Situation awareness based on eye movements in relation to the task environment

de Winter, J. C.F.; Eisma, Y. B.; Cabrall, C. D.D.; Hancock, P. A.; Stanton, N. A.

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

ORIGINAL ARTICLE

CrossMark

Situation awareness based on eye movements in relation to the task environment

J. C. F. de Winter[1] · Y. B. Eisma[1,2] · C. D. D. Cabrall[1,3] · P. A. Hancock[4] · N. A. Stanton[5]

Abstract
The topic of situation awareness has received continuing interest over the last decades. Freeze-probe methods, such as the Situation Awareness Global Assessment Technique (SAGAT), are commonly employed for measuring situation awareness. The aim of this paper was to review validity issues of the SAGAT and examine whether eye movements are a promising alternative for measuring situation awareness. First, we outlined six problems of freeze-probe methods, such as the fact that freeze-probe methods rely on what the operator has been able to remember and then explicitly recall. We propose an operationalization of situation awareness based on the eye movements of the person in relation to their task environment to circumvent shortfalls of memory mediation and task interruption. Next, we analyzed experimental data in which participants ($N = 86$) were tasked to observe a display of six dials for about 10 min, and press the space bar if a dial pointer crossed a threshold value. Every 90 s, the screen was blanked and participants had to report the state of the dials on a paper sheet. We assessed correlations of participants' task performance (% of threshold crossing detected) with visual sampling scores (% of dials glanced at during threshold crossings) and freeze-probe scores. Results showed that the visual-sampling score correlated with task performance at the threshold-crossing level ($r = 0.31$) and at the individual level ($r = 0.78$). Freeze-probe scores were low and showed weak associations with task performance. We conclude that the outlined limitations of the SAGAT impede measurement of situation awareness, which can be computed more effectively from eye movement measurements in relation to the state of the task environment. The present findings have practical value, as advances in eye-tracking cameras and ubiquitous computing lessen the need for interruptive tests such as SAGAT. Eye-based situation awareness is a predictor of performance, with the advantage that it is applicable through real-time feedback technologies.

Keywords Eye movements · Automatism · Operator state · Situated cognition

---

Joint first authors: J. C. F. de Winter and Y. B. Eisma.

✉ J. C. F. de Winter
  j.c.f.dewinter@tudelft.nl

1   Department of BioMechanical Engineering, Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, Delft, The Netherlands

2   Department of Control & Simulation, Faculty of Aerospace Engineering, Delft University of Technology, Delft, The Netherlands

3   Department of Cognitive Robotics, Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, Delft, The Netherlands

4   MIT2 Laboratories, Department of Psychology and the Institute for Simulation and Training, University of Central Florida, Orlando, FL, USA

5   Transportation Research Group, Faculty of Engineering and the Environment, Boldrewood Innovation Campus, University of Southampton, Southampton, UK

1 Introduction

1.1 Situation awareness

During the last three decades, an extensive body of research has appeared concerning situation awareness (SA). Although SA was initially characterized as "the buzzword of the '90s'" (Pew 1994), the term is now firmly embedded into the vocabulary of human factors and ergonomics. The construct of SA has received "strong endorsement" (Wickens 2015, p. 90) and is regarded as valuable in the research community (Parasuraman et al. 2008). At the same time, SA has its critics (Dekker 2015; Flach 1995) and its validity has been debated (Carsten and Vanderhaegen 2015; Millot 2015).

Interest in SA can be attributed to the fact that systems have become increasingly complex and automated (Hancock 2014; Parasuraman et al. 2008; Stanton et al. 2017). Wickens

([2008](#)) explained the growing importance of SA by noting that: "This trend reflects, on one hand, the growing extent to which automation does more, and the human operator often does (acts) less in many complex systems but is still responsible for understanding the state of such systems in case things go wrong and human intervention is required" (p. 397).

According to Endsley, SA reflects the extent to which the operator knows what is going on in their environment and is the product of mental processes including attention, perception, memory, and expectation (Endsley [2000a](#)). More formally, SA has been defined as "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. 792). Endsley's model of SA thus consists of three ascending levels (Endsley [2015a](#)). Level 1 denotes the perceptual process within the dynamic environment, Level 2 concerns a comprehension of those perceived elements from Level 1, and Level 3 SA is the projection of the future status.

1.2 The use and validity of the situation awareness global assessment technique (SAGAT)

Endsley ([2015b](#)) noted that "much of the disagreement on SA models that has been presented ultimately has boiled down to a disagreement on the best way to measure SA" (p. 108). It is a supportable assertion that the most often-used method to assess SA is the Situation Awareness Global Assessment Technique (SAGAT; Endsley [1988](#)). A Google Scholar search (August 2018) using the query "situation awareness global assessment technique" yielded 1850 papers, which proved to be considerably more than the number of hits for any competitor technique (e.g., "situation awareness rating technique" yielded 708 papers and "situation present assessment method" yielded 367 papers). SAGAT is a freeze-probe technique that requires operators to memorize and report on pre-defined aspects of their task environment via queries which interrogate aspects of either perception (Level 1 SA), comprehension (Level 2 SA), or projection (Level 3 SA). The higher the score with respect to a normative 'ground truth', the higher the operator's SA is considered to be.

As pointed out by Durso et al. ([2006](#)), "one of the arguments advanced for the importance of SA is that SA is a sensitive harbinger of performance" (p. 721). It has been shown that individual differences in task performance can be predicted from SAGAT scores to some extent. For example, it has been found that SAGAT scores correlate with performance on a military planning task ($r=0.66$, $N=20$; Salmon et al. [2009](#)), teamwork performance among medical trainees ($r=0.65$, $N=10$ teams; Gardner et al. [2017](#)), and performance in a surgical task ($r=0.47$, but two other correlations

were not statistically significant from zero, $N=97$; Bogossian et al. [2014](#), and $r=0.81$, $N=16$; Hogan et al. [2006](#)). SAGAT also relates to how well pilots handled in-flight emergencies in a simulator ($r=0.41$, $N=41$; Prince et al. [2007](#)), crash-avoidance performance in a low-fidelity driving simulator ($r=0.44$, $N=190$; Gugerty [1997](#)), scores on a driving-based hazard perception test ($r=0.56$, N about 38; McGowan and Banbury [2004](#)), performance in submarine track management (β between -0.02 and 0.41, $N=171$; Loft et al. [2015](#)), and performance in air traffic control ($r=0.52$, $N=18$; O'Brien and O'Hare [2007](#)).

However, other studies are less positive regarding the validity of the SAGAT. Durso et al. ([1998](#)) found that SAGAT correlated only weakly with performance of air traffic controllers (β between -0.01 and 0.24, $N=12$), whereas Lo et al. ([2016](#), p. 335) found "a general tendency across conditions for a negative relation between SA probes and multiple performance indicators" ($N<10$). Similarly, Pierce et al. ([2008](#)) found that participants with higher SAGAT scores committed fewer procedural errors and violations in an air traffic control task, but these effects were not statistically significant (N about 20, $p \geq 0.08$). Similarly, Strybel et al. ([2008](#)) found no significant association between SAGAT scores and air traffic control performance ($N=13$). Additionally, Cummings and Guerlain ([2007](#)) found that overall performance scores in a missile control task were not statistically significantly correlated with SAGAT scores ($N=42$), whereas Ikuma et al. ([2014](#)) found no significant correlations between SAGAT scores and control room operator performance ($N=36$).

We argue that the above-mentioned small-sample correlations may not be statistically reliable, due to measurement error and possible selective reporting bias. According to the principle of aggregation (Rushton et al. [1983](#)), predictive validity is increased if the predictor and criterion are averaged across multiple measurement instances. Looking at the largest sample study (Gugerty [1997](#)), the relatively strong correlation of 0.44 could be due to the fact that SAGAT scores and performance scores were averaged across a large number of trials per person (84 or more).

From the above observations, the question arises as to whether some of the stronger predictive correlations are inflated due to common method variance. To illustrate, McGowan and Banbury ([2004](#)) observed that SAGAT scores were strongly predictive of hazard anticipation performance ($r=0.56$). This strong correlation is to be expected, as the term 'hazard anticipation' is often equated with SA (Horswill and McKenna [2004](#); Underwood et al. [2013](#)). McGowan and Banbury argued that the correlation could be even stronger than 0.56: "if all the probe queries were to measure projection then a higher correlation will be found". In other words, it is no surprise that responses to SAGAT queries (e.g., 'what will happen next' queries) show strong

associations with scores on a hazard anticipation test; the criterion and predictor variable are conceptually similar, and no independent performance is predicted. Also, it can be questioned whether the SAGAT has additional predictive validity, also called 'incremental validity' (Sechrest 1963), with respect to standard psychometric tests, such as tests of working memory and spatial ability (Pew 1994). This topic has been previously investigated by Durso et al. (2006). In a study using 89 participants, they found that SAGAT was not a sufficiently strong predictor of air traffic control performance to enter a stepwise regression model after diverse cognitive and non-cognitive tests had been allowed to enter the model first. This led these authors to conclude that "typical cognitive measures already capture much of what off-line measures contribute" (p. 731). Indeed, it is known that psychometric test scores show positive inter-correlations (Van der Maas et al. 2006), and it is plausible that operators who possess high working memory capacity will perform well on any task, and thus will perform well on the SAGAT also (Gugerty and Tirre 2000; Sohn and Doane 2004). In other words, a statistical association between SAGAT scores and task performance may be due to a common cause such as general intelligence (*g*) rather than anything that is necessarily situational.

1.3 Aim of this study

As indicated above, the SAGAT is a widely used freeze-probe technique. SAGAT scores appear to be moderately correlated with task performance, while incremental validity is contentious. At present, it is unknown why the SAGAT has imperfect validity with regard to task performance. Accordingly, the research question that this paper sets out to answer is: "What are the limitations of SAGAT?", and secondly: "Is an alternative body-based measure of SA more predictive of task performance than a freeze-probe method?" More specifically, we propose here that SA can alternatively be operationalized via eye movements of the operator in relation to the task environment.

The idea of using eye-trackers for inferring SA is not a new one per se. In their work, "Development of a novel measure of situation awareness: The case for eye movement analysis", Moore and Gugerty (2010) found that the higher the percentage of time air traffic controllers fixated on important aircraft, the higher their task performance and SAGAT performance. Our present work aims to follow up on this type of analysis by focusing on eye movements in a dynamic environment. We postulate that eye movements reflect the extent to which an operator exerts a grip on the current environment (cf. Merleau-Ponty 1945) as part of the perception–action cycle (Neisser 1976), thus also being a predicate of task performance. In order to establish the concept of SA by means of eye movements and task relations,

we have included the results of an experiment with 86 participants who performed a visual monitoring task of a dynamic system. We examined the correlations between a freeze-probe method and eye-based SA on the one hand, and task performance, on the other.

2 Problems with SAGAT

When using SAGAT, the ongoing task is frozen and the simulation screen is blanked out. The operator then answers queries about the task environment. SAGAT queries need not necessarily be textual (see Endsley 2000b, for a review). An example of non-textual queries is the work of Gugerty (1997) in which participants had to pinpoint the location of cars in a top-down view of the simulated road.

Six problems arise from the SAGAT, and they can be considered common to all freeze-probe techniques: (1) memory decay/bias, (2) task resumption deviations, (3) removal from the ongoing task, (4) explicit representations, (5) intermittency, and (6) non-situated cognition.

First, there is an inherent and inevitable time delay between the moment of freezing and the subsequent completion of all the required queries. This makes such measurements susceptible to memory decay and the biases associated with it. Thus, the most immediate and familiar situational features are remembered best (and these do not necessarily reflect those with the greatest task relevancy). Gugerty (1998) found that "information was forgotten from dynamic spatial memory over the 14 s that it took participants to recall whole report trials" (p. 498).

Second, after the simulation freezes, participants have to subsequently resume the task, and so post-freeze task performance and SA almost certainly deviate from non-interrupted task performance. It has been argued by Endsley (1995) that these two problems may not be fatal to measuring SA; she empirically found that the length of the time interval and task interruption have only minor effects on SAGAT scores. McGowan and Banbury (2004), on the other hand, found a negative effect of SAGAT interruption on task performance as compared to the same task without interruption.

Third, as most researchers in general seem to agree that SA refers to "the level of awareness that an individual has of a situation" (Salmon et al. 2008, p. 297 *awareness*, the experience of awareness should ideally be reflected in the nature and character of the measurement method(s) themselves (Smith and Hancock 1995). How people respond to paper and pencil SAGAT queries, however, is only an indirect reflection of their phenomenological awareness, because they are removed from the situation by blanking the screen and interrupting the ongoing flow of behaviour. The task of completing SAGAT queries is temporally (i.e., the operator completes queries every few minutes while the simulator is
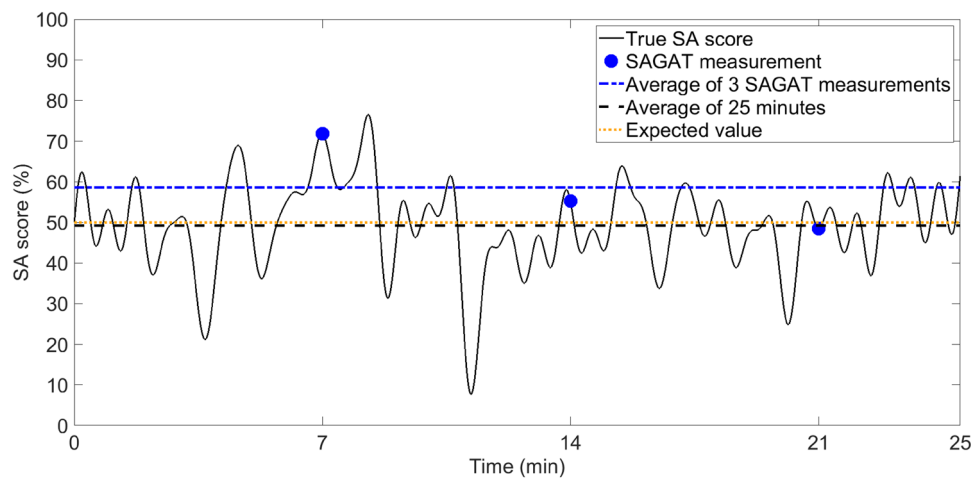
Fig. 1 Hypothetical illustration of a human's true SA score during a 25 min task. Three simulation freezes were assumed during which the SAGAT score was probed (at 7, 14, and 21 min). Here, we assumed that SA varies continuously, which is plausible, given that the state of technological systems (velocity, mass flow, etc.) is necessarily con-

tinuous due to laws of physics. However, SA could also change in discrete steps because the system state may manifest in discrete forms (e.g., warning lights) and because perception may resemble discrete steps also (as illustrated with multi-stable perception; Leopold and Logothetis 1999)

frozen) and functionally (i.e., the operator completes queries by means of a pencil, keyboard, or touchscreen) separate from the actual task.

Fourth, the SAGAT requires the participant to bring aspects of the task environment forward into conscious attention and to answer corresponding queries. However, what an operator reports in a query does not necessarily reflect his/her knowledge of the situation. According to dual-processing theories, which distinguish between unconscious (i.e., implicit, automatic) and conscious (i.e., explicit, controlled) processes (Evans 2003; Kahneman 2011; Kihlstrom 2008), it is the unconscious processes that are evoked based on situational triggers. Reflexes and instincts are the most basic examples of non-conscious behaviors in response to environmental stimuli. Implicit cognitive processes may also be acquired through practice. For example, after sufficient practice, drivers perform certain elementary tasks, such as changing gears, without overt conscious attention (Shinar et al. 1998, see also Morgan and Hancock 2008). Other familiar paradigms, such as the Stroop task, provide a further illustration that participants process the meaning of stimuli unconsciously, whether they want to or not. Endsley (1995) acknowledged that "data may be processed in a highly automated fashion and thus not be in the subject's awareness" (p. 72). However, she argued that the intrusion of unconscious processes represents only a small threat to SAGAT, by invoking three lines of reasoning. First, she argued that participants who fill out a SAGAT response sheet are able to extract situational content from long-term memory despite the fact that information has been processed automatically. Second, she reasoned that the multiple-choice response style of SAGAT facilitates access from memory,

as opposed to when being asked open-ended questions. The third argument was that participants are likely aware that they will complete a SAGAT query, which in turn enhances memorization and recall. Whether these assertions are true, and whether the recognition associated with the third argument does not interfere with memory capacity in the first place, requires further research. In sum, from the preceding observations, it would appear that the individual responds to environments often founded upon information not readily available to conscious introspection.

The fifth issue with SAGAT is that it measures SA intermittently rather than continuously, and therefore, it does not capture the dynamics of SA (Stanton et al. 2015). According to the law of large numbers, when administering the SAGAT on a small number of instances, one obtains a relatively imprecise estimate of the long-run expected value (Fig. 1). Moreover, when sampling at a limited rate, one does not capture higher frequencies in the signal. It is the fluctuations in SA that can be valuable sources of information for assessing cause-and-effect relationships regarding how changes of the environment, inter-operator communication, or task feedback influence SA.

Finally, the SAGAT task-freeze approach fails to take account of the situated cognition phenomenon (Stanton et al. 2015). People rely on artifacts to hold information on their behalf (Hutchins 1995; Sparrow et al. 2011). A study by Walker et al. (2009) comparing the communication modes of voice-only (i.e., no video, no data), video, and data-link in a distributed planning task showed that the SAGAT method could lead to the decision to use voice only. This was due to the fact that as the communication media became richer the SAGAT scores became poorer. As Stanton et al. (2015)

reported, "The explanation lies in that the greater the support from the environment, the less the person has to remember as the artifacts in the system hold the information" (p. 46). It seems a falsehood to divorce cognition from context. Similarly, Chiappe et al. (2015) argued that SAGAT is an inappropriate method to measure SA as blanking the screens prevents operators "from accessing externally represented information that they are used to obtaining in this way when engaged in a task" (p. 40).

3 Towards SA estimation from eye movements in relation to the task environment

We have indicated that it would be of considerable value to be able to assess SA in real-time. Here, we select eye movements as a candidate variable for the dynamic measurement of SA. The use of eye movement counteracts each of the above limitations of the SAGAT, as eye movement measurements are available on a continuous basis, can be obtained without interrupting or disturbing the ongoing task, do not require the operator to bring task elements to explicit memory, and are, therefore, free from issues of memory decay.

Humans rotate their eyes to orient the high-resolution fovea to the part of their scene that promises to render the greatest information. According to the eye-mind hypothesis, gaze direction is a strong correlate of cognitive activity (Just and Carpenter 1980; Yarbus 1967). Furthermore, according to the thesis of situated cognition, cognitive activity routinely exploits structure in the natural and social environment (Robbins and Aydede 2009). Given such an assumption, it should be feasible to identify some aspects of SA from eye-movements in relation to the task environment.

First, we illustrate the potential of eye movements through the lens of driving, which is a common task with strong safety implications (World Health Organization 2015). Driving is predominantly a visual task (Sivak 1996; Van der Horst 2004). In a review of more than half a century of driving safety research, Lee (2008) concluded that most crashes occur because "drivers fail to look at the right thing at the right time" (p. 525). Car driving involves much more than mere object detection, as drivers look ahead (i.e., 'preview') to anticipate and respond to what will happen next (e.g., Deng et al. 2016; Donges 1978). Research on how drivers extract relevant information from the task environment has often been reported under the heading of 'hazard perception' or 'hazard anticipation', which are terms now often equated with SA (Underwood et al. 2013; Horswill and McKenna 2004).

Recent research in this area has indicated that hazard precursors are discriminative between inexperienced and experienced drivers (Garay-Vega and Fisher 2005; Underwood

Fig. 2 A precursor used in previous SA research. Participants watch an unfolding scene. "This moped rider is about to pass a sports car with a driver in it and the front wheels turned to the left. If this sports car pulls out, the moped rider has to brake or swerve to the left. Has the participant driver noticed the sports car?" (from Vlakveld 2011)

et al. 2011). Precursors are visual cues that place critical demands on the driver's understanding and projection of an unfolding situation (cf. Levels 2 and 3 SA), such as the example shown in Fig. 2. Drivers with high SA are expected to be more likely to glance at the sports car (Level 1 SA), because the state of the sports car is informative about future collision risks (Levels 2 and 3 SA). Thus, in order to compute a driver's SA, an algorithm first has to establish critical features in the environment (e.g., a sports car is inching out), and whether the driver has attended to this feature. To clarify, a lot of eye movements in an environment with many task-relevant objects may signal high SA (because the driver scans these task-relevant objects), whereas the same eye movements in an environment with a small number of critical objects may signal low SA (i.e., the driver is distracted).

4 An empirical demonstration of measuring SA by means of eye movements in relation to the task environment

Here, we provide a demonstration by means of experimental results as to how SA can be extracted from eye movements in relation to task conditions. The results herein are based on an experiment presented in Eisma et al. (2018).

We used a visual sampling paradigm in which participants viewed a series of moving dials (Senders 1983). The participant's head was fixed via a head support (i.e., no postural changes). Thus, the human rotated the eyes to perceive the status of the display. Even though the task was chosen to be simple, it encapsulates the essential monitoring features of supervisory control of a dynamic system. This paradigm has its origins in a study by Fitts et al. (1950), which has been called "the first major Human Factors study" (Senders 2016).

Fig. 3 Screenshot of one of the seven videos. The dashed line is the threshold. The solid line is the pointer
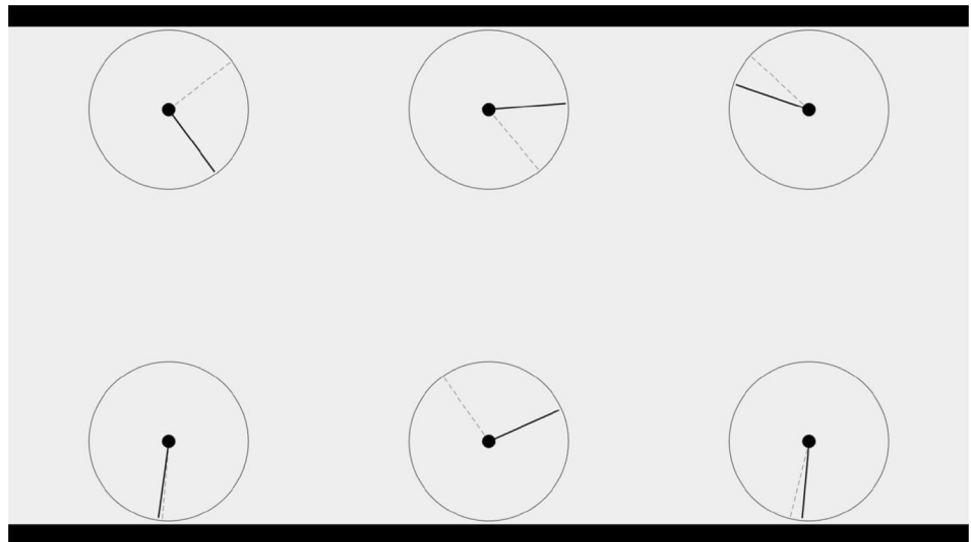
We express the amount of 'grip' on the environment as the percentage of resemblance between observed and ideal conditions, where 100% means optimal performance, and a low or zero percentage means that the operator's mind is wandering or the operator is asleep or unconscious, being completely disengaged or oblivious to the task. Accordingly, we define a 'sampling score' that defines how well the human observer has scanned the status of the dynamic displays.

4.1 Experimental methods

4.1.1 Participants

Participants were 86 university students (21 female, 65 male) with a mean age of 23.44 years (SD = 1.52) (Eisma et al. 2018). The original sample consisted of 91 participants, but data for five participants proved invalid due to computer faults, eye-tracker limitations, or data storage errors. The research was approved by the Ethics Committee of the TU Delft under the title 'Update of Visual Sampling Behavior and Performance with Changing Information Bandwidth'. All participants provided written informed consent.

4.1.2 Experimental tasks

Participants viewed seven 90-s videos on a 24-inch monitor having a resolution of 1920×1080 pixels. An EyeLink 1000 Plus was used to track the participants' eye movements. Each video showed six circular dials with moving pointers (as in Senders 1983). The pointer movement was a random signal with a bandwidth that differed between the six dials (0.03, 0.05, 0.12, 0.20, 0.32, and 0.48 Hz; as in Senders 1983). The threshold (dashed line, see Fig. 3) was a random angle that differed for each of the 42 dials (7 videos×6 dials). In

each of the seven videos, the pointer signals had a mean of 0 deg (i.e., relative to where the threshold was defined) and a standard deviation of 50.1 deg. The signal realization was different for each of the 42 dials, and the bandwidth ordering per dial was different for the seven videos. Each participant viewed the same seven videos in randomized order. An example video is provided in the supplementary materials.

4.1.3 Experimental procedures

Participants first completed a training of 20 s during which a single dial was shown. Participants were instructed to press the spacebar when a pointer crossed the threshold from either direction. The screen blanked after each video, and participants immediately completed a paper and pencil test about the current (Question 1), past (Question 2), and future (Question 3) states of the pointers (Fig. 4).

4.1.4 Dependent measures

First, we calculated a performance score per participant. This score was defined as the percentage of threshold crossings for which the participant pressed the spacebar. In total, there were between 74 and 115 threshold crossings per video. Per crossing, a 'hit' was counted if the participant pressed the spacebar within 0.5 s (i.e., between -0.5 and $+0.5$ s) of the moment of the crossing (Eisma et al. 2018). A spacebar press could not be assigned to more than one threshold crossing, and no more than one hit could be assigned to a threshold crossing.

Second, we calculated a visual sampling score per participant. This measure of SA was defined as the percentage of threshold crossings for which the participant fixated on a 420×420 pixel square surrounding the dial, within 0.5 s of the moment of the threshold crossing.
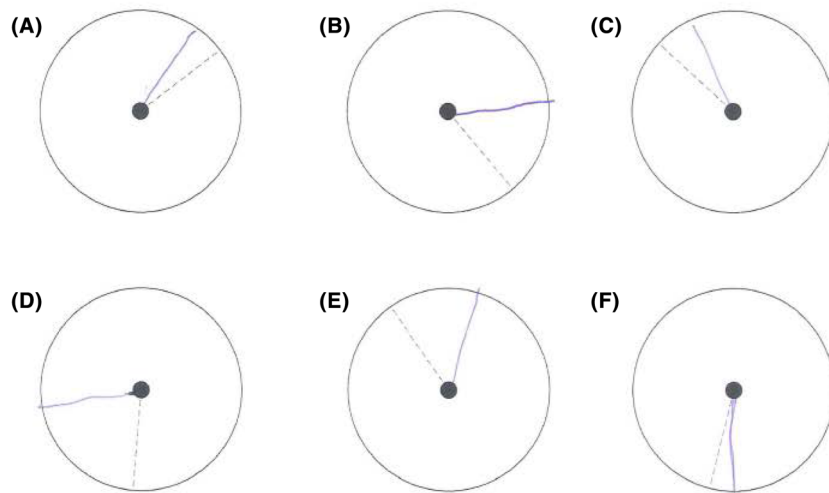
Fig. 4 The form completed by a participant (using a blue pencil) after one of the seven videos. In Question (1) participants drew a line, while in Questions (2–5) they circled an answer (Color figure online)

Third, we calculated a freeze-probe score for each participant. This score was defined as the percentage of 42 dials for which the participant drew a line on the correct side of the threshold.[1] The correct side meant that the line drawn by the participant occurred within the same clockwise or counterclockwise angular direction (i.e., from the threshold at 0° to ± 180°) as the 'ground truth' (i.e., the pointer position at the end of the video). If a participant did not draw a line (which happened in six out of 3588 dials) the score for this particular dial was marked as incorrect. We chose this binary definition (correct vs. incorrect side from the threshold) of the freeze probe score because alternative measures (e.g., absolute difference between the drawn angle and the threshold angle) may be prone to bias. More specifically, we observed that participants tended to draw the line near the threshold (if they were uncertain); this approach would yield a low error score (because the pointer indeed moves around the threshold) even when the participant was merely guessing. Furthermore, a binary scoring corresponds with the

SAGAT, where participants have to tick a response which can be either correct or incorrect.

For three of 86 participants, freeze-probe data were unavailable in one to two out of seven forms. Furthermore, for three other participants, due to computer/calibration issues, eye-tracking data for one to three out of seven videos were unavailable. These participants were retained in the analysis, using only relevant and acceptable data.

4.2 Experimental results

Participants viewing behavior was found to strongly relate to the state and dynamics of the dials. With high replication correspondence to the results of Senders (1983), glance frequency, dwell time, and dwell time per glance were evidenced as a function of task signal bandwidth (for details, see Eisma et al. 2018).

Table 1 shows a cross-tabulation of the sampling and performance score per threshold crossings. It can be seen that if a dial was not visually sampled in the right 1-s time frame (i.e., surrounding when a pointer crossed a threshold), then it was unlikely (28.4%) that the participant pressed the spacebar in that same 1-s time frame. Conversely, if a dial was sampled, then the participant pressed the spacebar in more than 50% (60.8%) of the threshold crossings. The phi coefficient (equivalent to the Pearson product-moment correlation

---

[1] We used image recognition in MATLAB to extract where participants had drawn the line. Participants used a blue pen, which could be relatively easily differentiated from the black/white background. The image recognition was found to have a mean accuracy of 0.14° (determined from the threshold which was printed on paper versus the known location of the threshold).

Table 1 Cross-tabulation of the number of times a dial was (not) sampled and a spacebar was (not) pressed, for each threshold crossing

	Dial not sampled	Dial sampled
Spacebar not pressed	13,135 (71.6%)	13,445 (39.2%)
Spacebar pressed	5208 (28.4%)	20,839 (60.8%)
Total	18,343 (100%)	34,284 (100%)

coefficient) between the visual sampling score and the performance score equaled 0.31. The correlation between the visual sampling score and the performance score at the level of participants was 0.78 (see Fig. 5, right).

The average freeze-probe score among participants was 57.7% (SD = 8.6%), which is slightly better than the expected value of 50% if participants were simply guessing. Participants had little confidence in their answers (Question 4 in Fig. 4): The average score was 4.08 (SD = 1.50) on the scale from 1 (very unsure) to 10 (very sure) (Fig. 4). Participants' freeze-probe score exhibited a moderate correlation with their performance score, $r = 0.20$ (Fig. 5, left).

The mean score on Question 2 (last dial) was 31.3% (SD = 18.6%) with respect to the last threshold crossing, and 29.6% (SD = 16.0%) with respect to the last space bar 'hit', whereas the mean score on Question 3 (next dial the participant will respond to) was 17.1% (SD = 13.6%), where 16.7% would be expected based on guessing alone. The scores on Questions 2 and 3 did not correlate significantly with the visual sampling score or freeze-probe score (all rs between −0.10 and 0.13).

In summary, we have shown that there is a moderate correlation ($r = 0.31$) between visual sampling and task performance at the level of threshold crossings, and a strong correlation at the level of participants ($r = 0.78$). Furthermore, it appears that participants had difficulty memorizing the state of the dials even though they filled out the form immediately after completing the task. In other words, how people sampled the dials was more strongly predictive of performance than what they memorized about the dials.

5 Discussion

5.1 Main findings

This paper aimed to outline several fundamental limitations of SAGAT and examine whether an eye-based measure of SA can be more predictive of task performance than a freeze-probe method. We argued that the SAGAT has the following limitations: (1) time delays between the freeze moment and the moment of answering the queries, (2) task interruption/disruption, (3) a disconnect from the ongoing task, (4) the need to bring the situation to conscious memory, (5) intermittent rather than continuous SA measurement, and (6) a failure to take situated cognition into account. Such fundamental limitations can help account for contentious empirical results regarding the validity of the SAGAT found in the literature (as reviewed in Sect. 1.2).

Building upon earlier work by Moore and Gugerty (2010), we have here shown that task performance can be predicted through eye-tracking measurements in relation to the state of the task environment in a more accurate manner than achieved by SAGAT. More specifically, correlations between visual sampling scores and performance scores were 0.31 at the level of threshold crossings and 0.78 at the level of individuals. In contrast, freeze-probe scores were low and showed weak associations with task performance. These results may be insufficiently compelling for real-time feedback applications, as the number of false positives and misses were rather high. However, we note that these calculations are binary (the timing or likelihood of glances were not considered), and therefore, there are multiple opportunities for improvement in both the sensitivity and specificity.

Fig. 5 The association between freeze-probe score and performance score (left panel, $r = 0.20$), and the association between visual sampling score and performance score (right panel, $r = 0.78$). Each marker represents a participant. The dashed line is a linear least-squares fit
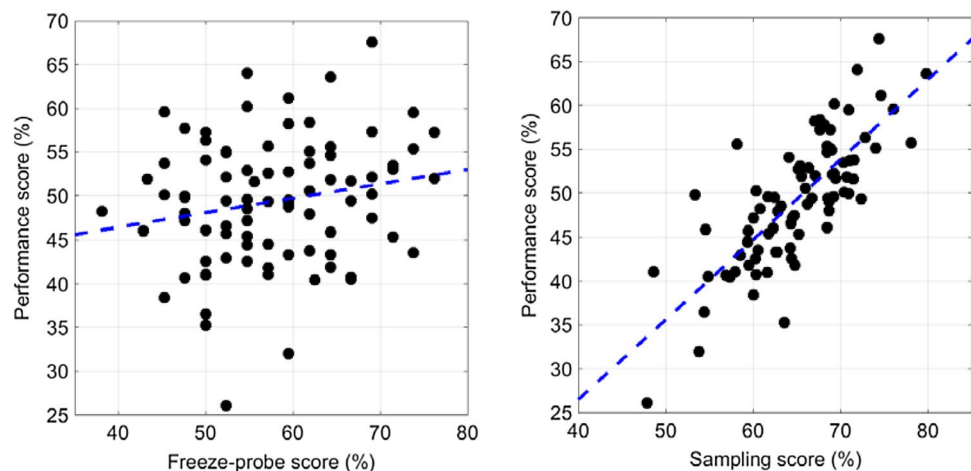
5.2 Hardware and software requirements

What hardware and software would be needed to implement a real-time SA assessment method based on eye movements in real-life situations? If the present approach were to be implemented in car driving, for example, high-end cameras would be needed that capture eye movements regardless of vibrations, lighting conditions, and driver's headgear such as caps, eyeglasses, and sunglasses. In the 1980s, physiological measurement tools were often bulky with limited capabilities (see Moray and Rotenberg 1989, for a study on human-automation interaction with gaze analyses at only 2 Hz). Consistent with Moore's law, however, (1965), computers have become considerably smaller and faster, and it is perhaps only a matter of time until we have the availability of ubiquitous eye-tracking cameras.

Additionally, the state of the environment has to be known. The ground truth could be human-generated as in SAGAT (choosing what to measure from the eyes and the task environment) or it could be computer-generated (e.g., using algorithms to determine what are relevant objects to look at). The latter approach requires databases (e.g., maps), sensors (e.g., cameras, radar), and analysis methods (e.g., instance segmentation of camera images). These capabilities are already being developed, for example for autonomous driving applications (Uhrig et al. 2016). A computer-generated ground truth should be able to establish that the turning of the sports car wheels shown in Fig. 2 is a hazard precursor, and that a situationally aware driver can be expected to have had their eyes towards this cue. Other operators (e.g., road users) may be part of the environment and so their states and dynamics should also be inputs for the model. Wickens et al. (2003, 2008) previously introduced a computational model of attention and SA based on the prior works of Senders. In their model, the probability of attending to an area is a weighted average of not only bandwidth as in Senders (1964, 1983), but also saliency (i.e., the conspicuity of information), effort (i.e., the visual angle between areas, where a larger angle is expected to inhibit scanning), and value (i.e., the importance of tasks served by the attended event). Attention to an area (i.e., Level 1 SA) is used to update human understanding of the current and future state of the system. This model appears to be a useful point of departure for developing a comprehensive algorithm for real-time SA assessment.

In real-life situations, multiple bodily signals (e.g., posture, see Riener et al. 2008) may need to be considered simultaneously as an input to a computational model, in order to infer SA. For example, it may be hard to extract SA related to strategies with long time constants from eye movements only. Additionally, the eye-mind hypothesis does not hold in a strong sense. In driving, a sizeable portion of collisions are caused by the looked-but-failed-to-see phenomenon, as well as related phenomena such as staring, mind wandering, and inattentional blindness (Herslund and Jørgensen 2003; White and Caird 2010). In other words, although the driver is fixated on a relevant stimulus, attention may covertly reside elsewhere. More research then appears to be needed to examine the validity of eye-based SA in complex supervisory tasks. In particular, it needs to be examined how eye-based SA can be employed in teams, especially in situations where different human actors and cognitive artifacts have conflicting information or intentions, and where task knowledge needs to be communicated between those agents (e.g., Salmon et al. 2008; Stanton et al. 2017; Vanderhaegen and Carsten 2017).

In sum, real-time SA assessment in outdoor environments is an engineering challenge, but not an unrealistic one considering the ongoing developments in sensors and artificial intelligence. So framed, our method is not fundamentally different from SAGAT, as both incorporate a comparison with ground truth. The difference is that SAGAT responses are explicitly reported by participants and cannot be extracted from veridical situations but only from simulated ones. In our case, the ground truth concerned the moments of threshold crossings of the pointer, whereas Moore and Gugerty (2010) defined specific aircraft as "important" within their air traffic control task environment upon which to evaluate the SA (estimation) construct. We recommend that researchers move beyond the use of paper and pencil tests of SA, and address and embrace the above developments to achieve the goal of ubiquitous SA assessment.

5.3 Differences from performance measurements and operator state assessments

Our proposal differs from performance-based measures of SA (Durso and Gronlund 1999; Gutzwiller and Clegg 2013; Prince et al. 2007; Sarter and Woods 1995). Performance-based SA suffers from circular reasoning, in the sense that it defines SA in terms of performance, but performance is what SA should prospectively predict in the first place (see Warm et al. 2008 recognizing the same paradox when mental resources are defined as task performance). Furthermore, in real-life tasks, such as supervision of highly automated systems, continuous performance measurements are often simply unavailable because the operator provides input only occasionally. In the present experiment, we asked participants to press the space bar when the pointers exceeded a threshold value. In reality, humans are often passive supervisors without an active performance task or overt responses to record.

Our approach also differs from operator-state assessment systems in general. For example, in driving, several sensor technologies exist that detect whether a driver is fatigued

or distracted (Barr et al. 2009; Blanco et al. 2009; Dong et al. 2011). Such systems may make use of measures of head movement, blink rate, eyelid closure, or gaze direction in any and all combinations and then provide feedback according to a multivariate algorithm (optionally combined with physiological and performance measures). The problem is that many of these systems measure the operator's behaviors without considering the environmental context in which behavior is embedded, and so may attack the issue of awareness per se, but do not reflect *situation* awareness specifically.

5.4 Future prospects

Hoffman and Hancock (2014) lamented that in many Human Factors investigations that are aimed at investigating why participants behave the way they do, researchers apparently never "bothered to ask the participants any questions after the experiment was over." Thus, there is clearly an inherent value in self-report and freeze-probe techniques for measuring SA, but we regard our approach to be in the long-term more promising and valuable for engineering applications that rely on real-time SA assessment, such as training and adaptive automation. Finally, we believe that the shortcomings of SAGAT, such as its reliance on memory skills and its disruption, also apply to many other SA procedures. For example, online probe measures, such as the situation awareness rating technique (SART), may be even more disruptive than SAGAT, but are likely less susceptible to issues of memory decay. As Salmon et al. (2006) noted, the SAGAT is "by far the most commonly used approach, and also the technique with the most associated validation evidence" (p. 228). Thus, it appears to be fair that we featured SAGAT as a target to which a new SA measure should be compared.

We have provided a demonstration as to how predictive-valid SA can be computed from eye movements and task features alone. From an engineering viewpoint, the human can be viewed as a machine (albeit a machine made of living tissue) and therefore all of a human's behavior has to have physical causes. The more accurate and information-rich the eye-movement and environment measurements become, the more opportunities arise for observing SA from these measurements. Concomitantly, the need for invoking indirect measures such as SAGAT then diminishes.

5.5 Limitations of the present experiment

The present task, in which participants had to watch a number of dials, may be regarded as arbitrary and unrepresentative of complex real-life situations such as control rooms and cockpits. However, our supervisory control task was intentionally designed to be abstract to provide a generic account of SA measurement. Moreover, our task replicated previous research of Senders (1983) and resembles the seminal work of Fitts et al. (1950), wherein pilots monitored a number of flight instruments (e.g., airspeed, directional gyro, engine instruments, altitude, vertical speed). We argue that our sampling task captures the essence of supervisory control—an area that Sheridan (1980) forecasted as increasingly relevant—in that operators have to monitor automation/instruments and detect anomalies (i.e., threshold crossings).

It may also be argued that our present freeze-probe measurement does not capture whether operators understand the situation (Level 2 SA) and anticipate what will happen (Level 3 SA). However, a review of the SAGAT shows that it is often used in simple tasks and includes simple items, such as items where participants have to recall the location of aircraft or cars (Endsley 2000b). That is, it seems that the use of our freeze-probe method does not fundamentally differ from the use of a typical SAGAT.

Participants performed poorly on the freeze-probe task and had little confidence in their answers. It is plausible that participants would score higher on freeze-probe queries if the supervisory task were interactive and meaningful (e.g., operating a nuclear power plant). As explained by Durso and Gronlund (1999), operators apply several strategies to reduce demands on working memory. Such strategies include focusing on the important information only, chunking of meaningful information, and restructuring the environment. Although our supervisory task did not allow for such strategies, our results do illustrate that participants were hardly able to remember the situation they had seen a few seconds before, a finding that is consistent with the notion that operators process information unconsciously (for explanation see Sect. 2). Eye-tracking seems a viable tool for measuring whether/when an operator has looked at specific objects (e.g., aircraft, cars), and provides a more direct indicator of SA than self-reported recall of the presence of objects or system states. Future research should establish whether SA based on eye movements in relation to the task environment can predict future, as opposed to concurrent performance, whether the criterion validity upholds in semantically rich tasks with longer time constants and correlated signals, and whether real-time feedback/control provided based on SA can enhance safety and productivity in operational settings.

Another limitation of the present study is that the participants were students at a technical university. As shown by Wai et al. (2009), engineering students score highly on intelligence-related tests, including tests of spatial ability. Accordingly, it is likely that engineering students have higher working memory capacity and would score better on the freeze-probe task than the general population. Because freeze-probe scores would likely be even lower in a sample that is representative of the entire population, our postulations and results against freeze-probe SA measurements are conservatively drawn. Another limitation of using

engineering students is restriction of range (Hunter et al. 2006). That is, because of the relatively homogenous sample, correlations between task-performance scores, visual sampling scores, and freeze-probe scores are likely attenuated as compared to correlations in a sample with a broad range of abilities. The issue of range restriction is especially pertinent for SA research, which is often concerned with specific groups of experts, such as pilots, military personnel, or air traffic control operators (Durso and Gronlund 1999).

6 Conclusions

It is concluded that the SAGAT suffers from time delays, task disruption, a disconnect from the ongoing task, a bias towards conscious recall, intermittent measurement, and a lack of measuring the situatedness of SA. We advanced a method to circumvent these limitations by calculating SA based on eye movements in relation to the task environment. We conclude that real-time SA based on eyes in relation to the task environment is moderately correlated with performance at the event level and strongly correlated with task performance at the level of individual participants.

Supplementary material

Supplementary material is available for download here: https://doi.org/10.4121/uuid:15d436f4-dbb2-468f-b66b-6d134ea2821c.

References

Barr L, Popkin S, Howarth H, Carroll RJ (2009) An evaluation of emerging driver fatigue detection measures and technologies: final report (report no. FMCSA-RRR-09-005). Volpe National Transportation Systems Center, Cambridge

Blanco M, Bocanegra JL, Morgan JF, Fitch GM, Medina A, Olson RL et al (2009) Assessment of a drowsy driver warning system for heavy-vehicle drivers: final report (report no. DOT 811). Virginia Tech Transportation Institute, Blacksburg, p 117

Bogossian F, Cooper S, Cant R, Beauchamp A, Porter J, Kain V et al (2014) Undergraduate nursing students' performance in recognising and responding to sudden patient deterioration in high psychological fidelity simulated environments: an Australian multi-centre study. Nurse Educ Today 34:691–696

Carsten O, Vanderhaegen F (2015) Situation awareness: valid or fallacious? Cogn Technol Work 17:157–158

Chiappe D, Strybel TZ, Vu KPL (2015) A situated approach to the understanding of dynamic situations. J Cogn Eng Decis Mak 9:33–43

Cummings ML, Guerlain S (2007) Developing operator capacity estimates for supervisory control of autonomous vehicles. Hum Factors 49:1–15

Dekker SW (2015) The danger of losing situation awareness. Cogn Technol Work 17:159–161

Deng T, Yang K, Li Y (2016) Where does the driver look? Top-down-based saliency detection in a traffic driving environment. IEEE Trans Intell Transp Syst 17:2051–2062

Dong Y, Hu Z, Uchimura K, Murayama N (2011) Driver inattention monitoring system for intelligent vehicles: a review. IEEE Trans Intell Transp Syst 12:596–614

Donges E (1978) A two-level model of driver steering behavior. Hum Factors 20:691–701

Durso FT, Gronlund SD (1999) Situation awareness. In: Durso FT, Nickerson RS, Schvaneveldt RW, Dumais ST, Lindsay DS, Chi MTH (eds) Handbook of applied cognition. Wiley, New York, pp 283–314

Durso FT, Hackworth CA, Truitt TR, Crutchfield J, Nikolic D, Manning CA (1998) Situation awareness as a predictor of performance for en route air traffic controllers. Air Traffic Control Q 6:1–20

Durso FT, Bleckley MK, Dattel AR (2006) Does situation awareness add to the validity of cognitive tests? Hum Factors 48:721–733

Eisma YB, Cabrall CDD, De Winter JCF (2018) Visual sampling processes revisited: replicating and extending Senders (1983) using modern eye-tracking equipment. IEEE Trans Hum Mach Syst. https://doi.org/10.1109/THMS.2018.2806200

Endsley MR (1988) Design and evaluation for situation awareness enhancement. Proc Hum Factors Ergon Soc Ann Meet 32:97–101

Endsley MR (1995) Measurement of situation awareness in dynamic systems. Hum Factors 37:65–84

Endsley MR (2000a) Theoretical underpinnings of situation awareness: a critical review. In: Endsley M, Garland DJ (eds) Situation awareness: analysis and measurement. Lawrence Erlbaum Associates, Mahwah, pp 3–32

Endsley MR (2000b) Direct measurement of situation awareness: validity and use of SAGAT. In: Endsley MR, Garland DJ (eds) Situation awareness analysis and measurement. Lawrence Erlbaum Associates, Mahwah, pp 147–174

Endsley MR (2015a) Situation awareness misconceptions and misunderstandings. J Cogn Eng Decis Mak 9:4–32

Endsley MR (2015b) Final reflections situation awareness models and measures. J Cogn Eng Decis Mak 9:101–111

Evans JSB (2003) In two minds: dual-process accounts of reasoning. Trends Cogn Sci 7:454–459

Fitts PM, Jones RE, Milton JL (1950) Eye movements of aircraft pilots during instrument-landing approaches. Aeronaut Eng Rev 9:24–29

Flach JM (1995) Situation awareness: proceed with caution. Hum Factors 37:149–157

Garay-Vega L, Fisher DL (2005) Can novice drivers recognize foreshadowing risks as easily as experienced drivers? Proceedings of the third international driving symposium on human factors in driver assessment, training and vehicle design. Rockport, ME

Gardner AK, Kosemund M, Martinez J (2017) Examining the feasibility and predictive validity of the SAGAT tool to assess situational awareness among surgical trainees. Simul Healthc J Soc Simul Healthc 12:17–21

Gugerty LJ (1997) Situation awareness during driving: explicit and implicit knowledge in dynamic spatial memory. J Exp Psychol Appl 3:42–66

Gugerty LJ (1998) Evidence from a partial report task for forgetting in dynamic spatial memory. Hum Factors 40:498–508

Gugerty LJ, Tirre WC (2000) Individual differences in situation awareness. In: Endsley MR, Garland DJ (eds) Situational awareness analysis and measurement. Erlbaum, Mahwah, pp 249–276

Gutzwiller RS, Clegg BA (2013) The role of working memory in levels of situation awareness. J Cogn Eng Decis Mak 7:141–154

Hancock PA (2014) Automation: how much is too much? Ergonomics 57:449–454

Herslund M-B, Jørgensen NO (2003) Looked-but-failed-to-see-errors in traffic. Accid Anal Prev 35:885–891

Hoffman RR, Hancock PA (2014) Words matter. Hum Factors Ergon Soc Bull 57:3–7

Hogan MP, Pace DE, Hapgood J, Boone DC (2006) Use of human patient simulation and the situation awareness global assessment technique in practical trauma skills assessment. J Trauma Acute Care Surg 61:1047–1052

Horswill MS, McKenna FP (2004) Drivers' hazard perception ability: situation awareness on the road. In: Banbury S, Tremblay S (eds) A cognitive approach to situation awareness. Ashgate, Aldershot, pp 155–175

Hunter JE, Schmidt FL, Le H (2006) Implications of direct and indirect range restriction for meta-analysis methods and findings. J Appl Psychol 91:594–612

Hutchins E (1995) How a cockpit remembers its speeds. Cogn Sci 19:265–288

Ikuma LH, Harvey C, Taylor CF, Handal C (2014) A guide for assessing control room operator performance using speed and accuracy, perceived workload, situation awareness, and eye tracking. J Loss Prev Process Ind 32:454–465

Just MA, Carpenter PA (1980) A theory of reading: from eye fixation to comprehension. Psychol Rev 87:329–354

Kahneman D (2011) Thinking, fast and slow. Farrar, Straus and Giroux, New York

Kihlstrom JF (2008) The automaticity juggernaut. In: Baer J, Kaufman JC, Baumeister RF (eds) Are we free? Psychology and free will. Oxford University Press, New York, pp 155–180

Lee JD (2008) 50 years of driving safety research. Hum Factors 50:521–528

Leopold DA, Logothetis NK (1999) Multistable phenomena: changing views in perception. Trends Cogn Sci 3:254–264

Lo JC, Sehic E, Brookhuis KA, Meijer SA (2016) Explicit or implicit situation awareness? Measuring the situation awareness of train traffic controllers. Transp Res Part F 43:325–338

Loft S, Bowden V, Braithwaite J, Morrell DB, Huf S, Durso FT (2015) Situation awareness measures for simulated submarine track management. Hum Factors 57:298–310

McGowan A, Banbury S (2004) Interruption and reorientation effects of a situation awareness probe on driving hazard anticipation. Proc Hum Factors Ergon Soc Ann Meet 48:290–294

Merleau-Ponty M (1945) Phenomenology of perception. Routledge, Abingdon

Millot P (2015) Situation awareness: Is the glass half empty or half full? Cogn Technol Work 17:169–177

Moore GE (1965) Cramming more components onto integrated circuits. Electronics 38:114–117

Moore K, Gugerty L (2010) Development of a novel measure of situation awareness: the case for eye movement analysis. Proc Hum Factors Ergon Soc Ann Meet 54:1650–1654

Moray N, Rotenberg I (1989) Fault management in process control: eye movements and action. Ergonomics 32:1319–1342

Morgan JF, Hancock PA (2008) Estimations in driving. In: Castro C (ed) Human factors of visual and cognitive performance in driving. Taylor and Francis/CRC, Boca Raton, pp 51–62

Neisser U (1976) Cognition and reality: principles and implications of cognitive psychology. W. H. Freeman/Times Books/Henry Holt & Co, New York

O'Brien KS, O'Hare D (2007) Situational awareness ability and cognitive skills training in a complex real-world task. Ergonomics 50:1064–1091

Parasuraman R, Sheridan TB, Wickens CD (2008) Situation awareness, mental workload, and trust in automation: viable, empirically supported cognitive engineering constructs. J Cogn Eng Decis Mak 2:140–160

Pew RW (1994) Situation awareness: the buzzword of the 90s. Gateway 5:1–4

Pierce RS, Strybel TZ, Vu KPL (2008) Comparing situation awareness measurement techniques in a low fidelity air traffic control simulation. Proceedings of the 26th international congress of the aeronautical sciences (ICAS). Anchorage, AS

Prince C, Ellis E, Brannick MT, Salas E (2007) Measurement of team situation awareness in low experience level aviators. Int J Aviat Psychol 17:41–57

Riener A, Ferscha A, Matscheko M (2008) Intelligent vehicle handling: steering and body postures while cornering. In: Brinkschule U, Ungerer T, Hochberger C, Spallek RG (eds) Architecture of computing systems–ARCS 2008. Springer, Berlin, pp 68–81

Robbins P, Aydede M (eds) (2009) The Cambridge handbook of situated cognition. Cambridge University Press, Cambridge

Rushton JP, Brainerd CJ, Pressley M (1983) Behavioral development and construct validity: the principle of aggregation. Psychol Bull 94:18–38

Salmon P, Stanton N, Walker G, Green D (2006) Situation awareness measurement: a review of applicability for C4i environments. Appl Ergon 37:225–238

Salmon PM, Stanton NA, Walker GH, Baber C, Jenkins DP, McMaster R, Young MS (2008) What really is going on? Review of situation awareness models for individuals and teams. Theor Issues Ergon Sci 9:297–323

Salmon PM, Stanton NA, Walker GH, Jenkins D, Ladva D, Rafferty L, Young M (2009) Measuring situation awareness in complex systems: comparison of measures study. Int J Ind Ergon 39:490–500

Sarter NB, Woods DD (1995) How in the world did we ever get into that mode? Mode error and awareness in supervisory control. Hum Factors 37:5–19

Sechrest L (1963) Incremental validity: a recommendation. Educ Psychol Meas 23:153–158

Senders JW (1964) The human operator as a monitor and controller of multidegree of freedom systems. IEEE Trans Hum Factors Electron 1:2–5

Senders JW (1983) Visual sampling processes (Doctoral dissertation). Katholieke Hogeschool Tilburg

Senders JW (2016) History of human factors [video file]. https://www.youtube.com/watch?v=QZH97Xy6NA8&t=2801s. Accessed 1 Feb 2018

Sheridan TB (1980) Computer control and human alienation. Technol Rev 83:60–73

Shinar D, Meir M, Ben-Shoham I (1998) How automatic is manual gear shifting? Hum Factors 40:647–654

Sivak M (1996) The information that drivers use: is it indeed 90% visual? Perception 25:1081–1089

Smith K, Hancock PA (1995) Situation awareness is adaptive, externally directed consciousness. Hum Factors 37:137–148

Sohn YW, Doane SM (2004) Memory processes of flight situation awareness: Interactive roles of working memory capacity, long-term working memory, and expertise. Hum Factors 46:461–475

Sparrow B, Liu J, Wegner DM (2011) Google effects on memory: cognitive consequences of having information at our fingertips. Science 333:776–778

Stanton NA, Salmon PM, Walker GH (2015) Let the reader decide: a paradigm shift for situation awareness in socio-technical systems. J Cogn Eng Decis Mak 9:44–50

Stanton NA, Salmon PM, Walker GH, Salas E, Hancock PA (2017) State-of-science: situation awareness in individuals, teams and systems. Ergonomics 60:449–466

Strybel TZ, Vu KPL, Kraft J, Minakata K (2008) Assessing the situation awareness of pilots engaged in self spacing. Proc Hum Factors Ergon Soc Ann Meet 52:11–15

Uhrig J, Cordts M, Franke U, Brox T (2016) Pixel-level encoding and depth layering for instance-level semantic labeling. arXiv:1604.05096

Underwood G, Crundall D, Chapman P (2011) Driving simulator validation with hazard perception. Transp Res Part F 14:435–446

Underwood G, Ngai A, Underwood J (2013) Driving experience and situation awareness in hazard detection. Saf Sci 56:29–35

Van der Horst R (2004) Occlusion as a measure for visual workload: an overview of TNO occlusion research in car driving. Appl Ergon 35:189–196

Van der Maas HL, Dolan CV, Grasman RP, Wicherts JM, Huizenga HM, Raijmakers ME (2006) A dynamical model of general intelligence: the positive manifold of intelligence by mutualism. Psychol Rev 113:842–861

Vanderhaegen F, Carsten O (2017) Can dissonance engineering improve risk analysis of human-machine systems? Cogn Technol Work 19:1–12

Vlakveld WP (2011) Hazard anticipation of young novice drivers (Doctoral dissertation). University of Groningen. SWOV dissertation series. SWOV Institute for Road Safety Research, Leidschendam, The Netherlands

Wai J, Lubinski D, Benbow CP (2009) Spatial ability for STEM domains: aligning over 50 years of cumulative psychological knowledge solidifies its importance. J Educ Psychol 101:817–835

Walker GH, Stanton NA, Salmon P, Jenkins D (2009) How can we support the commander's involvement in the planning process? An exploratory study into remote and co-located command planning. Int J Ind Ergon 39:456–464

Warm JS, Parasuraman R, Matthews G (2008) Vigilance requires hard mental work and is stressful. Hum Factors 50:433–441

White CB, Caird JK (2010) The blind date: the effects of change blindness, passenger conversation and gender on looked-but-failed-to-see (LBFTS) errors. Accid Anal Prev 42:1822–1830

Wickens CD (2008) Situation awareness: review of Mica Endsley's 1995 articles on situation awareness theory and measurement. Hum Factors 50:397–403

Wickens CD (2015) Situation awareness its applications value and its fuzzy dichotomies. J Cogn Eng Decis Mak 9:90–94

Wickens C, McCarley J, Thomas L (2003) Attention-situation awareness (A-SA) model. In: Foyle DC, Goodman A, Hooey BL (eds) Proceedings of the 2003 conference on human performance modeling of approach and landing with augmented displays (NASA/CP-2003-212267). NASA, Moffett Field, pp 189–225

Wickens CD, McCarley JS, Alexander AL, Thomas LC, Ambinder M, Zheng S (2008) Attention-situation awareness (A-SA) model of pilot error. In: Foyle DC, Hooey BL (eds) Human performance modeling in aviation. Taylor and Francis/CRC, Boca Raton, pp 213–242

World Health Organization (2015) WHO global status report on road safety 2015. World Health Organization, Geneva

Yarbus AJ (1967) Eye movements and vision. Plenum Press, New York