University of Groningen

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

Lo, Julia C.; Sehic, Emdzad; Brookhuis, Karel A.; Meijer, Sebastiaan A.

Published in:
Transportation Research. Part F: Traffic Psychology and Behaviour

DOI:
[10.1016/j.trf.2016.09.006](https://doi.org/10.1016/j.trf.2016.09.006)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2016

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

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. <https://doi.org/10.1016/j.trf.2016.09.006>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

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

Julia C. Lo^{a,b,*}, Emdzad Sehic^{a,b,1}, Karel A. Brookhuis^{a,c,2}, Sebastiaan A. Meijer^{a,d,3}

^a Delft University of Technology, Department of Multi-Actor Systems, Delft, The Netherlands

^b ProRail, Department of Innovation and Development, Utrecht, The Netherlands

^c University of Groningen, Department of Psychology, Groningen, The Netherlands

^d KTH Royal Institute of Technology, Department of Health Systems Engineering, Stockholm, Sweden

ARTICLE INFO

Article history:

Received 22 December 2015

Received in revised form 27 June 2016

Accepted 2 September 2016

Available online 4 October 2016

Keywords:

Decision-making

Dynamic environment

Intuition

Situation awareness

Socio-technical system

Tacit knowledge

ABSTRACT

Previous research on situation awareness (SA) predominantly focused on its explicit, reasoned, conscious features rather than on the implicit, intuitive, unconscious aspects that are often identified with expert operators. This research investigated implicit levels of SA of train traffic controllers (TTCs) in order to contribute to the body of knowledge on rail human factors research and SA. 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.

© 2016 Elsevier Ltd. All rights reserved.

* Corresponding author at: Moreelsepark 3, 3511 EP Utrecht, The Netherlands.

E-mail addresses: julia.lo@prorail.nl (J.C. Lo), emdazad.sehic@prorail.nl (E. Sehic), k.a.brookhuis@rug.nl (K.A. Brookhuis), smeijer@kth.se (S.A. Meijer).

¹ Address: Moreelsepark 3, 3511 EP Utrecht, The Netherlands.

² Address: Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands.

³ Address: Alfred Nobels Alle 10, 14 152 Stockholm, Sweden.

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.

1.1. 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).

1.2. 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 (Knies, 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 adaptations 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 train paths. 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.

2. 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.

3. 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 undistruptive, and therefore took on a monitoring role during the scenarios. This study has also been described in Lo et al. (2014).

3.1. Method

3.1.1. 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 1 describes the characteristics of the simulator; see Lo et al. (2014) for a more elaborate description.

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.

3.1.2. 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.

3.1.3. 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 2. Probes were presented in a multiple-choice answering format in line with Strater, Endsley, Pleban, and Matthews (2001).

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, $\alpha = .43$) 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 ($\alpha = .75$), 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 ($\alpha = .73$), 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.

Table 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

Table 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: ...

Mental workload. Five workload items based on the NASA-TLX (Hart & Staveland, 1988) were administered after each scenario ($\alpha = .81$). 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.

3.1.4. Procedure

The session started with a general introduction on the purpose of the simulator session and a description of the simulator's functionalities (see Fig. 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.

3.2. 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 ($M = 4.4$, $SD = 1.01$).

3.2.1. 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.

3.2.2. 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.

Fig. 1. Simulator setup with the simulator set with the new timetable (foreground) and the set with the current timetable (background).

3.2.3. Simulator validity

The findings show that participants have a rather positive to a positive perception of the simulator (see Table 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.

3.2.4. Situation awareness and performance

Table 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.

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.

3.2.5. 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 ($\rho = .64$, $p = .06$). 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, $\rho = .32$, $p = .42$. 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 ($\rho = .37$, $p = .33$; $\rho = -.42$, $p = .27$). 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 $\rho = .68$, $p = .04$; $\rho = .66$, $p = .08$. 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; $\rho = -.57$, $p = .14$, 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, $\rho = -.54$, $p = .14$ and a moderate negative correlation in scenario 2, $\rho = -.46$, $p = .22$; 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 $\rho = -.60$, $p = .12$ and $\rho = -.68$, $p = .06$. Similar correlations were also found in scenario 2, respectively $\rho = -.36$, $p = .28$ and

Table 3

Validity dimensions of the simulator for the current task in the pilot study.

	<i>N</i>	<i>M</i>	<i>SD</i>
Structural validity	9	3.7	.45
Process validity	9	3.9	.47
Psychological reality	9	3.9	.40

Table 4

Measurements of situation awareness, performance and mental workload.

	Scenario 1				Scenario 2			
	<i>N</i>	Mean rank	<i>M</i>	<i>SD</i>	<i>N</i>	Mean rank	<i>M</i>	<i>SD</i>
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.

$\rho = -.64$, $p = .03$. 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, 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.

3.3. 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).

4. 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.

Table 5SA 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

4.1. Method

4.1.1. 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 6](#) describes the design characteristics of the simulator session.

4.1.2. 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.

4.1.3. 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* ($\alpha = .83$).

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, $\alpha = .65$ after the removal of one item) and process validity ($\alpha = .60$). Psychological reality was measured with three items ($\alpha = .67$).

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.

4.1.4. 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 [Fig. 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

Table 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

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.

4.2. 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 ($M = 3.9$, $SD = 1.04$).

4.2.1. 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.

4.2.2. 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.

4.2.3. 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 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.

4.2.4. Situation awareness and performance

Variables related to SA, mental workload, and performance are provided in Table 8 in terms of mean ranks and means to provide an overview of the normal and non-normal distributed values.

Fig. 2. Simulator setup with the “turn” and “through” TTCs and two facilitators.

Table 7

Validity dimensions of the simulator for the current task in the main study.

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

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).

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.

4.2.5. 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; $\rho = -.49, p = .18$.

Table 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 ^{a,b}	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 ^{a,b}	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 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 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 level-1 SA scores might support the notion of implicit SA.

4.3. 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.

Table 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. General discussion and conclusion

5.1. 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).

5.2. 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.

5.3. 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).

5.4. 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 decision-support 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.

Acknowledgments

This work was funded through the Railway Gaming Suite program, a joint project by ProRail and Delft University of Technology. The authors thank Kari Pluyter for the data preparation of the main study, Giel van Lankveld for his involvement in the simulator sessions of the main study, Alexander Verbraeck for the fruitful discussions, and Gert-Jan Stolk, Berend Wouda and many other colleagues from the GameLab for their support in the development of the simulator.

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. Zsombok & G. Klein (Eds.), *Naturalistic decision making* (pp. 269–283). Mahwah, NJ: Lawrence Erlbaum Associates.
- ESSAI (2000). *WPI orientation on situation awareness and crisis management*. Workpackage Report. ESSAI/NLR/WPR/WP1. EC DG-TREN. Contract No.: 2000-GRD1-10450.
- 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., McLernon, 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. Zsombok & 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. In *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.