could not be answered. This restricted further communications and therefore contributed to a delay in coordination between operators.

In sum, the applied situation awareness theories and related measurement techniques each provided unique findings on different levels of analysis (individual, team, network). The findings on individual situation awareness showed limitations using the classical information-theory paradigm, while the macrocognition paradigm on a network level provided useful findings. Relevant findings in support of both paradigms were found on a team level. As such, the notion of pluralism, i.e. the parallel existence of different paradigms and theoretical streams, is highlighted in this work. This work has pointed out that there is variety of theories, each with their own unique contributions and limitations.

3. What are the requirements on research gaming simulations in order to measure the mental models and situation awareness of operators?

Validity is a key requirement when designing for research gaming simulations that can measure mental models and situation awareness. To generalize findings of operator's mental models and situation awareness to the actual work environment, research gaming simulations require high validity levels. In ensuring validity, three aspects on validity need particular attention:

1. Validity can be found in the *design of the gaming simulation environment* (ecological validity) and validity is also relevant within the *gaming simulation session* (external, test and internal validity) (Chapter 2). In order to obtain a high validity, all these validity types need to be managed.
2. The relevance of *ecological validity* in terms of the gaming simulation design also relates to the representation of the gaming simulation environment. The most intuitive but most resource intensive option is to develop a research gaming simulation that is highly comparable to the actual operational environment. The assumption is that valid and accurate results can be obtained that can be generalized. However, this may not be necessary when the validity of the gaming simulation is high on the dimension of structural validity, process validity and psychological reality, and in relation to the task at hand. For instance, a gaming simulation focusing on mitigating disruptions should have all required information and tools for an operator to execute relevant tasks for such an event. The gaming simulation design does not necessarily need to include additional information and tooling for tasks that are not relevant.

   In terms of the representation of the environment it can also be noted that the use of indexical and symbolic simulation principles in the gaming simulation may capture the essence of the actual work environment. As such, analog, tabletop representations can be used as a valid simulated environment (see Chapters 7 and 8). It can also be argued that in a large multi-actor gaming simulation setting, the emphasis is on the communication and coordination in terms of network situation awareness. As such, the gaming simulation can be designed accordingly to these requirements.
3. A pivotal question for gaming simulation designers is to identify whether the gaming simulation serves *one or multiple purposes*. If multiple purposes are served, the gaming simulation is most likely a *hybrid gaming simulation*. In the case of a hybrid gaming simulation, careful considerations need to be made with respect to the gaming simulation design, research design and research procedures during the gaming simulation session. These conditions need to be strictly followed to

comply to the high validity requirements that are set in measuring mental models and situation awareness (Chapter 11).

In sum, research gaming simulations that are used to measure mental models and situation awareness of operators should ideally comply to with the highest validity research standards. These validity requirements can be compared to requirements in psychological research experiments in terms of their research design and research procedures. By striving for the highest validity standards, findings from measurement studies can be better generalized to the actual operational environment. It will be more challenging, however not unlikely, for hybrid gaming simulations to adhere to these validity standards. The current work provided a framework that identified validity requirements of gaming simulations for research studies. Validity threats can be found throughout the design and the execution of the research gaming simulation.

4. To what extent can gaming simulation be used as a formal research environment for complex operational environments?

In reflecting on the use of gaming simulation as a formal research environment in Chapters 4 to 8, the following observations can be drawn:

1. There is a need to define a research gaming simulation more specifically. A distinction can be made between research games for *explorative testing* and research games for *explanatory/hypothesis testing* (Chapters 3 and 10). Explorative research gaming simulations focus on testing the conceptual design and finetuning this design, while explanatory gaming simulations use a final version of the conceptual design and investigate causal relations.

   In analyzing the conducted studies (Chapter 11), it was observed that the intended explanatory research gaming simulation often turned out to be an explorative research gaming simulation in hindsight: the conceptual design often could be further finetuned by the input that operators provided (Chapters 6 to 8). In another study, the explanatory gaming simulation developed during the session into a design gaming simulation (Chapter 8). As such, it can be stated that the studies showed that gaming simulations can be used as a research environment, however there has been an incongruence between the intended purpose of the gaming simulation (i.e. being an explanatory gaming simulation with a final conceptual design) and the outcomes of the gaming simulation (e.g. obtaining more input and subsequent research questions).
2. *Applying* and *executing experimental research designs* in an organizational context is *difficult to manage*. The studies in Chapters 4 and 8 showed limitations in the research design, for instance a limited number of scenarios, non-random selection and number of participants. As a result, quantitative data from the gaming simulation session was

not (often) used. Instead, the research gaming simulations that investigated organizational questions strongly focused on the depth of the obtained qualitative input (Chapter 11).

In sum, although this work has showed the practical value of research gaming simulations for organizational research and human factors research (Chapters 4 to 8), there are practical challenges that are encountered. Both the mismatch between intended (explanatory) and executed (exploratory) research gaming simulation type in hindsight, and the difficulty to apply experimental research designs are limiting research gaming simulation to the design and explorative types of games.

5. How can mental models and situation awareness contribute to system design processes using gaming simulations?

Small- and large-scale changes in socio-technical systems allall have an operational impact and ideally should be co-designed together with their operators. Due to their operational experience, operators hold unique knowledge on the system's characteristics and its dynamics. This valuable knowledge can be used as input for the system's conceptual design (Chapter 10). Gaming simulation can be used as an environment to extract this knowledge as part of a participatory system design process. The knowledge that is extracted depends on two criteria:

1. *When it is extracted*. The system design process consists of different stages. Each stage (design, testing/research, training) requires and attends to operator knowledge in a different manner and therefore also puts requirements on the design of the *gaming simulation type* (design, exploratory research, explanatory research, training) to extract the knowledge. In a design stage, the desired outcome consists of ((creative) ideas from expert operators that can form, shape and contribute to the system's conceptual design. After a design concept is sufficiently developed it can be iteratively explored in an exploratory gaming environment. The desired outcome is a finetuned design concept based on operator knowledge. Subsequently, a mature design concept can be experimentally tested in a gaming simulation with an explanatory/ hypothesis testing focus. In this stage, the desired outcome aims at testing the operational impact of the matured design concept with ideally limited feedback based on highly detaileddetailed operator knowledge. Finally, in a training stage, the desired outcome focuses on developing operator knowledge in relation to the implications of the chosen design concept rather than extracting knowledge. Training takes place before the design concepted is put into operation.

2. *Where it is extracted.* The validity and level of detail of the extracted knowledge depends on whether it is elicited from a mental model or from the situation awareness of an operator When the extracted knowledge is based on mental models, this knowledge may be predominantly based on previous experiences, while knowledge based on situation awareness most likely is context-rich and detailed.

   In a design stage of a system design process, knowledge that is extracted from an expert operator's well-developed mental model may be abstract and open-minded. In an explorative research game, knowledge from well-developed mental models is required while operator's situation awareness does not need to reach a highly representative state. However, explanatory research games require that the operator knowledge is based on well-established situation awareness, which comes with a high representative level of detailed knowledge based on the setting and its tasks. In a training stage, operator mental models are changed as operators need to learn how the change will affect their tasks, roles and/or systems.

In sum, the knowledge of operators is crucial throughout the system (re)design process. Not attending to operator knowledge as input for the system design process may lead to suboptimal solutions and possibly a higher resistance to organizational change. The framework in Chapter 10 focuses on the methodological challenge to optimally make use of the role of operator knowledge throughout a system design process. Mental models and situation awareness are crucial requirements for establishing validity and level of detail of the extracted and attended operator knowledge.

In terms of stakeholders in the system design process other than the operators, strategic decision-makers have an influencial role in defining requirements in terms of validity and level of detail of the extracted operator knowledge. In turn, the gaming simulation designers are responsible for developing a gaming simulation that meets these requirements.

12.5 General conclusion

The balance between the relevance of investigations for research/academic purposes and that for practice/organizational purposes has been a returning challenge in the conducted studies. In reflecting on a general conclusion based on the findings, frameworks and research questions, a distinction will be made between these perspectives.

12.5.1 Researchers

In scientific research a key point is the advancement of theory and methods. The current work introduced frameworks on a multi-disciplinary intersection of gaming simulation - particularly in relation to participatory (system) design - and human factors concepts and research.

For the human factors/cognitive engineering field, situation awareness theories from the traditional information-processing paradigm and macrocognition paradigm have been applied. Herein, the notion of pluralism is embraced: this work has pointed out that there are various theories rooted in different paradigms. The studies each underlined the value and application of different theories, opposed to following one, as an ultimate paradigm, theoretical stream and/or theory. These theories each have their own unique contribution and limitation and also may exist in parallel.

The studies in this dissertation also have illustrated the challenge of conducting human factors research in an applied setting. Herein a fine balance between the research design requirements and the organizational demands needs to be found.

For the gaming simulation field, theoretical advancements have been made in the area of research games, in which a further distinction has been made between exploratory and explanatory research games. This distinction between a bottom-up, open versus a hypothesis, controlled research environment may be seen as trivial, however it can have crucial implications in terms of validity and generalizability of the findings.

The importance of experimental rigidity in the gaming simulation design of explanatory research games and the importance of debriefing in exploratory research games in particular contributed to the body of knowledge on research games. The existence of hybrid gaming simulation and its design requirements is another theoretical development in the gaming simulation field that yet needs to be further explored.

12.5.2 Practitioners

The knowledge that operators possess of the dynamics in an operational setting can contribute to different stages of a participatory system design process, in which gaming simulation can be used as a tool to facilitate this process.

For strategic decision-makers this work is predominantly relevant in terms of guidelines and insights on the level of influence they have in requirement formulation for the gaming simulation and the level of quality of the conducted gaming simulation by providing resources. The use of intelligent agents to reduce load on resources can be a solution to limit restrictions in the organization of gaming simulations, although more research is needed in the cognitive modeling of operators.

For gaming simulation designers this work has provided guidelines on increasing the validity of (hybrid) gaming simulations and how cognitive concepts such as mental models and situation awareness influence the design of research games.

For human factors/cognitive engineering practitioners the findings on operator's mental models and situation awareness can be useful in optimizing and designing operational systems, procedures and in supporting gaming simulation designers to translate validity requirements in the design of explanatory research gaming simulations.

12.5.3 Limitations

The current work exists of multiple investigations, in which a number of deficiencies can be identified. A common thread in the cause of limitations in these studies can be ascribed to the challenges that are faced when conducting research in an organizational setting. Trade-offs in the research design have been made for instance in the case of number of participants, participant selection and number of gaming simulation sessions.

A number of limitations throughout the studies can be ascribed to technical issues in the simulator or ad-hoc changes in the gaming simulation rules, leading to a qualitative oriented approach in some studies (e.g. Chapter 6 and Chapter 8). However, the qualitatively obtained findings still provided valuable insights into the development of shared mental model and team situation awareness.

Further on, the cognitive model focused on in Chapter 9 is based on an extension of Endley's three-level model of situation awareness. Based on the findings from Chapter 5 this model may need revision to include the role of implicit situation awareness in the cognitive model.

12.5.4 Future work

This dissertation has introduced multiple frameworks that connected literature on gaming simulation design and human factors research. These frameworks have been illustrated and supported by the studies in Chapters 4 to 8. Further research could support the substantiation of the frameworks.

The applied theories in this work were of different paradigms, varying from classical information process theories to macrocognitive perspectives. Following the latter perspective, research in the area of naturalistic decision-making and resilience engineering could provide insights into and development of the adaptivity and resilience on a multi-level scale of the railway system.

The human factors knowledge that was gained from these studies could provide valuable input to the system and interface design of operators. With the stronger role of digitalization and intelligent solutions in the rolling stock to increase capacity, e.g. through new technologies such as the European safety system (European Rail Traffic Management System – ERTMS) or Automated Train

Operations (ATO), the role of automation and the role of human operators is challenged. However, opportunities also arise to redesign roles towards more balanced workload conditions.

References

Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement. *Proceedings of the 32nd Human Factors Society Annual Meeting*, 32, 97-101.

Gorman, J. C., Cooke, N. J., & Winner, J. L. (2006). Measuring team situation awareness in decentralized command and control environments. *Ergonomics, 49*(12-13), 1312-1325.

Warfel, T. Z. 2009. *Prototyping: A Practitioner's Guide*. New York: Rosenfeld Media.

Summary

The most dominant reason for the Dutch railways to innovate has been the expected increase in railway passengers and freight demand. To achieve this a large scale (re)design process of the railways as complex socio-technical system is needed. Since infrastructural expansion alone is not a sustainable option, solutions are also sought in innovative process optimizations.

Train traffic and network controllers are responsible operators for the operational management of the railway system. With their experience, these operators hold unique knowledge about the system's actual characteristics and its dynamics. Therefore, it is crucial to test and train any new system design with them, for example by gaming simulations.

This dissertation focuses on the role of the human operator as part of a large-scale socio-technical system (re)design process of the Dutch railway system.

Thesis topics

The following topics in this dissertation are covered: mental models and situation awareness, (hybrid) gaming simulations and validity.

Mental models and *situation awareness* are established cognitive constructs within the human factors/cognitive engineering field. These concepts are linked to operator knowledge, in which mental models are often defined as knowledge structures that hold an operator's representation of a physical system (e.g. what is a switch, how does a switch work etc.). Situation awareness builds on well-developed mental models, in which this knowledge is put into context. In turn, situation awareness is also an indicator of good operational decision-making.

Gaming simulations can be used as a research environment to design, test and train users with these future system designs. Hybrid gaming simulations come into play when multiple gaming simulation purposes (e.g. design, test or train) occur in one gaming simulation session.

Validity is especially relevant for research gaming simulations. Exploratory research gaming simulations are focused on testing a system design, in which insights are gathered to improve the system design. Hypothesis testing or explanatory research gaming simulations are focused on testing a system design, in which insights are gathered on the expected impact of a future system design. While exploratory research gaming simulation are more lenient towards validity requirements, explanatory research gaming simulations must be compliant towards strict experimental rigidity. The highest validity level is required with measurements on for instance mental models and situation awareness.

Using (hybrid) gaming simulations and a human factors perspective this dissertation focuses on:

1. Measurements of (individual, team and/or network) mental models and situation awareness of train traffic and network controllers
2. The extraction of operator knowledge as feedback for system design, in which validity is ensured by mental models and situation awareness
3. The use of situation awareness theory to develop intelligent agents to represent human operators in gaming simulations

Thesis structure

This work is comprised of three sections. Section 1 focuses on the theoretical foundation of this dissertation by reviewing literature on gaming simulation and validity (Chapter 2) and gaming simulation types and mental models and situation awareness (Chapter 3).

- Chapter 2 reviewed different validity types in computer sciences, social sciences and human factors, and provided a framework that captures the different stages and types of validity that are required in gaming simulations.
- Chapter 3 reviewed literature on the cognitive concepts of situation awareness and mental models from the human factors field and literature on the different gaming simulation types. A framework is provided that captures the four gaming simulation types (design, research, training and policy) in relation to mental models and situation awareness.

Section 2 focuses on measurements of mental models and situation awareness. The investigation of mental models and situation awareness of operators is key in the optimization and design of the operator's task space, for instance rules and procedures and system interfaces. This is not only limited to the individual level (Chapter 4 and 5), but also reaches to the team level - e.g. operators within a control center (Chapter 6) and network level - e.g. operators from regional and national control centers (Chapter 7 and 8). The relevance of different situation awareness theories was also investigated throughout Chapters 5 to 7.

- Chapter 4 investigates the diversity in train traffic controller's goals and strategic mental models. The findings revealed an incongruence between the primary goal set by the organization and the primary goal set by train traffic controllers, i.e. only 14% of the operators agreed with the primary organizational goal being arrival punctuality. Another finding was that train traffic controllers applied up to five different completion strategies (i.e. an operationalization of a strategic mental model) in a more disrupted train traffic condition. These findings provided indications of a relatively strong diversity in primary operator goals and strategic mental models. This level of variation between operators could be seen as an indication of weak resilience.

- Chapter 5 measures situation awareness in line with the popular three-level model by Endsley and its related Situation Awareness Global Assessment Technique (SAGAT) as method. A high SAGAT score would indicate a high situation awareness level of an operator through explicit, reasoned and conscious processes. However, indications of predominantly implicit situation awareness were found for train traffic controllers. This was explored through a set of analyses: (1) fairly low absolute values (ca. 45%) of the SAGAT probes with relatively high performance (i.e. average punctuality of 88%) and correlations between SAGAT scores and multiple performance indicators, (2) the negative relation between work experience and SAGAT scores and (3) deviations between the three SA levels (level-1 SA scores were lower than level-2 SA scores). These findings indicated the presence of intuitive, unconscious processes, i.e. implicit situation awareness as opposed to conscious, reasoned process with train traffic controllers.
- Chapter 6 explores team collaboration and situation awareness development at a regional control center. An unexpected finding was that team identity was not associated with the regional control center, but was rather shared by train traffic controllers who worked on overlapping geographical areas of infrastructure and shared the same information displays. The findings indicated that the development of team situation awareness was facilitated by both the shared displays, communication, as well as coordination between train traffic controllers.
- Chapter 7 focuses on workload and situation awareness on a network level, in which social network analysis was used as a technique to respectively investigate communication flow and content. The findings showed so-called 'spiders' and 'gatekeepers' in the network. During a disruption, the train traffic controller in the disrupted area served as an important node and 'gatekeeper' of information between different subgroups in the network. It is notable that overall, largely similar information was shared across the entire railway traffic network. In general, the regional network controller was a 'spider' in the traffic control network through his/her central position. The identification of 'spiders' and 'gatekeepers' can be particularly useful in assessing the efficiency and workload of the network, particularly in relation to the number of failed communication attempts.
- Chapter 8 provides insights in the process of shared mental model development on a network level. Findings from this study led to three propositions on the development of shared mental models between operators in a gaming simulation: firstly, operators developed a shared understanding regarding the game environment. They then developed

a shared understanding of each other's goals. Finally, operators developed a shared understanding on the characteristics (strengths and weaknesses) of (part of) the railway system and related implications for different types of disruptions.

Section 3 focuses on the application of mental models and situation awareness and gaming simulations for cognitive modelling (Chapter 9), extracting operator knowledge in different system design stages (Chapter 10), and implications of validity in hybrid gaming simulations (Chapter 11).

- Chapter 9 focuses on developing cognitive models based on situation awareness for gaming simulation. In the organizational use of gaming simulations, resources and particularly the availability of operators are scarce. One of the solutions to tackle this issue is through the application of intelligent agents that can make decisions on behalf of an unavailable operator. As situation awareness is an indicator for good decision-making, the current work investigated the use of situation awareness theories to develop a cognitive model of an operator in a negotiation setting. Agent-based modeling is used as a technique to create an initial cognitive model of a national network controller.
- Chapter 10 focuses on the role of mental models and situation awareness in extracting operator knowledge*.* Using the participatory design perspective in the (re)design of the railway system, the professional knowledge from operators to shape and finetune the system design is acknowledged as important input for strategic decision-making. Operators can be involved in different phases in the system design process, varying from designing, testing to training. Operator knowledge derived from gaming simulations in each of these system design stages can be used by strategic decision-makers to further develop the system design concept.

  The chapter provide a framework, in which the different system design stages are linked to different requirements in operator knowledge and to operator's cognitive state and validity. A gaming simulation that has a high resemblance in comparison to the actual work environment is most likely to trigger a high situation awareness: when operator knowledge is retrieved from operators in this cognitive state, this knowledge will probably hold rich environmental and situational details with a higher accuracy in the description and thereby a higher validity. However, to obtain this validity level much effort needs to be invested in eliminating validity threats in the study. On the contrary, gaming simulations that deviate from the resemblance with the actual work environment by using stepwise decision-making rounds instead of real-time most likely will not trigger a representative situation awareness. Instead, operators will rely on their mental models when knowledge is extracted.

The framework can be used by strategic decision-makers and gaming simulation designers to define the design requirements of a gaming simulation. Strategic decision-makers play an important role in the requirements of the gaming simulation purpose by the formulation of the research questions. Secondly, the research design can by influenced by strategic decision-makers through the allocation of resources. Gaming simulation designers are dependent on the scope and resources, however they are responsible for the output. The framework also provides insights in the design conditions and requirements that need to be met to achieve a certain level of output. As such, gaming simulation designers can use these requirements to find a balance between the gaming simulation design and desired output expressed by strategic decision-makers.

- Chapter 11 explores the gaming simulations in Chapters 4 to 8, which combined multiple purposes in a single gaming simulation. These multi-purposed or hybrid gaming simulations were caused by (1) multiple research questions in a single gaming simulation, (2) testing of future system design without a separate training gaming simulation and (3) the openness of gaming simulation. The predominant form was the existence of multiple research questions, which was caused by an organizational wish to use operator knowledge to investigate organizational research questions in combination with an academic oriented focus to investigate human factors research questions.

  Most challenging in the design of hybrid gaming simulations is to ensure validity of these gaming simulation to be able to fulfill the gaming simulation purposes. A synthesis of different frameworks is provided in this chapter that touches upon the design of the research design and gaming simulation session and the level of exerted control in these two stages, in order to design an explanatory gaming simulation.

Chapter 12 concludes with a discussion and conclusion. In this chapter, findings from the research studies and frameworks are summarized, research questions are answered, and a general conclusion is provided.

Samenvatting

De verwachte toename in reizigers- en goederenvervoer vereist een capaciteitsvergroting van de Nederlandse spoorweginfrastructuur. Om deze te bereiken vindt een grootschalig (her)ontwerpproces van het spoor als complex sociaal-technisch systeem plaats. Aangezien alleen uitbreiding van de infrastructuur geen werkbare optie is, worden ook oplossingen gezocht in innovatieve procesverbeteringen die zich richten op het operationele management van de spoorweginfrastructuur.

Treindienstleiders en verkeersleiders zijn hierin als operators verantwoordelijk voor het operationele management van de spoorweg. Door hun ervaring beschikken zij over unieke kennis van de kenmerken en dynamiek van dit complex sociaal-technisch systeem. Hierdoor is het cruciaal om een nieuw systeemontwerp met hun te testen en trainen, bijvoorbeeld door spelsimulaties.

Dit proefschrift onderzoekt middels spelsimulaties de rol van de menselijke operator in het grootschalig (her)ontwerpproces van het Nederlandse spoorwegsysteem.

Onderwerpen in dit proefschrift

De volgende onderwerpen komen in dit proefschrift aan bod: mentale modellen en situatiebewustzijn, (hybride) spelsimulaties en validiteit.

Mentale modellen en *situatiebewustzijn* zijn bekende cognitieve concepten binnen het vakgebied van human factors/cognitive engineering. Deze concepten kunnen gekoppeld worden aan de kennis van operators, waarbij mentale modellen vaak worden gedefinieerd als kennis die een operator heeft van een fysiek systeem (bijvoorbeeld wat is een wissel, hoe werkt een wissel etc.). Situatiebewustzijn bouwt voort op goed ontwikkelde mentale modellen, waarin deze kennis in context wordt geplaatst. Daarbij is situatiebewustzijn ook een indicator voor goede operationele besluitvorming.

Spelsimulaties oftewel *gaming simulaties* kunnen gebruikt worden als een onderzoeksomgeving om samen met gebruikers toekomstige systeemontwerpen te ontwikkelen, testen en trainen. Als één spelsimulatiesessie meerdere doeleinden (bijvoorbeeld ontwerpen, testen of trainen) heeft, is er sprake van een hybride spelsimulatie.

Validiteit is vooral relevant voor onderzoekssimulaties. Verkennende onderzoekssimulaties hebben als doel om een systeemontwerp te toetsen en daarbij inzichten te verkrijgen om het ontwerp te verbeteren. Verklarende/ hypothese toetsende onderzoekssimulaties hebben het doel om een systeemontwerp te toetsen en inzichten te verkrijgen in de verwachte impact van een toekomstig systeemontwerp. Hoewel verkennende onderzoeks-simulaties minder strenge valditeitseisen hebben, moeten onderzoekssimulaties die een hypothese over een systeemontwerp testen, voldoen aan strenge experimentele onderzoekseisen. Het hoogste validiteitssniveau is vereist bij bijvoorbeeld metingen van mentale modellen en situatiebewustzijn.

Dit proefschrift richt zich middels (hybride) spelsimulaties en een human factors perspectief op:

1. Onderzoek naar (individuele, team- en/of netwerk) mentale modellen en situatiebewustzijn bij treindienstleiders en verkeersleiders
2. Het verkrijgen van operatorkennis als feedback voor een systeemontwerp, waarbij validiteit van deze kennis gewaarborgd wordt via mentale modellen en situatiebewustzijn
3. Het gebruik van situatiebewustzijn theorieën om kunstmatige intelligentie te ontwikkelen die menselijke operators in spelsimulaties kunnen representeren

Structuur van dit proefschrift

Dit werk bestaat uit drie delen. Deel 1 richt zich op het theoretische uitgangspunt van dit proefschrift door literatuuronderzoek over spelsimulatie en validiteit (hoofdstuk 2) en spelsimulatietypen en mentale modellen en situatiebewustzijn (hoofdstuk 3).

- Hoofdstuk 2 bestaat uit literatuuronderzoek naar verschillende validiteitstypes in vakgebieden zoals informatica, sociale wetenschappen en human factors, en biedt een raamwerk dat de verschillende stadia en soorten validiteit vastlegt die vereist zijn in spelsimulaties.
- Hoofdstuk 3 bestaat uit literatuuronderzoek naar de cognitieve concepten van situatiebewustzijn en mentale modellen uit het human factors vakgebied en verschillende spelsimulatiesoorten. Er wordt een kader geboden dat spelsimulatiesoorten (bijvoorbeeld onderzoek of training) verbindt met mentale modellen en situatiebewustzijn.

Deel 2 richt zich op onderzoeken naar mentale modellen en situatiebewustzijn. Onderzoek naar mentale modellen en situatiebewustzijn van operators is cruciaal in het ontwerp en de optimalisatie van taken van de operator met bijvoorbeeld betrekking tot regels, procedures en infrastructuurwijzigingen. De verrichte onderzoeken beperken zich niet alleen tot de mentale modellen en situatiebewustzijn van individuele operators (hoofdstuk 4 en 5), maar richten zich ook op het teamniveau - bijv. operators binnen een regionaal verkeersleidingspost (hoofdstuk 6) en op het netwerkniveau - bijv. operators van zowel regionale als landelijke verkeersleidingspost(en) (hoofdstuk 7 en 8). De toepasbaarheid van verschillende theorieën over situatiebewustzijn werd ook onderzocht in hoofdstukken 5 tot en met 7.

- Hoofdstuk 4 onderzoekt de diversiteit in de doelen en strategische mentale modellen van treindienstleiders. De bevindingen toonden een incongruentie tussen het primaire doel van de organisatie en het primaire doel van de treindienstleiders, d.w.z. slechts 14% van de treindienstleiders was het eens met het primaire organisatiedoel: aankomstpunctualiteit. Een andere bevinding was dat treindienstleiders tot vijf verschillende afhandelingsstrategieën toepasten (d.w.z. een operationalisatie van een strategisch mentaal model) in een meer verstoorde toestand van het treinverkeer. Deze bevindingen gaven aanwijzingen voor een relatief sterke diversiteit in primaire operatordoelen en strategische mentale modellen. Deze variatie tussen treindienstleiders worden gezien als een indicatie voor zwakke veerkracht.
- Hoofdstuk 5 meet situatiebewustzijn volgens het bekende model van Endsley en de bijbehorende Situation Awareness Global Assessment Technique (SAGAT) meetmethode. Een hoge SAGAT-score duidt op een hoog situatiebewustzijnsniveau van een operator door expliciete, beredeneerde en bewuste processen. Er zijn echter aanwijzingen gevonden voor vooral impliciete situatiebewustzijn bij treindienstleiders door een reeks van analyses: (1) redelijk lage absolute waarden (ca. 45%) van de SAGAT-score met relatief hoge prestaties (bijv. gemiddelde punctualiteit van treinen van 88%) en correlaties tussen SAGAT-scores en meerdere prestatie-indicatoren, (2) de negatieve relatie tussen werkervaring en SAGAT-scores en (3) afwijkingen tussen de drie SA-niveaus (SA-scores van niveau 1 waren lager dan SA-scores van niveau 2). Deze bevindingen wezen op de aanwezigheid van intuïtieve, onbewuste processen, d.w.z. impliciete situatiebewustzijn in tegenstelling tot een bewust, beredeneerd proces bij treindienstleiders.
- Hoofdstuk 6 onderzoekt de samenwerking en ontwikkeling van situatiebewustzijn op een regionale verkeersleidingspost. Een onverwachte bevinding was dat de team-identiteit niet gelinkt was aan de verkeersleidingspost, maar werd geassocieerd door overlappende geografische werkgebieden en gedeelde displays tussen treindienstleiders. De bevindingen gaven aan dat de ontwikkeling van teamsituatiebewustzijn werd bevorderd door zowel de gedeelde displays, communicatie als coördinatie tussen treindienstleiders.
- Hoofdstuk 7 richt zich op werkbelasting en situatiebewustzijn op netwerkniveau, waarbij sociale netwerkanalyse is gebruikt als een techniek om zowel de communicatiestroom als inhoud te onderzoeken. De bevindingen toonden zogenoemde 'spinnen' en 'gatekeepers' in het netwerk aan. Tijdens een storing heeft de treindienstleider in het verstoorde gebied een belangrijke rol en is deze 'gatekeeper' van informatie tussen verschillende subgroepen in het netwerk. Opvallend is

dat grotendeels vergelijkbare informatie over de gehele spoorketen is gedeeld. De decentrale verkeersleider is typisch een 'spin' in het verkeerleidingsnetwerk door zijn/haar centrale positie. De identificatie van ‘spinnen’ en ‘gatekeepers’ kan bijzonder nuttig zijn bij het beoordelen van de efficiëntie en werkbelasting van het netwerk, met name met betrekking tot het aantal mislukte communicatiepogingen.

- Hoofdstuk 8 geeft inzicht in de ontwikkeling van gedeelde mentale modellen op netwerkniveau. De bevindingen van dit onderzoek hebben geleid tot drie stellingen over de ontwikkeling van gedeelde mentale modellen tussen operators in een spelsimulatie: (1) operators ontwikkelden eerst een gedeeld begrip van de spelomgeving, (2) operators ontwikkelden vervolgens een gedeeld begrip van elkaars doelen, (3) operators ontwikkelden ten slotte een gedeeld mentaal model van de kenmerken (sterke en zwakke punten) van (een deel van) het spoorwegsysteem en de bijbehorende implicaties voor verschillende soorten verstoringen.

Deel 3 richt zich op de toepassing van mentale modellen en situatiebewustzijn en spelsimulaties voor cognitieve modellen (hoofdstuk 9), het verkrijgen van operatorkennis in verschillende systeemontwerp stadia (hoofdstuk 10), en implicaties van validiteit in hybride spelsimulaties (hoofdstuk 11).

- Hoofdstuk 9 richt zich op het ontwikkelen van cognitieve modellen van operators voor spelsimulatie op basis van situatiebewustzijn. Bij het uitvoeren van spelsimulaties is de beschikbaarheid van operators vaak schaars. Eén van de oplossingen voor dit probleem is de toepassing van kunstmatige intelligentie die beslissingen namens een afwezige operator kan nemen. Aangezien situatiebewustzijn een indicator is voor goede besluitvorming is in dit hoofdstuk onderzocht hoe theorieën van situatiebewustzijn gebruikt kunnen worden om een cognitief model van een operator (agent) te ontwikkelen. Agent-based modelleren wordt gebruikt als een techniek om een eerste cognitief model van een landelijke verkeersleider te ontwikkelen.
- Hoofdstuk 10 richt zich op de rol van mentale modellen en situatiebewustzijn bij het verkrijgen van operatorkennis. Bij het (her)ontwerp van het spoorwegsysteem wordt de professionele kennis van gebruikers middels het ‘participatory design’ perspectief erkend als belangrijke input voor strategische besluitvorming. Daarbij kan de kennis van operators gebruikt worden om het systeemontwerp vorm te geven en te optimaliseren.

  Het hoofdstuk biedt een raamwerk, waarin verschillende fasen van het systeemontwerp gekoppeld zijn aan verschillende eisen wat betreft de validiteit van de spelsimulatie en representativiteit van deze operatorkennis (op basis van mentale modellen of situatiebewustzijn). Een spelsimulatie die vergelijkbaar is met de werkelijke werkomgeving leidt over het algemeen tot een hoog situatiebewustzijn. Wanneer kennis van operators afkomstig is van een hoog situatiebewustzijn, zal deze kennis rijke omgevings- en situationele details bevatten met een hogere nauwkeurigheid in de beschrijving en daarmee een hogere geldigheid. Om dit validiteitsniveau te bereiken, moet echter veel moeite worden gedaan om validiteitsbedreigingen in het onderzoek te reduceren. Spelsimulaties die daarentegen afwijken van de gelijkenis met de werkelijke werkomgeving, door bijvoorbeeld stapsgewijze besluitvormingsrondes te gebruiken in plaats van real-time, zullen niet tot een representatief situatiebewustzijn leiden. In plaats daarvan zal de verkregen operatorkennis gebaseerd zijn op hun mentale modellen.

Het raamwerk kan door strategische besluitvormers en spelsimulatieontwerpers worden gebruikt om de ontwerpeisen van een spelsimulatie te bepalen. Bij het definiëren van de ontwerpeisen van de spelsimulatie spelen strategische besluitvormers een belangrijke rol. Onderzoeksvragen dienen daarbij scherp geformuleerd te worden. Daarnaast kunnen besluitvormers het onderzoeksontwerp beïnvloeden door het beschikbaar stellen van middelen (bijvoorbeeld deelnemers). Ontwerpers van spelsimulatie zijn niet verantwoordelijk voor de scope en middelen, maar wel verantwoordelijk voor de output. Het raamwerk biedt ook inzichten in de ontwerpvoorwaarden en eisen waaraan moet worden voldaan om een bepaald kwaliteitsniveau te bereiken. Spelsimulatieontwerpers kunnen dit raamwerk gebruiken om een optimale spelsimulatie te ontwerpen.

- Hoofdstuk 11 onderzoekt de spelsimulaties uit hoofdstukken 4 tot en met 8, die meerdere doeleinden in een enkele spelsimulatie hebben gecombineerd. Deze veelzijdige of hybride spelsimulaties ontstaan door (1) meerdere onderzoeksvragen in één enkele spelsimulatie te onderzoeken, (2) het testen van een toekomstig systeemontwerp zonder deze voorafgaand getraind te hebben en (3) de openheid van de spelsimulatie. De meest voorkomende vorm was het bestaan van meerdere type vragen in één enkele spelsimulatie om zowel vragen vanuit de organisatie, als academische (human factors) onderzoeksvragen te beantwoorden.

De uitdaging in het ontwerp van hybride spelsimulaties is het waarborgen van de validiteit van deze spelsimulaties om de spelsimulatiedoeleinden te kunnen bereiken. Voor het ontwerpen van verklarende onderzoekssimulatie wordt verder in dit hoofdstuk een synthese van verschillende raamwerken gegeven. Deze richten zich o.a. op het ontwerp van de onderzoeksopzet, het ontwerp van de spelsimulatiesessie en de mate van experimentele controle in deze twee fasen.

Hoofdstuk 12 sluit af met een discussie en conclusie. In dit hoofdstuk worden de bevindingen uit de onderzoeken en raamwerken samengevat, worden onderzoeksvragen beantwoord en een algemene conclusie gegeven.

List of publications

Journal articles

Bekebrede, G., Lo, J. C. & Lukosch, H. K. (2015). Understanding complexity: The use of simulation games for engineering systems. *Simulation & Gaming 46(5)*, 447-454.

Bekebrede, G., Lo, J. C. & Lukosch, H. K. (2015). Understanding complex systems through mental models and shared experiences: The case of SIMPORT-MV2. *Simulation & Gaming 46(5)*, 536-562.

Lankveld, G. van, Sehic, E., Lo, J. C. & Meijer, S. A. (2017). Assessing gaming simulation validity for training traffic controllers. *Simulation & Gaming 48(2)*, 219-235.

Lo, J. C., & Meijer, S. A. (2020). Assessing network cognition in the Dutch railway system: insights into network situation awareness and workload using social network analysis. *Cognition, Technology & Work, 22*(1), 57-73.

Lo, J. C. & Meijer, S. A. (in review). Participatory design in large-scale railway infrastructure using gaming simulations: The role of shared mental models.

Lo, J. C., Philipsen, E. P. & Schraagen, J. (in review). Using hierarchical task analysis to support the digital transformation of railway traffic control: Applications and limitations.

Lo, J. C., Pluyter, K. R. & Meijer, S. A. (2016). Individual markers of resilience in train traffic control: The role of operators' goals and strategic mental models and their implications for variation, expertise and performance*. Human factors 58(1)*, 80-91.

Lo, J. C., Sehic, E., Brookhuis, K. A. & Meijer, S. A. (2016). Explicit or implicit situation awareness? Measuring the situation awareness of train traffic controllers. *Transportation research part F: traffic psychology and behaviour 43*, 325-338.

Lo, J. C., Sehic, E., & Meijer, S. A. (2019). Balancing organizational and academic research: Investigating train traffic Controller's geographical workspace design and team situation awareness using gaming simulations. *Journal of Rail Transport Planning & Management, 10*, 34-45.

Lo, J. C., Sehic, E. & Meijer, S. A. (2017). Measuring mental workload with low-cost and wearable sensors: Insights into the accuracy, obtrusiveness, and research usability of three instruments. *Journal of Cognitive Engineering and Decision Making 11(4)*, 323-336

Mayer, I. S., Bekebrede, G., Harteveld, C., Warmelink, H. J. G., Zhou, Q., Ruijven, T. W. J. van, Lo, J. C., Kortmann, L. J. & Wenzler, I. (2014). The research and evaluation of serious games: Toward a comprehensive methodology. *British Journal of Educational Technology, 45(3),* 502-527.

Mayer, I. S., Zhou, Q., Lo, J. C., Abspoel, L., Keijser, X., Olsen, E., Nixon, E. & Kannen, A. (2013). Integrated, ecosystem-based marine spatial planning:

Design and results of a game-based, quasi-experiment. *Ocean & Coastal Management, 82*, 7-26.
Page, L., Lo, J., Velazquez, M. & Claudio, D. (in review). Introducing a validation framework to standardize and improve human factors and ergonomics research.
Van den Hoogen, J., Lo, J. C. & Meijer, S. A. (2016). Debriefing research games: Context, substance and method. *Simulation & Gaming 47(3)*, 368-388.

Conference proceedings

Aydogan, R., Lo, J. C., Meijer, S. A. & Jonker, C. M. (2014). Modeling network controller decisions based upon situation awareness through agent-based negotiation. In S. Meijer & R. Smeds (Eds.), *Frontiers in Gaming Simulation* (pp. 191-200). Berlin: Springer.
Aydogan, R., Sharpanskykh, O. A. & Lo, J. C. (2014). A trust-based situation awareness model. In M. A. Janssen, L. Na'ia Alessa & C. M. Barton (Eds.), *Proceedings of the 12th European Conference on Multi-Agent Systems* (pp. 1-16).
Harteveld, C., Bekebrede, G., Lo, J. C., Plomber, A. J. & Jordaan, B. (2012). Make it fun or real: Design dilemmas and their consequences on the learning experience. In W. Bielecki, J. Gandziarowska-Ziolecka, A. Pikos & M. Wardaszko (Eds.), *Facing the Challenges of the Globalizing World with the Use of Simulation and Gaming* (pp. 153-164). Warsaw: Poltext.
Hillege, R. H. L., Lo, J. C., Romeijn, N., & Janssen, C. P. (in press). The mental machine: Classifying mental workload state from unobtrusive heart rate measures using machine learning.
Lo, J. C. & Meijer, S. A. (2014). Gaming simulation design for individual and team situation awareness. In S. A. Meijer & R. Smeds (Eds.), *Frontiers in Gaming Simulation* (pp. 121-128). Berlin: Springer.
Lo, J. C. & Meijer, S. A. (2013). Measuring group situation awareness in a multi-actor gaming simulation: A pilot study of railway and passenger traffic operations. In A. Bisantz (Ed.), *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (pp. 177-181). Los Angeles: SAGE.
Lo, J. C. & Meijer, S. A. (2013). Situation awareness measurement techniques for gaming simulations: An overview and application for railway traffic controllers. *Proceedings of the International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (COGSIMA)* (pp. 238-245). Piscatway: IEEE.
Lo, J. C., Sehic, E. & Meijer, S. A. (2014). Explicit or implicit situation awareness? Situation awareness measurements of train traffic controllers in a monitoring mode. In D. Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics* (pp. 511-525). Berlin: Springer.
Lo, J. C., Van den Hoogen, J. & Meijer, S. A. (2014). Testing changes in the railway system through gaming simulation: How different types of innovations affect operators' mental models. In T. Ahram & T. Marek (Eds.), *Proceedings of the 5th International Conference on Applied Human Factors and Ergonomics (AHFE)* (pp. 8054-8065).

Lo, J. C., Van den Hoogen, J. & Meijer, S. A. (2013). Using gaming simulation experiments to test railways innovations: Implications for validity. In R. Pasupathy, S. H. Kim, A. Tolk, R. Hill & M. E. Kuhl (Eds.), *Proceedings of the 2013 Winter Simulation Conference (WSC)* (pp. 1766-1777).

Mayer, I. S., Zhou, Q., Lo, J. C., Abspoel, L., Keijser, X., Olsen, E., Nixon, E. & Kannen, A. (2012). Integrated, ecosystem-based marine spatial planning: First results from an international simulation-game experiment. *3rd International Engineering Systems Symposium (CESUN)* (pp. 1-10).

Roungas B., Lo J.C., Angeletti R., Meijer S., & Verbraeck A. (2019). Eliciting requirements of a knowledge management system for gaming in an organization: The role of tacit knowledge. In: R. Hamada et al., (Eds.), *Neo-Simulation and Gaming Toward Active Learning. Translational Systems Sciences, 18*. Singapore: Springer.

Van den Hoogen, J., Lo, J. C. & Meijer, S. A. (2014). The debriefing of research games: A structured approach for the validation of gaming simulation outcomes. In W. C. Kritz (Ed.), *Proceedings of the 45th Conference of the International Simulation and Gaming Association (ISAGA)* (pp. 599-610).

Van den Hoogen, J., Lo, J. C., & Meijer, S. A. (2014). Debriefing in gaming simulation for research: Opening the black box of the non-trivial machine to assess validity and reliability. In A. Tolk, S. Y. Diallo, O. Ryzhov, L. Yilmaz, S. Buckley & J. A. Miller (Eds.), *Proceedings of the 2014 Winter Simulation Conference (WSC)* (pp. 3505-3516). Piscataway: IEEE.

Conference posters

Albers, S., De Groot, B., Lo, J., Mug, J., Philipsen, E., Sehic, E., Vermeire, J. & Van 't Woudt, C. (2017). Testing a high frequency timetable concept with railway traffic operators in a multi-actor simulation setting. *Poster presentation at the 5th International Rail Human Factors Conference*.

Albers, S., Lo, J., Sehic, E. & Van 't Woudt, C. (2018). Development & use of a multi-actor simulation environment for Dutch railways. *Poster presentation at WinterSim 2018.*

Lo, J. C. & Meijer, S. A. (2015). Assessing network cognition in the dutch railway system through communication: Theoretical and methodological approaches using social network analysis. *Poster Presentation at the 94th Annual Meeting Transport Research Board (TRB).* Washington D.C., USA.

Lo, J. C. & Meijer, S. A. (2014). A framework to assess cognition in different types of gaming simulations. *Poster Presentation at the Human Factors and Ergonomic Society Europe Chapter*.

Lo, J. C., Sehic, E. & Meijer, S. A. (2015). Mental workload measurements through low-cost and wearable sensors. *Poster Presentation at the Human Factors and Ergonomic Society Europe Chapter*.

Lo, J. C., Sehic, E., van Luipen, J. J. W. & Meijer, S. A. (2015). Railway traffic control in the netherlands: Recent and future human factors developments at ProRail. *Poster presentation at the 5th International Rail Human Factors Conference*.