An exploratory analysis of eye tracking methods for measuring situation awareness in police operational settings

Øyunn Sofie Fencl

MAPSYK330, masterprogram i psykologi,

Studieretning: Psykologisk Vitenskap

Ved

UNIVERSITETET I BERGEN

DET PSYKOLOGISKE FAKULTET

VÅR 2019

Antall ord: 13746

Veileder: Asle Makoto Sandvik, Politihøgskolen i Stavern

Biveileder: Bjørn Sætrevik, Universitetet i Bergen

Abstract

This exploratory analysis has aimed to assess the applicability of eye tracking methods for measuring situation awareness in police operational settings. The data used in this analysis is based on a pilot study of eye tracking in a stress-enhancing police operational simulator scenario, conducted on final-year bachelor students at the Norwegian Police University College in Stavern. By using eye tracking data which was retrieved from two groups of participants exposed to different pre-scenario stress conditions and which measured gaze duration to one task-relevant and one task-irrelevant visual area of interest, four coarse hypotheses were developed and explored using group means comparisons, correlation and regression analyses. The analyses aimed to examine relationships between eye tracking data, self-reported situation awareness scores (measured using Situation Awareness Rating Scales (SARS)), and performance data (hit rate and reaction time), as well as the effect of stress on the two groups' performance and perception of situation awareness. Findings did not conclusively support our hypotheses, although the results did show a significant relationship between self-reported attention to danger cues and the task-relevant visual area of interest identified from the eye tracking data. Results also showed group differences, reflecting a number of possible effects of stress on situation awareness. These findings are discussed in light of current research and literature review, and limitations and suggestions for further research are addressed.

Keywords: exploratory, pilot, eye tracking, situation awareness, operational psychology, simulator, police, decision making

Sammendrag

Denne utforskende analysen har hatt som mål å undersøke og vurdere anvendelsen av blikksporing (eye tracking) som metode i måling av situasjonsbevissthet i politiets operative settinger. Analysen baserte seg på data fra en pilotstudie der blikksporing ble målt i et stressbelastet operativt simulatorscenario, med tredjeårs bachelorstudenter ved Politihøgskolen i Stavern som frivillige deltakere. Blikksporingsdata fra to grupper utsatt for ulike stressbetingelser ble anvendt for å måle varigheten av blikkfiksering på et oppdragsrelevant og et ikke-oppdragsrelevant visuelt fikseringsområde. Ut ifra disse dataene ble fire grove hypoteser utviklet og analysert ved bruk av gruppegjennomsnitt samt korrelasjons- og regresjonsanalyser. Analysene hadde som mål å utforske forholdet mellom blikksporingsdata, selvrapportert situasjonsbevissthet (målt i Situation Awareness Rating Scales (SARS)) og prestasjonsdata (målt i antall treff og reaksjonstid), samt å vurdere effekten av stress på de to gruppenes prestasjoner og opplevd situasjonsbevissthet. Funnene støttet ikke hypotesene, men resultatene viste signifikante forhold mellom selvrapportert oppmerksomhet på faresignaler og det oppdragsrelevante visuelle fikseringsområdet hentet ut fra blikksporingsdataene. Resultatene pekte også på funn i gruppeforskjeller, noe som indikerer at stress har en mulig effekt på situasjonsbevissthet. Funnene diskuteres i lys av nyere forskning og litteraturgjennomgang, der svakheter ved pilotstudien og analysen samt forslag til videre forskning er presentert og diskutert.

Forord

Jeg vil takke min biveileder Bjørn Sætrevik for god oppfølging og innspill til analyse og skriving av oppgaven, samt for å ha inkludert meg i prosessen med å analysere eye tracking dataene. Takk også til tredjeårsstudentene ved Politihøgskolen i Stavern for deltakelse i studien. Takk til mine foreldre for god støtte og gjennomlesing av oppgaven.

En spesiell takk rettes til min veileder Asle Makoto Sandvik for at jeg fikk delta i dette prosjektet. Din genuine interesse for fagfeltet har vært inspirerende. Samtidig har du vist tålmodighet og forståelse gjennom skriveprosessen og kommet med grundige og raske tilbakemeldinger.

Table of Contents

As a part of their work, law enforcement officials must often make quick decisions in ambiguous, unclear and stressful situations. At times, their decisions lead to fatal consequences. A recent case in point is that of Swedish Eric Torell, a 20-year old, autistic man with Downs Syndrome who was erroneously shot and killed by three police officers. One of the factors that led to the incorrect decision to shoot Mr. Torell, was that the police officers misinterpreted the harmless toy gun he was carrying as being an actual weapon (Brustad, 2018).

Several factors may lead police officers into making errors in their decision making, including time pressure, operating in ambiguous and constantly changing environments, as well as misinterpretations of situations that arise (Johnsen, 2018). Reports show that lack of situation awareness (SA) is one of the major contributors to human error in critical situations (Johnsen, 2018). Although the concept of situation awareness continues to be debated in the literature, it has become a mainstay of the sciences of operational psychology and human factors and safety (van Winsen, Henriqson, Schuler & Dekker, 2015). Situation awareness has been extensively researched, particularly within operational domains such as the military (Graham, Endsley, Weeks & Strater, 2001; Endsley, Holder, Leibrecht, Garland, Wampler & Matthews, 2000), aviation (Muehlethaler & Knecht, 2016), maritime operations (Saus, Johnsen & Eid, 2010) and healthcare (Cooper, Browning, Ross, Sparkes, Williams, Munro, O'Meara, Black & Bogossian, 2014). The primary focus of many of these studies has been on learning outcomes and developing training programs to improve situation awareness. Despite the abundance of research, however, not only is a universal set of criteria as to what defines good situation awareness lacking, but there appears to be no consensus as to how SA can be specifically measured (see Flach, 1995). Following hypotheses that attention follows gaze fixation and vice-versa (Just & Carpenter, 1980; Deubel, 1996; Cooper et al., 2014) and an understanding that attention is critical to attaining and maintaining situation awareness, eye tracking has been proposed as potential means of SA measurement. Using data collected from a pilot study of eye tracking in a training simulator scenario conducted at the Stavern division of the Norwegian Police University College, this thesis seeks to investigate to what extent eye tracking methods can be used as a measure of situation awareness in police operational settings.

Situation awareness has a number of theoretical underpinnings. Although it is a topic of study spanning multiple research domains, within law enforcement it is approached primarily from an operational perspective. A short overview of operational psychology will therefore be provided below, followed by a summary of the primary research methods employed. The psychological processes involved in decision making are also critical to SA. The level of an individual's or group's SA is most clearly reflected in the kinds of decisions they make and the consequences of those decisions. As such, a review of research and theory on decision making relevant to SA will also be presented.

Operational Psychology

Operational psychology is a field within psychological research that seeks to investigate and account for the various basic psychological principles and processes which influence problem solving and cooperation in teams and organizations during the execution of demanding tasks in unsecured and high-pressure environments (Eid, Johnsen & Laberg, 2018).

One of the main factors that differentiates operational psychology from other areas of research is that it primarily examines how people behave and react in situations that are uncertain, complex, dynamic and critical. Therefore, operational psychological research specifically focuses on acquiring knowledge about the relevant psychological processes involved in individual appraisal of risk, safety, perception of danger and threats, and how decisions are carried out (Kobbeltvedt et al., 2002). According to Sommerfelt-Pettersen (2018), readiness and adequate training are key elements for success in operational contexts, as well as the ability to professionally manage the unwanted, unexpected and unclarified. Eid, Johnsen and Laberg (2018) define operational psychology as "the systematic knowledge about individual or contextual factors that affect human behavior in operational environments and operational situations where health, basic values or even life is at stake" (p. 15). This definition implies that both personal traits, abilities, knowledge, and training in combination with physical, technical and social factors in the environment can affect how people behave and react in operational situations.

Operational psychology is based on general psychological theories within areas such as personality psychology, cognitive psychology, social psychology and physiological aspects of psychology. Furthermore, the field also incorporates experiential knowledge from the military, police, emergency medical dispatcher and response services, fire and rescue services, as well as humanitarian response teams (Eid, Johnsen & Laberg, 2018).

Given the focus area and history of the field, participants in studies are often referred to as *operators* in operational psychology literature, a term which will be used interchangeably with subjects, participants etc. within this thesis.

Common research methods used in operatioal psychology. Due to the complex and dynamic nature of its object of study, operational psychological research focuses on examining operator behavior and decision making in the field or in experimental contexts which emulate the uncertainty and stress of an actual operative context. Some of the research methods used in operational psychology are interviews, surveys, experiments and personality assessment (Eid, Johnsen & Laberg, 2018).

Interviewing methods allow researchers to study individuals' experiences and perceptions of real, critical and rare events that are challenging to study using experimental methods. Interviews allow accessing valuable insights into how individuals think, plan and react in unexpected and unclear situations, and elicit the narrative behind individual experience (Kvale, 1996). However, the usefulness of interview methods on a wider scale is limited. For example, it is not possible to generalize results from interview-based data to a wider population, since the data is based on the subjective experiences of individuals. Additionally, interview methods are fraught with a number of pitfalls, including interviewer bias and challenges associated with alterations of reported events or experiences due to memory issues, a desire to be a "good respondent" and please the interviewer, or simply responding out of self-interest or group pressure (Weiss, 1994; Eid, Johnsen & Laberg 2018).

The survey method is often used to map out features of an entire organization. Within operational psychology, the surveys are usually utilized when researchers want to examine attitudes, values, and norms within an operational organization or department. Survey methods can provide valuable information which can be used to develop action plans to improve working environments and develop trust in operational teams.

Experimental studies are the most accurate means to examine causal relationships and are therefore used extensively in operational psychological research. The development of simulators, combined with advances in monitoring and tracking technology have made it easier to maintain the demand for control in experimental environments while at the same time increasing the impression of reality in experiments (Eid, Johnsen & Laberg, 2018).

Personality assessments are often used in the recruitment process of personnel to operational teams and usually consist of a combination of interviews, observation, and testing. Candidates may, for example, be placed in a pre-constructed experimental situation, asked to solve various tasks while being observed by the experimental team, and followed up by post-trial interviews. Personality assessments have also been employed to identify specific traits and attributes and how these can affect operational outcomes as well as determine team member suitability and group composition in specific operational settings (e.g. Kjærgaard, Leon, Venables & Fink, 2013; Wright & Fallacaro, 2011; Chidester, Kanki, Foushee & Dickinson, 1990).

Decision Making

Central to all operational tasks is the ability to make decisions. Although a universal among all humans, the decisions made in operational professions are often high stake, and errors in judgement can have devastating consequences. Additionally, operational decisions must often be made in very short time and in chaotic environments that leave little or no space to consider options (Johnsen, 2018).

Decision making can be defined as selecting “choices among options” (Klein, 2008). Many theories on decision making have emerged through the years, a number of which have their conceptual basis in a rational choice paradigm, where decision making is understood as a weighing of options and generation of “decision sets” (Klein, 2008). The literature differentiates between *analytical decision making strategies* and *intuitive decision making strategies*. Analytical decision making strategies entail cognitively demanding processes such as assessments, calculations, and critical thinking. In contrast to intuitive strategies, analytical decision making strategies are time-consuming and demand that decision makers have an overview of the various alternatives and the ability to predict different outcomes. Although analytical strategies have been considered the best possible approach to decision making, research has shown that only 10% of decisions made in operational environments are the results of analytical decision making. In the late 1980’s, researchers conducted extensive studies of how people in high-stakes environments, including jurors, navy commanders, airline pilots, nurses and others, made decisions in the situations and settings they engaged in. They found that operators used their prior experiences to make judgements and decisions – referred to as intuitive or naturalistic decision making (NDM) (Klein, 2008) – rather than taking an analytical approach to a much greater degree than previously assumed.

Recognition-Primed Decision Model. The Recognition-Primed Decision Model (RPD) (Klein, Calderwood & Clinton-Cirocco, 1986, as cited in Klein 2008) is a sub-theory of NDM that formulates how people make decisions by matching the situation at hand with mental patterns of previously experienced situations. Based on prior experiences, operators are able to recognize the central features of new situations. These features or patterns generate expectations, relevant descriptions, definitions of possible goals, and common actions (Johnsen, 2018, p. 258). Prior experiences can not only be formed by participation in concrete events but also through training. The RPD model holds that the matching of prior experiences to new situations is based on four features; *cues*, *expectancies*, *goals*, and *typical actions* (Klein & Calderwood, 1991, as cited in Lundberg, 2015). *Cues* can be defined as “aspects of the environment that can be used to recognize situations” (Lundberg, 2015, p. 5). *Expectancies* are defined as the ability to “generate explanations for events that have occurred, tying them together in a story” (Klein et al., 2005, p. 21, as cited in Lundberg, 2015). Expectancies include the observation of cues to evaluate the understanding of the accuracy of the situation. A decision maker’s *goals* (what to achieve) and *typical actions* (what to do) are also important in understanding situations. These features relate the future projection of the situation to the goals and actions of the decision maker (Lundberg, 2015).

The RPD model consists of a three-stage process where each stage is characterized by the operator’s degree of familiarity with the situation. *Stage 1* (simple comparison) refers to a familiar situation in which the operator readily recognizes its central elements and forms realistic expectations, definitions of possible goals, and a repertoire of common actions. At *Stage 2* (diagnosing the situation), the degree of uncertainty is greater, and the operator will often generate “if-then” questions, depending on his or her knowledge of the situation and the available courses of action. Here, the operator is aware of the possible choices but uncertain about important factors concerning the situation which may influence the outcome. Finally, *Stage 3* (evaluating course of action) refers to situations where the circumstances are clear, but the operator has little or no knowledge of what actions to take. Since these are often entirely new situations to the operator, he or she is forced to engage in mental simulations of the consequences of different patterns of action prior to choosing an option (Klein, 1993, p. 144; Johnsen, 2018). Mental simulation is a process where the operator imagines how a situation will play out within the constraints and context of the situation (Klein, 2008) and evaluates available options serially, *satisficing* (Simon, 1956, p. 129) for the first solution that works. The selected option may not necessarily be the best option, but one which is good enough (Klein, 1993).

Errors and pitfalls in decision making. Errors due to lack of experience may lead an operator to misinterpret the situation or to opt for a wrong course of action. Lack of information may make it difficult for an operator to recognize features of the situation and therefore lead to decision errors (Johnsen, 2018). Incomplete simulation may lead to action errors in cases where the operator cannot or does not have the possibility to predict the outcome of an action, usually while a situation is still developing. This error may arise as the result of time pressure or excessive workload (Orasanu, Martin & Davidson, 1998).

Time pressure can also lead to "the information trap", a specific kind of decision error that stems from an operator's desire to gain more information about a situation, and how this desire can come into conflict with the need to arrive at a decision. The more time that passes before a decision is made and the more information an operator receives, the number of available options potentially increases. After a period of time, however, the number of available options will begin to decrease due to an escalation of the situation or previous options no longer being viable. In critical situations, therefore, adequate decisions need to be made during a "window of opportunity" (Johnsen, 2018) when the information available is sufficient to act, if not necessarily complete.

Situation Awareness

The concept of situation awareness (hereafter abbreviated as SA) constitutes the foundation on which decisions are made. Endsley (1995) defines SA 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 near future"* (Endsley, 1987b, 1988b in Endsley, 1995, p. 36). This is also the most widely accepted definition to date. The definition implies that SA consist of three different mental processes; *perception, comprehension* and *projection* (Salmon, Stanton, Walker, Baber, 2006, as cited in Johnsen, 2018). The theory does not address all of an individual's knowledge, but rather the knowledge about the status of different components that make up a dynamic environment (Endsley, 1995). It should be noted that elements, environments, time, and space are highly variable across operational situations and professions. A law enforcement worker and an aircraft pilot may, for example, perceive different elements in the same environment, ascribe their importance differently, or have different experiences of time and space.

Endsley's theory of situation awareness in dynamic systems. According to Endsley (1995), it is necessary to distinguish SA as a state of knowledge from the processes involved in obtaining SA, referred to as *situation assessment*. The three different mental processes involved in obtaining SA can be divided into three different levels. *Level 1* (perception) refers to the perception of elements in the environment relevant for the situation. For example, a police officer needs to know the location and appearance of suspects, if they are armed and if there are casualties or injured people. At *Level 2* (comprehension) an operator needs an understanding of the elements and how they are related to and affect the operational goals. Level 2 SA incorporates knowledge from Level 1 to form a holistic picture of the environment to understand the importance of the elements in the environment. According to

Endsley (1995), the difference between a novice and an expert decision maker becomes evident in Level 2 SA. While a novice decision maker may completely perceive the elements of a particular situation, he or she may not be able to understand the significance or status of these elements and how they may affect operational goals. The highest level of SA, *Level 3* (projection), refers to the ability to predict the future actions of the elements within the specific situation. The operator must therefore try to predict possible outcomes of the situation and their consequences. The ability to project the location and actions of the elements also provides the operator with the time and knowledge necessary to make the best decision possible to achieve the desired goals (Endsley, 1995). Because operational environments are dynamic, this also means that the situation can change at any time. Therefore, operators must be able to change their SA according to the situation by updating their perception and understanding of the elements (Johnsen, 2018).

Endsley (1995) argues that the human properties of short-term sensory memory, working memory, long-term memory, and perception in combination both underlie and affect SA. Short-term sensory memory, also called “preattentive processing” (Endsley, 1995), is the process of sensing elements in the environment such as movement, color, shapes, and spatial proximity (Neisser, 1976). The elements that are most salient will provide cues for processing using attention ultimately leading to perception (Endsley, 1995). Humans have limited attentional capacity, which may pose problems for operators in complex and dynamic operational environments. Short-term sensory memory therefore has implications for the development of Level 1 SA. According to research, operators that work in complex environments often avoid the problem of limited attention capacity by deciding themselves where to direct their attention (Braune & Trollip, 1982, as cited in Endsley, 1995). For example, a police officer involved in a robbery case may decide to direct his or her attention to footprints outside the building only and, for the time being, ignore other information. So, to a certain extent, individuals can determine which elements in the environment that become a part of their Level 1 SA based on their operational goals as well as long-term memory and working memory (Endsley, 1995). Preconceptions, experience, and expectations also affect how an individual perceives the environment (Jones, 1977, as cited in Endsley, 1995). An experienced operator will have more accurate expectations about how a specific situation will develop compared to a novice. If the situation plays out according to expectations, the experienced operator will therefore be able to process the information faster. However, the operator will be more prone to make errors if the situation does not play out according to expectations. After information has been perceived, it is processed in working memory. Both

Level 2 and Level 3 SA processes, as well as decision making and subsequent actions occur in working memory. The capacity of working memory is limited, especially when dealing with new situations (Endsley, 1995). The processes involved in long-term memory, however, such as the creation of *mental models*, do not have these capacity limitations. Mental models are defined by Rouse and Morris (1985, p. 7) as “the mechanisms whereby humans are able to generate descriptions of system functioning and observed system states, and predictions of future states”. Mental models are fundamental to the processes described in the Recognition-Primed Decision Model (RPD) (see above) as well as to the ability to achieve higher levels of SA without overburdening working memory. In order to successfully achieve higher SA using mental models, an individual must be able to perceive and identify “critical cues” in the environment (Endsley, 1995).

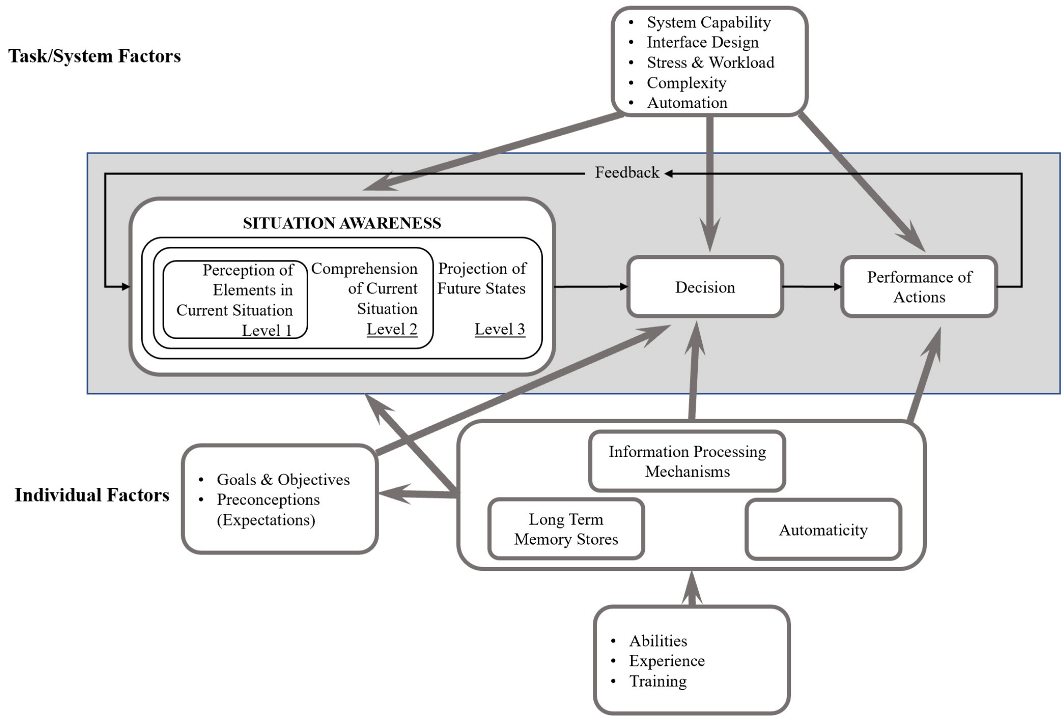


Figure 1. Illustration of Endsley’s (1995) theory of situation awareness in dynamic systems (redrawn by the author).

Lundberg’s (2015) holistic model of SA. In a recent proposal, Lundberg (2015) points out that much of earlier SA theorizing has tended to favor descriptions of SA from the perspective of a state (or level), a set of processes (the mechanisms involved in attaining and maintaining an SA state), or system (the context of human and technological relations in which SA is achieved and supported). He argues that these aspects of SA are interdependent and proposes a holistic framework which allows for analysis from all three perspectives simultaneously (p. 13).

As with earlier SA definitions, including Endsley's (1995), the concepts of mental models and schemata are central to Lundberg's framework. Mental models are necessary to achieve higher levels of SA and may be *declarative* (describing static elements of the environment of basic facts), *procedural* (describing relationships between actions, tasks, and goals), or *strategic* (describing the efficacy of actions and tasks in practice in attaining goals). In SA, mental models are central in determining a course of action or devising alternative procedures.

The concepts of mental model and schema are closely knit. Lundberg (2015) distinguishes between two classes of schemata. *Genotype* schemata inform what situation has occurred, how to make sense of it, and expectancies of what objects in the environment to explore further. They can be described as a "preparedness for exploration" (p. 3). *Phenotype* schemata inform the specifics of what is known about the situation which has occurred. They are specific manifestations of genotype schema and denote "implications of the explored environmental objects interpreted through the [genotype] schema" – that is, the meaning derived from the current status of objects in the environment in a particular situation. Lundberg also characterizes them as "particular instances of active exploration" (p. 3).

In the process of making sense of a situation, a schema can function as a frame – or script or story "that accounts for the data and guides the search for additional data" (Klein et al., 2005, p. 20, as cited in Lundberg, 2015). Framing is described by Klein, Wiggins, and Dominguez (2010) as the "process of fitting data into a frame, and fitting a frame around the data" (p. 308) and is integral to not only sense-making of objects in the environment but also impacts which objects are attended to in an environment and what importance they are ascribed. Situations are recognized through three key framing processes: re-framing, elaboration, and questioning. *Re-framing* refers to the comparison of different understandings (or frames) or creation of new understandings about a situation. *Elaborating frames* is the process of gathering additional information or details deemed as relevant to the current frame or rejecting information which is irrelevant. Lastly, *questioning frames* refers to the raising of doubt as to the correctness or appropriateness of currently generated and accepted frames due to inconsistencies, anomalies, or poor quality of information.

SA is achieved and maintained through a circular process modelled after the perceptual cycle (Neisser, 1976) in which schemata are constructed and reconstructed when making sense of a situation. This happens through the process of *framing*, where frames (genotype schemas) are created, elaborated, questioned, or evoked by cues, answering the general question: *What is this situation at large?* Frames also represent overarching goals and judgments as to what needs to be done in a particular situation (Lundberg, 2015, p.12). An awareness of the status of particular objects in the environment activates a (phenotype) schema, or model of the actual situation, which in turn sets expectancies and directs exploration of the environment for more relevant information. *Implications* or understandings which address specifics of an identified situation emerge from interpreting objects within frames and answer the specific question: *What about this situation?* The information collected from the environment (or emergence of/changes in the status of objects) within the existing frame may, in turn, lead to a re-framing or substitution of the original frame, thus closing the circle, setting new expectancies and initiating new rounds of sampling and interpretation.

This process takes place along an *event horizon* (a term borrowed from Hollnagel's (1993) contextual control model) of past, current, and future plans, goals, actions, and developments. The event horizon also represents an "awareness of uncertainty, timing and spatial orientation of events, as well as their relations to plans and system mechanics" (Lundberg, 2015, p.12). Awareness may be focused on present progress along the event horizon but may also have a diagnostic (past: what happened?) or prognostic (future: what can happen?) orientation. Particularly in uncertain situations, the extent of the event horizon may be limited, bringing about an awareness of alternative trajectories or forks in the event horizon.

The processes of framing and awareness of implications of objects within frames are supported by the system in which SA processes and states take place. The specific design or organization of a system (e.g. a police dispatching unit, air traffic control function, emergency response team, etc.) may inhibit or facilitate the flow of information, its distribution, and the identification of objects in the environment and their implications. Weaknesses in systems are exposed when, for example, the maintenance of SA in dynamic environments is overly dependent on human memory to make sense of steady flows of information. Due to the limitations of short-term memory, available information may either be too massive or dynamic in order to be manageable (Lundberg, 2015, p. 10). Although such limitations can be mitigated through teamwork and task delegation, collective SA (shared SA or team SA) is dependent on the correct distribution and compatibility of SA within the system (Lundberg, 2015, p. 11).

In short, Lundberg's model synthesizes previous models of SA with research findings over several decades and constructs a fine-grained framework for theorizing SA. It accounts for the close interdependence of SA states, processes, and systems; the integration of past and future (imagined) events in the present situation; and, in its distinction between framing and implications (elaboration of frames), details the cyclical processes involved in attaining and maintaining SA in dynamic environments.

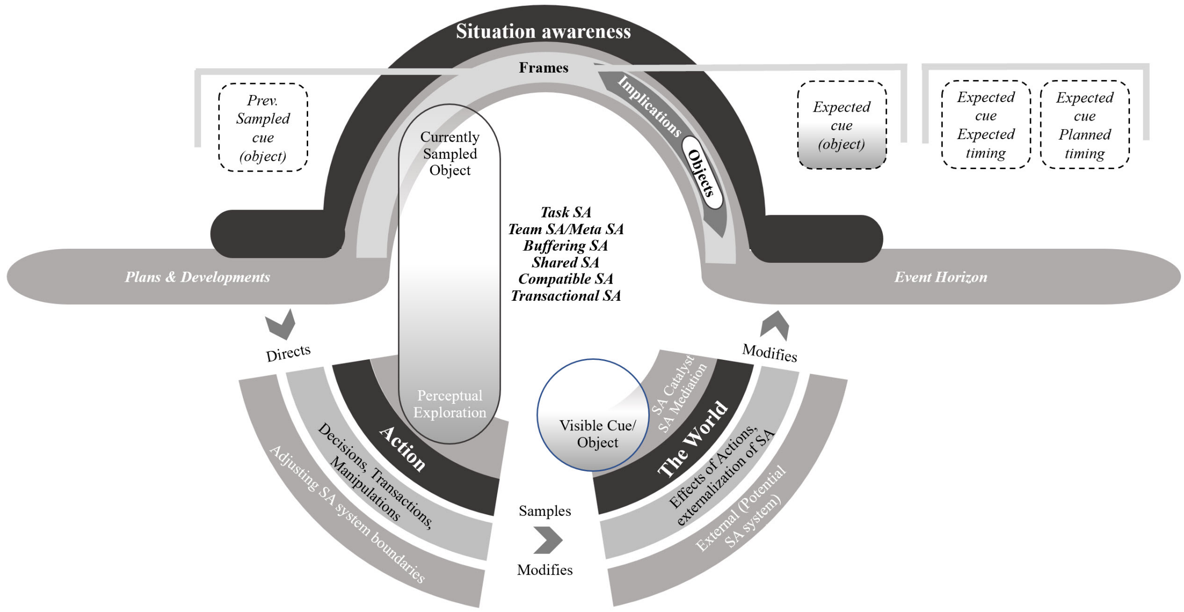


Figure 2. Illustration of Lundberg's (2015) holistic model of situation awareness (redrawn by the author).

Situation assessment. The Norwegian Police uses the concept *situation assessment*, which is closely related to SA but differs slightly from the definition of SA used within psychological research. *Situation assessment* is defined as "a review of the actual situation, an ongoing process from the commencement to the completion of the task" (Edvindsen, 2012, p. 138). According to Edvindsen (2012), several factors should be taken into consideration when conducting a situation assessment: the threat, one's own resources, time, assessment of the area and objects, weather conditions, lighting and driving conditions, and police communication and leadership.

Assessment of threat includes evaluating whether there has been a threat or if weapons have been used. Assessment of threat is appropriated by observation, intelligence, searching in the police's registers, and from other sources of information (Edvindsen, 2012). Furthermore, the culprit's actions, his or her ability to execute new actions, and his or her physical and mental state are important in the assessment of threats.

As part of the assessment process, the internal resources available to the police must be accounted for, including the number of officers present, what gear they have at their disposal, and whether the personnel is qualified for the task. Depending on the mission, reinforcement must be mobilized if some of these resources are lacking (Edvindsen, 2012). Critical situations often arise suddenly, and actions must be taken within a short period of time. However, police officers often have the opportunity to obtain more time by isolating, observing, and evacuating the area in order to plan and execute the mission. Furthermore, by definition, situation assessment also entails assessment of the area and object, the context, as well as police communication and leadership (Edvindsen, 2012). Situation assessment targets planning and assessing operations in advance of their implementation to a greater degree than psychological approaches to SA. In addition, it has a more practical orientation, identifying specific and detailed factors of a situation that needs to be considered before execution.

Stress and SA. Stress is viewed as a major challenge to SA. Although lower levels of stress may be beneficial to attaining SA due to increased attention to elements in a situation (Endsley, 1999, p. 265), the negative consequences of higher levels can be extreme. This is because stress, together with the increases in autonomic functioning which it causes, can tax an operator's limited attentional capacity (Hockey, 1986, as cited in Endsley, 1999), making it more difficult to achieve SA and more likely to decrease good decision making (see Sneddon, Mearns & Flin, 2013; Zhang, Jin, Garner, Mosaly & Kaber, 2009).

Endsley (1995, 1999) makes a distinction between two types of stress factors: *physical stressors*, including fatigue, lighting, weather, temperature, noise, vibration and boredom and *psychosocial stressors*, such as fear, anxiety, awareness of the gravity of a situation, prestige, danger of job loss, time pressure, mental load, and threats to self-esteem. Stress can affect SA in several ways. First, stress can decrease the capacity of available working memory and recall (Endsley, 1995, p. 53). Stress can also reduce the ability to take in information (Endsley, 1999, p. 265). On the cognitive level, this is demonstrated by the observation that operators in critical situations tend to retrieve information from probable or dominant sources at the expense of other peripheral sources which could provide valuable or essential information (Bacon, 1974; Weltman, Smith & Egstrom, 1971). Additionally, there is some evidence that physical stress can degrade the perception and comprehension of information (Perry, Sheik-Nainar, Segall, Ma & Kaber, 2008). Third, stress can cause a narrowing of the field of attention (Sneddon et al., 2013; Endsley, 1995, 1999).

The effect of stress on attention can be explained theoretically in terms of selective attention, also known as "tunnel vision". Selective attention can be defined as "the process of selecting and focusing our mental capacity on understanding and observing certain external stimuli"

(Eid, 2018, p. 95). The process of selective attention is important in operational settings because the ability to select and focus on relevant stimuli to the exclusion of irrelevant stimuli within dynamic and stressful environments may help the operator in developing an accurate SA, and arrive at a decision faster. However, there is an ongoing debate in the literature on how stress affects selective attention (Chajut & Algom, 2003). Several approaches to this topic have been proposed, among these are the Attention Approach, the Capacity-Resource Theory and the Thought Suppression Approach.

The Attention Approach derives from Easterbrook's (1959) influential work on how stress diminishes the number of cues used in performing a task. This approach's core assumption is that stress reduces the attentional resources available to an individual, thereby directing the remaining resources to process the task-relevant attributes of the situation. This will in turn result in improved selectivity in responding (Chajut & Algom, 2003).

The Capacity-Resource Theory also assumes that stress narrows attention, although in a directional way. This means that the narrowed attention is drawn to the more accessible, automatic dimensions of attention which are more easily activated in the absence of conscious attention (Chajut & Algom, 2003, p. 232). In other words, an operator under stress can risk narrowing his attention on irrelevant cues, thereby failing to develop selective attention to the task-relevant dimensions.

The third approach, the Thought Suppression Approach, holds that attention is a voluntary, conscious process directed towards target relevant stimuli. In contrast to other approaches, it includes a second, unconscious and automatic process (Wegner et al., 1994, as cited in Chajut & Algom, 2003), which suppresses attention to stimuli irrelevant to the task being performed. This does not mean, however, that there is no awareness of irrelevant stimuli, since ascertaining stimulus relevance presupposes some level of awareness. Under stress, the automatic "search for to-be-suppressed information" (Chajut & Algom, 2003, p. 233) ironically increases a person's sensitivity to the very information which he or she aims to ignore. Because the act of focusing on relevant stimuli depletes attentional resources, the unconscious suppression of irrelevant information may increasingly take precedence and come to the forefront as attentional resources weaken (Wegner et al., 1994, as cited in Chajut & Algom, 2003). The distinction between the two latter approaches is that the Capacity-Resource Approach emphasizes the stimuli as being processed in an automatic way, while the Thought Suppression Approach holds that it is the process itself that is automatic (Chajut & Algom, 2003).

Common Approaches to Studying SA

Much of the research on SA has been conducted within the aviation, military, health and sports domains. One reason for this is the continuously increasing use of advanced technology in aircraft, military, health and sports systems, and because a significant percentage of accidents within these domains are caused by errors in SA and human factors (Graham et al, 2001). For example, Endsley, Holder, Leibrecht, Garland, Wampler and Matthews (2000) developed a model of SA which included variables specific for military operations and tasks (e.g. tactics, fatigue and time pressure), which has been valuable for training on enhancing SA among soldiers (Graham et al., 2001). In sports, SA has been studied from the perspective of the official (e.g. referee/judge) who must be able to make high-stakes decisions in pressured situations (Neville & Simon, 2016). However, because of the wide application of SA as a construct in a number of different scientific and non-scientific disciplines, it has become increasingly difficult to define and understand (van Winsen et al., 2015).

Eye Tracking

Much of human behavior requires visual information and attention. By studying the eye movements of people performing actions or engaging in tasks, researchers can gain valuable insights into the workings of visual perception and its interactions with behavior, attention and cognition.

Initially developed within aviation and marketing, with the purpose of identifying a person's scan path, gaze fixation, and area of interest (Cooper et al., 2014), modern eye tracking methods have become increasingly popular in a wide variety of research fields (Horsley, 2014). Most people are not aware of where they direct their gaze. As research shows, the foveae (pupils) shift location almost every three seconds (Tatler, Kirtley, MacDonald, Mitchell & Savage, 2014). However, since attention follows what our eyes fixate on and vice-versa, eye tracking can be used as means of objectively capturing the frequency of consecutive fixations (referred to as *gazes*, Just & Carpenter, 1980, p. 329). According to Tatler, et al. (2014), perception and action are bidirectionally linked; the perception process provides us with information to perform actions, while the actions we perform influence the environment and therefore also our perception of it. Eye tracking methods in research have the advantage of being able to measure the perception of an individual without influencing the environment and can be employed in both laboratory and real-life settings (Tatler et al., 2014).

Eye tracking and SA measurement. Salmon, Stanton, Walker, Jenkins, Ladva Rafferty and Young (2009) have reported that there exist over twenty different approaches to measuring SA. These include freeze probe methods (where queries are administered during "time-outs" in the task or scenario), performance/outcome measures (where performance is tested on imbedded or external task), rating techniques (completed by participant and/or observer), and self-report questionnaires (completed in real-time or post-trial, e.g. SART, Taylor, 1990; SARS, Waag & Houck, 1994). According to Endsley's (1995) theory of Situation Awareness, the ability to perceive the elements in the environment, comprehend how these are related and how they affect goals, and to project and predict their actions are prerequisites to attaining and maintaining SA. Since attention is required for identification and comprehension of key elements in the environment, eye tracking data can provide detailed information about where a person's attention and focus is directed, and therefore which elements a person perceives in a given environment. Analysis of this data can provide not only an objective measure of SA prerequisite fulfilment, but also an indication of SA level attained.

Objectives

This thesis is an exploratory analysis which aims to assess the extent to which eye tracking methods can be used as a measure of situation awareness in police operational settings. As a basis for analysis, results from a pilot study carried out at the Norwegian Police University College in Stavern (PHS) will be used following a set of coarsely defined hypotheses. By complementing the objective eye tracking data with subjective data elicited from participants using the SARS self-report questionnaire (Waag & Houck, 1994), an analysis of how well eye tracking data compares to participant experiences of their own levels of SA will be provided. Additionally, the impact of stress as part of the pilot study will be analysed and discussed. In order to include all the possible variables which can indicate the applicability of eye tracking methods as a measure of SA, the hypothesis in this thesis will focus on two areas of interest (AoIs) retrieved from the eye tracking data from the PHS pilot study: "pillar" and "fallen hostage", with self-reported situation awareness and reaction time being predictors of the two. This generates four coarse hypotheses distributed into two executive pairs, presented below (p. 27).

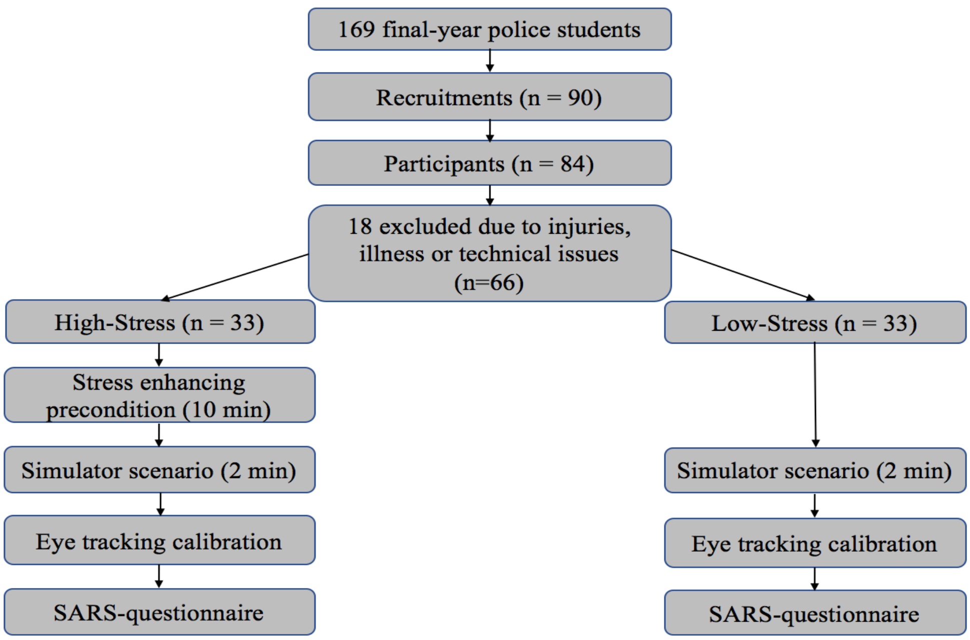
Method

The pilot study on which this thesis is based includes both objective data retrieved from eye tracking recordings (" visit pillar" and "visit fallen hostage"), performance data (the reaction time of the shots fired by the participants, measured in milliseconds) and subjective data collected from the SARS self-report questionnaire (see Table 1).

Participants and Participant Flow

The participants in the PHS study were recruited from among final-year police students enrolled in the bachelor's program at the Stavern division of the Norwegian Police University College (PHS). The first 90 students were recruited based on a convenience sample and the first 45 males and 45 females who volunteered to participate were included in the study. No incentives or monetary remuneration were offered. Participants were between the age of 22 to 33, with a mean of 24.25 years. Due to illnesses or injuries, six participants withdrew from the study before the data collection commenced (n = 84). During the experiment, twelve participants were eliminated from the scenario due to virtually being "shot" by the first assailant in the scenario (see below for more information). Since they could not provide any eye tracking data were, any subsequent results they provided (e.g. from the self-report questionnaire), was no longer relevant to this thesis and were removed from all analysis. Further, the scores from an additional six participants were excluded from the eye tracking data set due to technical issues. As part of the experiment, participants were randomly assigned into two groups. One group (High-Stress group) was subjected to a number of stress enhancing events prior to the main test in the simulator. The other group (Low-Stress group) was not exposed to stress enhancement prior to meeting at the locale where the simulation was held.

Figure 3. Participant flow through the pilot study.



Norwegian Police and Police Education

The aim of the official training of the Norwegian police is to provide a broad theoretical and practical foundation for police work. The central institution of education where this is provided is the Norwegian Police University College, which offers a three-year bachelor's degree program in police study, in-service training, post-graduate studies, including a master's program in police science. After completing a bachelor's degree, graduates qualify for employment as police officers in Norway. In general, the Norwegian police is unarmed while on service. If ordered to so do, however, officers can arm themselves with weapons accessible in their police cars or, in particularly urgent or critical cases, arm themselves without an order.

Training Simulator

The simulator used in the PHS study was a MILO® Range 4.8. Shooting Simulator produced by IES Interactive Training, which uses video and interactive technology to simulate realistic scenarios and environments for practicing use of firearms. The PHS study scenario was an ongoing, multi-perpetrator school shooting and included graphically violent scenes in which pupils were shot. Each participant completed the trial individually and the scenario lasted approximately for two minutes. As part of the scenario, three shooters appeared on the scene in succession, and the participant had to eliminate each assailant in order to successfully advance to the next stage. The scenario ended after the elimination of the third shooter or if the participant was shot by one of the assailants. This thesis focuses solely on the eye tracking data collected in conjunction with the second shooter who appears suddenly from behind a pillar in the interactive video. To ensure that all participants progressed to the second shooter (and were thus able to provide eye tracking data), the organizers of the study, if needed, remotely assisted the active participants without their knowledge on the takedown of the first shooter. The scenario was designed to be as realistic as possible, and the pupils who were portrayed in the interactive video as having been taken hostage or shot were fully visible to each participant. Between the first and second shooter scenes in the interactive video, one of the pupils gave a subtle, non-verbal cue to the participant, indicating that the shooter is behind the pillar. Out of a total of 84 participants, 71 were assisted in the takedown of the first shooter (36 in the High-Stress condition and 35 in the Low-Stress condition). There were several reasons why assistance was provided, one being the participants' shooting skills. Most of the participants had not yet undergone any extensive shooting training or practice, since these skills are usually taught towards the end of their education. Most of the training they had received prior to the study was scenario based, primarily focusing on the feasibility or legality of firearms use in service. There were also reasons of a more technical nature as to why assistance was provided; the shots fired were not always registered by the system, and the "hit zones" defined as part of the scenario were somewhat limited (e.g. a shot in the leg of one of the perpetrators would not register as a hit). The weapon was a Heckler & Koch P30L modified for use in the simulator. Heckler & Koch P30L is the standard firearm used in the Norwegian police.

Image 1. Screen shot from video recordings of the participants in the training simulator

Procedures

The study was approved by the Norwegian Centre for Research Data (NSD). The participants were informed of their rights to withdraw from the study at any time, and the informed consent statement was signed by all. Before the study commenced, the participants were randomly assigned, while matched on gender, into two different testing conditions; Low-Stress and High-Stress. The participants were not informed about which test condition they were assigned to and were only given a date, time and location for the simulator test. Immediately after the simulator test, a Situation Awareness Rating Scales (SARS) questionnaire was completed by all.

High-Stress condition. Prior to the main test in the simulator, the participants in the High-Stress group were exposed to a physical and mental stress enhancing condition, lasting for about 10 minutes. Following instructions, the participants packed their gear and prepared for a «normal» day on patrol. The participants were seated in the front passenger seat and were told that the driver of the patrol car (a silent partner and experienced police officer), would remain in the role of a driver and not leave the car or interact in any way. Participants communicated with dispatch on the radio and were ordered to respond as a one-person patrol.

Once they left the car, they would be on their own. While still in the car, dispatch ordered them to drive to a given address in response to a report of domestic disturbance in a third-floor apartment. Just before the participants reached the apartment, they received another urgent call from dispatch requiring them to abort the current mission and to immediately respond to reports of shots fired at a nearby high school (address of the building with the training simulator). As part of the response, participants were instructed to put on protective gear, arm themselves from the car, and head off on foot to the high school (located about 50m/165 ft from the apartment). Because of their close proximity, they were informed that they were the first of several patrols on the scene and had to act instantly to save lives. Once at the school gate, they could hear shots being fired. To open the gate, they were given a four-digit code over the radio. Once inside the school building, the scenario was briefly paused (about 30 seconds), while the participants were shown to the simulator training room. Their weapons were replaced by laser guns specifically designed for the simulator. They were further informed that the scenario would start back at the school entrance and were urged to act accordingly.

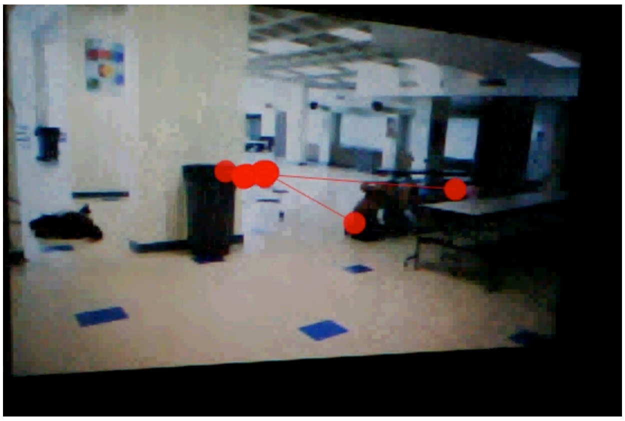
Low-Stress condition. The Low-Stress group participants met at the training simulator room (next to the simulator), where they were solely provided with oral instructions, and informed of several calls received regarding shots fired at Stavern High School. Similar to the High-Stress group, they were told that they were the first patrol on the scene and had to act instantly to save lives. They were already wearing protective gear and were provided with simulator guns. Once they entered the simulator room, the scenario started.

Measures and Data

Although there were in total three consecutive shooters in the simulator scenario, this exploratory analysis only focused on the eye tracking data collected for the second shooter in the scenario, as this was the most reliable and useful data for the purposes of this analysis.

Eye tracking and eye tracking glasses. Each of the participant's eye movements were recorded using the Tobii Eye Tracking Glasses, and analyzed using Tobii Studio. The data collected from the eye tracking tools together with the video recording of the scenario was analyzed by delimiting the spatial boundaries of two visual Areas of Interests (AoIs), the pillar and the fallen hostage, and extracting the cumulative duration of gaze directed within each AoI, measured in milliseconds (see Image 2). The glasses were calibrated immediately after the test scenario had ended according to standard procedures (see Tobii Technology, 2012, p. 25).

Image 2. Screen shot from video recordings



The pillar (left) and hostages (right), red dots are gaze fixations.

Situation Awareness Rating Scales (SARS). All participants completed a modified, Norwegian version of the Situational Awareness Rating Scale (SARS) (Waag & Houck, 1994, as cited in Endsley, 1996) questionnaire immediately after completing the simulator scenario. The questionnaire consisted of 17 questions about the scenario and was used to measure subjective SA on a six-point Likert's scale, where 1 = to a very little degree and 6 = to a great degree. The questions fell into six executive categories; *planning*, *management of gear*, *information*, *awareness of danger cues*, *awareness of the situation* and *learning experience* (see Table 1). Question number 15 was reversed in SPSS due to its negative formulation so that it corresponded with the rest of the SARS questionnaire.

Table 1. The SARS questionnaire divided into categories

Categories	Questions
Planning	1.“To what degree did the situation allow you to develop a plan for the mission?”
	2. “To what degree did you manage to stick to your plan?”
	3. “To what degree did you manage do make continuous changes to your plan?”
Management of gear	4. “To what degree did you manage to operate the available gear (weapons, armor etc.)?”
Information	5. “to what degree did you experience the quality of the initial information given to you by the assigner as being good?”
	6. “To what degree did you manage to make use of the information given to you?”

	7. "to what degree did you actively take initiative to gather relevant information to solve the mission?"
Awareness of danger cues	8. "To what degree did you detect cues to danger?"
	9. "To what degree did you manage to prioritize the order of situations that arose based on danger cues?"
	10. "To what degree did you manage to prioritize actions to critical signals?"
	11. "To what degree did you manage to initiate actions to critical signals?"
Awareness of the situation	12. "To what degree did you manage to create a picture of the situation?"
	13. "To what degree did you manage to discover changes in the situation?"
	14. "To what degree did the situation make you offensive (you acted prior to events)?"
	15. "To what degree did the situation make you defensive?"
Learning experience	16. "To what degree did you experience the scenario as realistic?"
	17. "To what degree did the scenario give you learning advantages?"

Hit rate. Hit rates were measured according to the number of hits each participant were able to fire at the shooter, where 0 equals no hits, 1 equals one hit and so on.

Reaction time. The reaction time was measured in milliseconds from the moment the shooter appeared from behind the pillar to the first shot was fired by the participant. In addition to the eye tracking data, reaction time will be an expression of the participants' objective SA, since reaction time is associated with where their attention is directed.

Hypotheses.

The first pair of hypotheses asserts that the higher the participants' self-reported SA scores (Hypothesis 1) and the shorter their reaction time (Hypothesis 2) is, the more participants' attention to pillar (as a potential assailant hiding place) will be. In contrast, the lower the participants' self-rated SA scores (Hypothesis 3) and the longer their reaction time (Hypothesis 4) is, the more their attention to fallen hostage (as a distraction cue) will be. Note that in formulating the above hypotheses, *attention to* pillar and *attention to* fallen hostage is used. Although the eye tracking data used in the analyses measured duration of gaze within the Areas of Interest *pillar* and *fallen hostage,* gaze duration reflects attention and ultimately awareness. Therefore, *attention to* was chosen to make the discussions below more salient as to the aims of this thesis. This analysis is exploratory, in that it not only seeks to discover if and to what extent eye tracking methods can measure SA as a whole, but also to discover and account for which SA factors can be measured by eye tracking technology. As such, this analysis will analyze which of the self-reported SARS questionnaire categories (collectively referred to as "SA scores") of "Planning", "Management of gear", "Information", "Awareness of danger cues", "Awareness of the situation" and "Learning experience", as well as the participants' reaction time can best predict the two AoI variables. Further, based on the assumption that stress can affect attention and therefore SA and the inclusion of stress as a grouping factor in the PHS scenario, the impact of stress on self-reported SA scores and eye tracking data will be explored.

Results

Table 2. This table shows descriptive statistics of the sample.

		N	***Male***	***Female***	
	Total sample	66	33	33	
	Low stress	33	16	17	
	High stress	33	17	16	
		Min	***Max***	***Mean***	***SD***
	Age (years)	22	33	24.25	2.15
	Age (male)	22	27	23.81	1.38
	Age (female)	22	33	24.69	2.65
Eye tracking data	Pillar (milliseconds)	0.20	5.28	2.18	1.26
	Fallen hostage (milliseconds)	0.03	3.21	0.68	0.61
Performance data	Hits	0	2	1.43	0.55
	Reaction time (milliseconds)	0.87	79.58	3.30	10.79
SARS	Situation awareness	9	24	16.07	3.12
	Planning	3	14	9.54	2.25
	Information	5	14	10.27	2.14
	Gear	1	6	4.09	1.06
	Danger cues	9	22	16.76	2.46
	Learning experience	4	12	9.98	1.83

In order to explore the effect of stress and whether there was a statistical difference between the High-Stress and the Low-Stress groups in their mean scores on the objective and self-rated SA scores, an independent-samples t-test was conducted (Pallant, 2016).

Table 3. This table shows an independent-samples t-test for differences between groups.

	Low Stress (*n* =42) Mean (*SD*)	High Stress (*n* = 42) Mean (*SD*)	*t (df)*	*p*	p^1
Pillar	2.35 (1.31)	1.98 (1.20)	-1.090 (55)	.279	1.000
Fallen hostage	0.41 (0.32)	0.93 (0.71)	3.066 (29.98)	.005	.045*
Hits	1.38 (0.55)	1.45 (0.55)	0.585 (69)	.560	-
Reaction time	1.38 (0.55)	1.45 (0.55)	-0.913 (49)	.366	-
Awareness of Situation	16.76 (2.84)	15.34 (3.27)	-2.133 (82)	.036	.288
Planning	9.46 (2.17)	9.63 (2.36)	0.341 (82)	.734	-
Information	10.64 (1.98)	9.85 (2.24)	-1.778 (83)	.079	.539
Gear	4.25 (1.09)	3.92 (1.02)	-1.425 (83)	.158	.790
Danger cues	17.23 (2.37)	16.28 (2.49)	-1.792 (83)	.077	.539
Learning experience	9.38 (1.96)	10.59 (1.48)	3.200 (82)	.002	.020*

p^1 Holm-Bonferroni adjusted sig. level   *$p^1 < .05$ **$p^1 < .01$ (two-tailed)

Independent-samples T-test.

Eleven independent-samples t-tests was conducted to compare the different scores for the Low-Stress and the High-Stress groups (see Table 3). A Holm-Bonferroni sequential correction (Holm, 1979) was used to control for Type-1 (false positive) errors in conducting multiple tests, since this correction method is regarded as more powerful than the Bonferroni method for testing multiple hypotheses. There was a statistically significant difference between the Low-Stress and the High-Stress groups in total visit duration to fallen hostage (t $(29.9) = 3.06$, $p' = .045$), indicating that the members of the High-Stress group paid longer attention to fallen hostage than the Low-Stress group. Further, self-reported learning experience (t $(82) = 3.20$, $p' = .020$) was also found as statistically different between the groups, implying that the Low-Stress group reported themselves as learning less from the experience than the High-Stress group. Interestingly, there was also a difference between the mean score for the Low-Stress and High-Stress groups in total visit duration to pillar (t $(55) = -1.09$, $p = 1.000$), although this was not statistically significant.

Correlation Analyses

The relationship between the different variables was investigated using Pearson product-moment correlation coefficient, for the total sample and for each group. Preliminary analyses were performed to ensure no violation of the assumptions of normality, linearity and homoscedacity (Pallant, 2016).

Table 4. This table shows an overview of Pearson's Product Moment correlation coefficients.

		Eye tracking data		**Performance data**		**SARS-questionnaire**				
		Pillar	Fallen hostage	Hits	Reaction time	Awareness of Situation	Planning	Information	Gear	Danger cues
Total sample	Pillar (milliseconds)									
	Fallen hostage (milliseconds)	-.135								
	Hits	-.113	.125							
	Reaction time (milliseconds)	.192	.010	-.371**						
	Awareness of Situation	.056	-.117	.208	.023					
	Planning	-.041	.124	.080	-.103	.262*				
	Information	.208	-.060	-.067	.210	.163	.144			
	Gear	.101	-.115	.307**	-.022	.298**	.216*	.208		
	Danger cues	.305*	-.161	.284*	-.111	.449**	.309*.	.203	.353**	
	Learning experience	-.056	.128	.142	-.191	.145	.175	.099	.203	.182
High-Stress	Pillar (milliseconds)									
	Fallen hostage (milliseconds)	-.232								
	Hits	.121	.199							
	Reaction time (milliseconds)	-.058	-.044	-.039						
	Awareness of Situation	.212	.122	.081	-.195					
	Planning	.097	.259	.032	.507*	.190				
	Information	-.104	.057	.158	-.182	.441**	.376*			
	Gear	.104	.050	.411*	.049	.147	.047	.346*		
	Danger cues	.422*	.110	.238	-.066	.411**	.332*	.426*	.363*	
	Learning experience	-.165	.169	.114	.321	.104	.204	.238	.109	-.126
Low-Stress	Pillar (milliseconds)									
	Fallen hostage (milliseconds)	.020								
	Hits	-.322	-.023							
	Reaction time (milliseconds)	.245	.130	-.490**						
	Awareness of Situation	-.159	-.093	.453**	-.031					
	Planning	-.144	-.146	.134	-.152	.384*				
	Information	.448*	.253	-.338	-.254	-.250	-.107			
	Gear	.026	.017	.267	-.094	.410**	.399**	.017		
	Danger cues	.133	-.381	.434*	-.259	.445**	.319*	-.120	.306*	
	Learning experience	.082	-.237	.149	-.193	.370*	.151	.120	.391*	.579**
	**p* <.05	***p* <.01								

Addressing the first hypothesis (*the higher (self-reported) SA score the participants report, the longer they will focus their attention on pillar*), the results of the correlation analysis show that only one of the SA variables (danger cues) was significantly associated with total visit duration to pillar (r = .305, p = .021). This indicate that the longer the participants focused on the pillar, the higher they scored on self-reported attention to danger cues. Looking at group differences, significant medium positive correlations between gaze duration to pillar and danger cues were found for the High-Stress group only (r = .422, p = .028). For the Low-Stress group, on the other hand, we found medium, positive correlations between self-reported scores on information and gaze duration to pillar to be statistically significant (r = .448, p = .013), with higher scores on the "information" self-report SARS questions associated with longer fixation on pillar.

For our second hypothesis (*the shorter the participants' reaction time is, the longer they focus their attention on pillar*), there were no statistically significant findings in the correlation analyses neither for the total sample (*r* = 192, *p* = .192), nor for the two groups (High-Stress: *r* = -.058, *p* = .830, Low Stress: *r* = 245, *p* = .248).

Concentrating on our third hypothesis (*the lower (self-reported) SA score the participants report, the longer the participants focus their attention on fallen hostage*), the analysis showed no statistically significant correlations between the SARS scores variables and gaze duration to fallen hostage. However, a closer look at the difference between the correlation coefficients for the High-Stress and the Low-Stress groups uncovers an interesting pattern (see Table 4), indicating that SARS scores correlated more negatively with fallen hostage for the Low-Stress group than for the High-Stress group. Additionally, a non-significant medium, negative correlation between danger cues and gaze duration to the fallen hostage for the Low-Stress group (*r* = -.381, *p* = .091), suggests that higher self-rated awareness of danger cues is associated with shorter duration visit to the fallen hostage.

We found no significant correlations for our fourth hypothesis *(the longer the participants' reaction time is, the longer they will focus their attention on fallen hostage)* for the total sample (*r* = .010, *p* = .956) or either of the groups (High-Stress: *r* = -.044, *p* = .887, Low-Stress: *r* = .130, *p* = .597).

Findings of additional interest include statistically significant correlations between categories of the SARS questionnaire awareness of danger cues and planning for both groups (High-Stress: *r* = .332, *p* =.034, Low-Stress: .319, *p* = .037) and awareness of danger cues and awareness of the situation for both groups (High-Stress: *r* = .411, *p* = .008, Low-Stress: *r* = .445, *p* = .003). In addition, awareness of danger cues and hits was found to be significantly associated for the Low-Stress group only (*r* =.434, *p* = .010).

Regression Analysis

Based on the results of the correlation analyses above, addressing out first and third hypotheses (*the higher (self-reported) SA score the participants report, the longer they will focus their attention on pillar* and *the lower (self-reported) SA score the participants report, the longer the participants focus their attention on fallen hostage,* correspondingly), we wanted to explore the ability of three of the self-reported SA variables that had the strongest correlations with gaze duration on pillar (planning, information and danger cues). Standard linear regression analyses were conducted (Pallant, 2016) in order to predict gaze duration to pillar and fallen hostage and determine which of these variables was the best predictor of gaze duration to pillar and fallen hostage. Therefore, standard linear regression analyses were conducted (Pallant, 2016). Likewise, for our second (*the shorter the participants' reaction* *time is, the longer they focus their attention on pillar*) and fourth (*he longer the participants' reaction time is, the longer they focus their attention on fallen hostage*) hypotheses, we wanted to determine the predictive ability of reaction time on gaze duration to pillar and fallen hostage. Linear regression analyses were also conducted for each group to investigate any differences between the groups in predictive power of the predictor variables due to the influence of stress.

Table 5. Summary of standard linear regression analyses.

Criterion		Predictors	*B*	*SE B*	β	*p*
	Total					
Pillar	R2 = .139, *p* = .045	Constant	-.722	1.284		.576
		Planning	-.092	.076	-.164	.227
		Information	.098	.077	.166	.210
		Danger cues	.166	.070	.322	.022*
	High-Stress					
	R2 = .278, *p* = .054	Constant	-.758	1.504		.619
		Planning	.024	.100	.046	816
		Information	-.193	.109	-.360	.091
		Danger cues	.272	.097	.560	.010*
	Low-Stress					
	R2 = .262, *p* = .045	Constant	2.225	2.219		.325
		Planning	-.104	.108	-.172	.344
		Information	.304	.133	.459	.012**
		Danger cues	.134	.099	.243	.186
	Total					
	R2 = .061, *p* = .488	Constant	1.203	.770		.127
		Planning	.054	.045	.196	.245
		Information	-.013	.046	-.045	.781
		Danger cues	-.053	.042	-.213	.213
	High-Stress					
Fallen Hostage	R2 = .071, *p* = .716	Constant	.132	1.143		.909
		Planning	.081	.076	.266	.302
		Information	-.020	.083	-.064	.809
		Danger cues	.014	.074	.049	.851
	Low-Stress					
	R2 = .189, *p* = .327	Constant	.898	.735		.239
		Planning	-.002	.036	-.010	.966
		Information	.035	.037	.210	.369
		Danger cues	-.048	.033	-.353	.159
	Total					
	R2 = .037, *p* = .235	Constant	2.107	.209		.000*
		Reaction time	.023	.019	.192	.235
	High-Stress					
Pillar	R2 = .003, *p* = .830	Constant	2.334	1.613		.170
		Reaction time	-.199	.911	-.058	.830
	Low-Stress					
	R2 = .060, *p* = .248	Constant	2.254	.280		.000
		Reaction time	.022	.019	.245	.248
	Total					
	R2 = .000, *p* = .956	Constant	.688	.116		.000*
		Reaction time	.001	.010	.010	.956
	High-Stress					
Fallen Hostage	R2 = .002, *p* = .887	Constant	1.090	1.078		.334
		Reaction time	-.088	.609	-.044	.887
	Low-Stress					
	R2 = .017, *p* = .597	Constant	.405	.080		.000*
		Reaction time	.003	.005	.130	.597

p < .05* *p* < .01**

Addressing our first hypothesis, findings from regression analyses (see Table 5) suggested that planning, information, and danger cues significantly explained 13.9% of the variance in gaze duration to pillar, where awareness of danger cues made a significant unique contribution (β = .332, *p* = .022). This was also true for the High-Stress group (β = .560, *p* = .010). The model significantly explained 27.8% of the variance in gaze duration to pillar for the Low-Stress group, with information making the biggest, significant contribution for the (β = .450, p = .012).

For our second hypothesis, reaction time was not found to be statistically significant in predicting variance in total visit duration to pillar (β = .192, p = .235).

Neither did regression analysis prove to significantly determine the predictive power of self-reported SA score on total length of gaze at fallen hostage, our third hypothesis. Here, the model only explained 0.6% of the variance in fallen hostage in the total sample, 0.7% for the High-Stress group and 18.9% for the Low-Stress group. Although not statistically significant, the variables awareness of danger cues for the Low-Stress group (β =-.353, p = .159) and planning for the High-Stress group (β = .266, p = .302) did make notable contributions to the equations.

Regression analyses of our fourth and final hypothesis were not found to be statistically significant for total and both groups, where reaction time explained as little as 0% of the variation in gaze duration to fallen hostage (β = .010, p = .950).

Discussion

Our predictions about the eye tracking data (pillar and fallen hostage) relative to the self-reported SA data and reaction time will be addressed in this section, followed by detailed discussions of differences observed between the High- and Low-Stress groups and the possible effects of stress on SA. In closing, some remarks on the limitations of the PHS study and eye tracking methods for measuring SA will be presented together with suggestions for further research.

With regards to our first hypotheses (*the higher the participants' (self-reported) SA scores are, the longer their gaze duration on pillar will be*), the results presented above indicate few straightforward associations between SA scores and total visit duration on pillar. Analyses generated only weak to medium correlations between scores on SARS questionnaire categories and gaze duration on pillar for the totality of the participants. Three of these were negative (scores on SARS "Planning" category questions for total sample and Low-Stress group, and "Awareness of situation" questions for High-Stress group), indicating that higher scores on these instances were actually associated with *lesser* total gaze duration on pillar. Although these results were inconclusive towards our first hypothesis, interesting differences between the High-Stress and Low-Stress groups were observed in relation to the possible role of stress on SA, discussed below. The regression model for the total sample showed statistically significant results, explaining 13.9% of the variance in gaze duration on pillar.

Although the explanatory contribution of the model was not particularly high, scores on SARS question categories planning, information, and awareness of danger cues had significant – although differing – predictive powers for the sample as a whole as well as for the Low- and High-Stress groups.

Addressing our second hypothesis (*the shorter the participants' reaction time is, the longer total gaze duration on pillar will be)*, correlation analyses (Table 4) showed no statistically significant findings related to participants' reaction time and attention to the pillar. Several reasons for this are possible. First, and as noted above (see p. 23), participants had limited or no experience in using firearms and thereby potentially also little knowledge of techniques for preparing and shooting, which may have impacted their reaction time. To use Lundberg's (2015) terms, the students had limited repertoire of schemata from earlier experiences with which they could frame situation and fully understand and predict the events and decisions with which they were confronted during the unfolding scenario, a factor which may also have affected their reaction time. Second, according to Klein (1993, 2008), most decisions made in operational settings are based on prior experiences (see p. 10). Since most participants had little practical experience in the field at the time of the PHS study, they may have had to rely on making decisions and acting based on an analytical approach. This would have forced them to make detailed plans and weigh their options for their next actions, rather than making quick and decisive strategies derived from experience, again affecting their reaction time. Moreover, the participants' lack of experience and practical training would also have made it difficult to develop any automaticity in key tasks relevant to the scenario, thus precluding quick reaction time. Automaticity is built up through repeating behaviors to such an extent that they become "second nature". As Endsley (1995, p. 45) explains, automatic processing tends to be effortless, fast, and unconscious, and can overcome the limitations of focused attention. Lastly, reaction time may have been affected by an error in decision making, "the information trap" (see p. 12 above), where the participants may have taken time to search the environment for more information, directing their attention to other elements than the pillar before deciding for a course of action. The delay caused by this search may also have affected their reaction time in the scenario.

The third hypothesis (*the lower the participants' (self-reported) SA score is, the longer the total gaze time on fallen hostage will be*) did not produce any statistically significant correlation or regression analyses. None of the scores from the SARS self-report question categories were found to be associated with or predict the amount of attention paid to the fallen hostage, suggesting that other factors were responsible for the variation observed.

However, some group differences were found. Scores on a majority of the SARS question categories showed more negative associations with attention to fallen hostage for the Low-Stress group than did scores for the High-Stress group (see Tables 3 and 4), indicating that the Low-Stress group may have ascribed less importance to the fallen hostage cue as part of the scenario than was the case among the participants of the High-Stress group.

The fourth hypothesis (*the longer the participants' reaction time is, the longer they focus their attention on fallen hostage*) also did not produce any statistically significant findings. The analyses found no relationship between the participants' reaction time and their gaze duration on fallen hostage, indicating that factors other than performance on reaction time accounted for the variation in gaze duration to fallen hostage in the PHS study.

Stress appears to have had some minor impact on both participant performance, focus of attention, and scores in several SARS question categories. Sandvik, Gjevestad, Aabrakk, Øhman, Kjendlie, Hystad, Bartone, Hansen and Johnsen (2019), in a separate article based on the PHS pilot study, discuss the link between physical fitness, hardiness and stress arousal. They show that the High-Stress group had significantly higher heart rate than the Low-Stress group. Further, they found that the High-Stress group reported significantly higher subjective stress than the Low-Stress group (see Sandvik et al., 2019, p. 6). This implies that the pre-scenario stress condition applied to the groups in advance of the main simulator test actually *did* affect the two groups on both levels of physical and mental stress. However, the correlation and regression analysis between groups show only small, weak relationships to the eye tracking data. Analyses revealed that the High-Stress group's self-reported awareness of danger cues predicted gaze duration on pillar, while the Low-Stress group's self-reported scores in the SARS information category questions was the best predictor of length of gaze at the pillar.

In the SARS questionnaire (see Table 1), the language used in the questions posed concerning awareness of danger cues to a large extent addressed the participants' memories of concrete, external elements during the scenario. These kinds of question could potentially have been more comprehensible and easier to answer by participants who did not have extensive experience in critical situations. In contrast, the questions from the SARS awareness of the situation category would have demanded at least some levels of introspection, imaginative thinking, and insight gained from previous experience in order to judge one's own level of "awareness of the situation". The concreteness and explicitness of being asked to what degree one was able to detect and act upon cues to danger or critical signals as part of the simulation may have elicited answers which reflected a more correct self-assessment of

SA than the abstract and general awareness of the situation questions. The questions identifying awareness of danger cues align very closely with Endsley's (1995) definition of Level 1 SA: *to perceive the elements in the environment*, perhaps more so than questions in other SARS categories. The danger cues questions also relate to Endsley's Level 2 SA, in that they reflect an understanding of the danger the perceived elements posed in the situation (otherwise the participants would not have assessed them as cues to danger). The fact that there was a positive association between the High-Stress group's attention paid to the pillar and scores on danger cue questions, could mean that the participants in the High-Stress group were able to perceive and understand the elements in the situation they were in. This may have reflected a certain level of SA even if their self-report scores to general (and abstract) awareness of the situation questions did not show such an association. Further, since awareness of danger cues were found to be significantly associated with gaze duration to pillar for the High-Stress group, one can argue that the eye tracking was, to some extent, able to measure degree of SA.

This could also indicate that participants in the High-Stress group, who were exposed to the stress-inducing treatment in advance of the main scenario, were primed to expect danger. In addition, the High-Stress group also received some advance warning over the dispatch radio reporting that shots had been fired at a nearby high school – information that was not provided to the Low-Stress group until right before the beginning of the scenario. This may have set certain expectations and mental images among members of the High-Stress group prior to the simulator test. In Lundberg's (2015) terms, members of the High-Stress group may already have developed – or were more likely to develop – a frame around the situation through which they could more readily make sense of and ascribe importance to danger cues due to the pre-scenario condition to which they were exposed.

Another explanation which could account for the differences between the two groups could be the effect of stress on attention. As noted above, stress can have taxing effects on attention processes. As such, a heightened level of stress may have caused the High-Stress group to develop selective attention for task-relevant attributes in the scenario (the pillar), in line with Easterbrook's (1959) Attention Approach.

Non-significant, but nonetheless interesting results show that the High-Stress group tended to pay more attention to the fallen hostage than the Low-Stress group. This could also be explained by the advance warning the High-Stress group received in an already stress-induced environment about shots being fired at a nearby high school and interpreted as a form of selective attention. In accordance with the Capacity-Resource Theory (Chajut & Algom, 2003), the pre-scenario stress the High-Stress group was exposed to and the advance information they were given, may have caused a narrowing in attention which led them to focus on task-irrelevant cues in the environment (fallen hostage) for a longer time than the Low-Stress group.

For the Low-Stress group, significant, moderate, positive correlations were also found between self-reported “information” and gaze duration on pillar. Information was also found to be predictive of gaze duration on pillar. This may indicate that the participants in the Low-Stress group were able to find a congruence between the information they received and the situation as it unfolded. Additionally, they were able to use that information and gather additional information from the environment, which may indicate that they took an analytical approach to making decisions rather than acting upon intuition. For the High-Stress group, however, the negative, although not significant, correlations between information and gaze duration to pillar may have been the result of the series of “changes of plans” they had experienced in advance of the main simulator scenario. This may have left them with little confidence in the information they were being provided or doubts as to what to expect next.

Furthermore, the significant correlations between learning experience and awareness of danger cues for the Low-Stress group can indicate that the participants in the Low-Stress group were able to ascribe importance to building an awareness of cues in the environment which signal potential danger as a learning experience compared to members of the High-Stress group.

Limitations

First, although the PHS study from which the data was taken was a pilot, its relatively small sample size ($n = 84$), might have affected the representability of the results attained (Bordens & Abbott, 2011). A larger sample may have given better estimates of the representability of the sample. The size of the sample may also have impacted the regression analyses, since in small samples the R square tends to overestimate the variance accounted for in the model due to sampling errors (Tabachnik & Fidell, 2001, as cited in Bordens & Abbott, 2011). To better estimate the variance explained by the different predictor variables, an adjusted R-square could have been reported.

Second, the participants in the study were police *students*, to a large degree without any experience of work in the field. According to Carreta, Perry and Ree (1994, as cited in Saus, Johnsen, Eid & Thayer, 2012), experience was found to be the best predictor of SA in a study of aviation. The lack of experience among the participants in the PHS study could have influenced the results on SA, in that participants had fewer mental models (Rouse & Morris, 1985; Endsley, 1995) or framing schemata (Lundberg, 2015) available to more readily assess the scenario and make appropriate and timely decisions.

Third, the use of data in this experiment based on a randomized two-group design has a number of limitations. The participants in the groups may have differed significantly on other characteristics than the ones accounted for in the study (e.g. personality traits and characteristics). The exclusion of such data from this analysis may have affected their performance on the dependent eye tracking variables (Bordens & Abbott, 2011). For example, Endsley and Garland (2000) have suggested that cognitive abilities, including mental simulation, analytical thinking, pattern matching, and story building are used in the process of developing SA. In a study of situation awareness among student anesthetics, Wright and Fallacaro (2011) found that participants with higher cognitive abilities showed higher levels of SA. Such individual differences in personality traits and abilities were not included in this exploratory analysis, and their effects on the results are thus unknown. Furthermore, small but crucial differences in the information the two groups received may have also contributed to skewing the results. As noted above, the High-Stress group received some degree of advance information about shots fired at a nearby high school which the Low-Stress group did not receive. This may have led participants in the High-Stress group to develop more solid expectations and mental images about the character of the situation before the scenario had begun, thus inserting an additional variable which was unaccounted for in the analyses.

Limitations of the SARS questionnaire. The manner in which participants perceived their own performance and outcome of the scenario could potentially have influenced their responses to the SARS questionnaire. For example, the participants who managed to shoot the second assailant (or all assailants), may have responded more positively to the SARS questions than participants who had been shot. It can be argued that the self-report design of the study conflated SA with outcomes, leading participants who had been shot to implicitly believe that they had low SA.

There are also a number of potential weaknesses with the questions addressing awareness of situation (see Table 1). As indicated above, some of them are particularly abstract (e.g. questions 14 and 15), and may therefore have been difficult to interpret. Since the participants were novices in this kind of setting, they may have had limited mental models from prior experiences to which they could relate the scenario. Therefore, one might ask to what extent participants could accurately judge whether the situation made them “offensive” or “defensive” without any frame of reference.

Additionally, self-ratings on a scale from zero to six may have led to several problems for the results. First, the participants may not have been honest in their self-reporting (Bordens & Abbot, 2011). Moreover, novices may not have had any scale on which to accurately judge their own subjective experience. In contrast to experienced law enforcement workers, who could have been in a better position to rate themselves in relation to prior experience in the field, novices may not necessarily have had this kind of knowledge yet, leading to inaccurate self-ratings.

Limitations of eye tracking methods for measuring SA. One limitation of measuring SA is that there exists no universally accepted criteria of what constitutes good SA, only a general assumption that the more information one manages to perceive and process, the better one's SA will be. This is, however, not necessarily the case in operational settings, since there is a greater demand for paying attention to specific task-related elements in the environment. It is therefore debatable if the objective SA variables used in the PHS study (eye tracking data for two defined AoIs, reaction time, and hits) have sufficient detail to function as a measure of SA, although the findings of this analysis point to some associations.

According to Endsley's (1995) three level theory of Situation Awareness, total visit duration on specific AoIs in the environment collected from eye tracking data can logically provide an indication of a perception of elements in the environment (level 1 SA), and possibly comprehension of those elements (level 2 SA), since neither perception nor comprehension are possible without attention (which is what eye tracking measures). However, it is doubtful that such data alone can give an indication of the projection and prediction of the future states of these elements (level 3 SA). This may have been possible if the eye tracking data was captured and analyzed more in terms of Lundberg's (2015) holistic framework of SA: as a series of events unfolding over an event horizon. A more detailed structuring and analysis of the data could have tracked the course of participant attention over time, what elements in the environment participants were aware of, and what decisions and actions were taken based on that awareness. For example, as part of the school shooting scenario, one of the pupils in the interactive video provided a subtle, non-verbal cue directing attention to the pillar as the location of the second shooter. The design of the experiment and the way that the data was collected does not allow us to know if this cue – or others like it – were perceived by the participant or not. While an experimental design which could capture such data could be technically challenging to execute, the more fine-grained information it would provide could prove to be more rewarding.

Suggestions for Further Research.

One of the limitations of the PHS pilot study is that it was solely conducted on a sample of students, who are generally novices when it comes to operational situations. A suggestion for further research would be to assess if the results of a repeat eye tracking experiment would show different results if conducted on experienced police officers. A meta-analysis by Gegenfurtner, Lehtinen & Säljö (2011), for example, showed that experts, compared to non-experts, had more gaze durations to task-relevant elements in the environment and better allocation of selective attention. Thus, an experiment conducted with experts may generate more accurate measures of SA without confounding variables due to inexperience.

Lastly, eye tracking in simulator environments can also be used in student feedback and learning. A study showed that eye tracking technology improved nurse and paramedic students self-reported learning and assisted feedback (Cooper et al., 2014).

Even though our hypotheses stated that self-reported SA and performance are predictors of eye tracking fixation were not supported, this does not necessarily mean that eye tracking is unable to measure SA. In this analysis, the results of the SARS questionnaire were subdivided into scores from topical categories. No overall score for the questionnaire as a whole was used, a shortcoming which could potentially have provided additional information and which should be included in future analyses using data from the SARS questionnaire.

Conclusion

Based on data derived from a simulator scenario pilot study, this exploratory analyses has examined the applicability of eye tracking methods for measuring situation awareness in police operational settings. Four different hypotheses developed from theoretical assumptions about eye tracking and situation awareness were explored. Results from statistical analyses of the data showed that these hypotheses were not conclusively supported. The factor that was found to be most associated with the eye tracking data was the score on the SARS category "awareness of danger cues", exposing potential weaknesses in the types of questions asked to measure SA in the study as well as in the study design. This thesis further addressed the role of stress in achieving and maintaining SA, the results of which showed that stress had a small, but interesting effect on attention and the perception of information. Limitations of the pilot study and SA measure in general were also addressed, together with suggestions for further research.

Even though SA is a widely used concept, there is a need in the literature for a generally accepted definition of SA and criteria of what constitutes good SA. This would not only help in solidifying the position of SA as a concept within several research areas but also allow for a better foundation for scholars to examine the eye tracking method in assessing SA. Eye tracking technology is a fascinating and advancing method, which has the potential to provide new information about individuals' situation awareness and attention, as well as assist in training students and workers in operational professions to develop and strengthen their awareness of the situations they face and the decisions they make. Hopefully, this exploratory analysis has contributed to insights about eye tracking methods and its usefulness in measuring situation awareness for future research.

References

Bacon, S. J. (1974). Arousal and the range of cue utilization. *Journal of Experimental Psuchology, 102*, 81-87.

Bordens, K. S., Abbott, B. B. (2011). *Research Design and Methods: a process approach*, (8th ed.). New York, USA: McGraw-Hill.

Brustad, L. (2018, August 2nd). Eric Torell (20) bar på en lekepistol. Ble skutt og drept av svensk politi. *Dagbladet*. Downloaded from https://www.dagbladet.no/nyheter/eric-torell-20-bar-pa-en-lekepistol-ble-skutt-og-drept-av-svensk-politi/70068829

Chajut, E., Algom, D. (2003). Selective Attention Improves Under Stress: Implications for Theories of Social Cognition. *Journal of Personality and Social Psychology, 28*(2). 231-248. Doi: 10.1037/0022-3514.85.2.231

Chidester, T. R., Kanki, B. G., Foushee, H. C., Dickinson, C. L. (1990). Volume 1. Leader characteristics and crew performance in a full-mission air transport simulation. In *Personality factors in flight operations. NASA Technical Memorandum, 102259.* Downloaded from https://ntrs.nasa.gov/search.jsp?R=19900014054.

Cooper, S., Browning, M., Ross, L., Sparkes, L., Williams, B., Munro, G., O'Meara, P., Black, B & Bogossian, F. (2014). *Can eye-tracking technology improve situational awareness and student feedback during simulation*? *Final Report*. (

Easterbrook, J. A. (1959). The Effect of Emotion on Cue Utilization and the Organization of Behaviour. *Psychological Review, 66*(3). 183-201.

Edvindsen, K. (2012). *Ordenstjeneste* (2nd ed.). Høvik: Forlaget Vett & Viten.

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

Endsley, M. R. (1996). Situation Awareness Measurement in Test and Evaluation. In T. G. O'Brien & S. G. Charlton (Eds.), *Handbook of human factors testing and evaluation* (pp. 159-180). Mahwah, NJ: Lawrence Erlbaum

Endsley, M. R. (1999). Situation Awareness in Aviation Systems. In D. J. Garland, J. A. Wise and V. D. Hopkins (Eds.) *Handbook of Aviation Human Factors*. Mahwah, NJ: Lawrence Erlbaum Associates.

Endsley, M. R., Holder, L. D., Leibrecht, B. C., Garland, D. J., Wampler, R. L., Matthews, M. D. (2000). Modelling and Measuring Situation Awareness in the Infantry Operational Environment. *TRW Systems Integration Group Fairfax VA Systems DIV.*

Endsley, M. R., Garland, D. J. (2000). *Situation Awareness Analysis and Measurement.* Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Eid, J., Johnsen, B. H., Laberg, J. C. (2018). Introdkusjon til operativ psykologi. In J. Eid, B. H. Johnsen (Eds.), *Operativ Psykologi* (3rd ed.) (pp. 11-30). Bergen: Fagbokforlaget

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

Graham, S. E., Endsley, M. R., Weeks, J. L., Strater, L. D. (2001). Situation Awareness for Military Ground Forces: Current Issues and Perspectives. *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting,* 351-355.

Gegenfurtner, A., Lehtinen, E., Säljö, R. (2011). Expertise Differences in the Comprehension of Visualizations: a Meta-Analysis of Eye-Tracking Research in Professional Domains. *Educational Psychology Review, 23*(4), 523-552.

Hollnagel, E. (1993). Human reliability analysis: context and control. In B. R. Gaines and A. Monk (Eds.), *Computer and People Services*. London, UK: Academic Press.

Holm, S. (1979). A simple sequential rejective method procedure. *Scandinavian Journal of Statistics, 6*, 65-70.

Johnsen, B. H. (2018). Beslutningstaking i operative situasjoner. In J. Eid & B. H. Johnsen (Eds.), *Operativ Psykologi* (3rd ed.) (p. 251-264). Bergen: Fagbokforlaget

Just, M. A., Carpenter, P. A. (1980). A theory of reading: from eye fixation to comprehension. *Psychological Review, 87*(4), 329-354.

Kjærgaard, A., Leon, G. R., Venables, N. C., Fink, B. A. (2013). Personality, Personal Values and Growth in Military Special Unit Patrol Teams Operating in a Polar Environment. *Military Psychology, 25*(1), 13-22. DOI: 10.1037/h0094753

Klein, G. (1993). A Recognition Primed Decision (RPD) Model of Rapid Decision Making. In G. Klein, J. Orasanu, R. Calderwood and C. E. Zsambok (Eds.), *Decision Making in Action* (pp. 138-147). Norwood, NJ: Ablex.

Klein, G. (2008). Naturalistic Decision Making. *Human Factors, 50*(3), 456-460. Doi: 10.1518/001872008X288385

Klein, G., Wiggins, S., Dominguez, C. O. (2010). Team Sensemaking. *Theoretical Issues in Ergonomics Science, 11*(4), 204-220.

Kvale, S. (1996). *InterViews: An Introduction to Qualitative Research Interviewing*. London: Sage Publications.

Lundberg, J. (2015). Situation Awareness Systems, States and Processes: A holistic framework. *Theoretical Issues in Ergonomics Science*. doi: 10.1080/1463922X.2015.1008601.

Muehlethaler, C. M., Knecht, C. P. (2016). Situation Awareness Training for General Aviation Pilots using Eye Tracking. *IFAC PapersOnLine*, *49*(19), 66-71.

Neisser, U. (1976). *Cognition and Reality: Principles and Implications of Cognitive Psychology.* San Fransisco, CA: W. H. Freeman and Company

Neville, T. J., Simon, P. M. (2016). Never blame the umpire – a review of Situation Awareness models and methods for examining the performance of officials in sport. *Ergonomics*, *59*(7), 962-975. DOI: 10.1080/00140139.2015.1100758

Orasanu, J., Martin, L., Davidson, J. (1998). Errors in Aviation Decision Making: Bad Decisions or Bad Luck? *Fourth Conference on Naturalistic Decision Making*; May 29-31. Warrenton, VA: United States.

Pallant, J. (2016). *SPSS survival manual: a step by step guide to data analysis using SPSS* (6th ed.). Maidenhead: McGraw-Hill Open University Press.

Perry, C. M., Sheik-Nainar, M. A., Segall, N., Ma, R., Kaber, D. B. (2008). Effects of physical workload on cognitive task performance and situation awareness. *Theoretical Issues in Ergonomics Science, 9*(2), 95-113. DOI: 10.1080/14639220600959237

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

Salmon, P. M., Stanton, M. 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.

Salmon, P. M., Stanton, M. A. (2013). Situation awareness and safety: Contribution or confusion? Situation awareness and safety editorial. *Safety Science, 56*, 1-5.

Sandvik, A. M., Gjevestad, E., Aabrekk, E., Øhman, P., Kjendlie, P. L., Hystad, S. W., Bartone, P. T., Hansen, A. L., Johnsen, B. H. (2019). Physical Fitness and Psychological Hardiness as Predictors of Parasympathetic Control in Response to Stress: a Norwegian Police Simulator Training Study. *Journal of Police and Criminal Psychology*. Doi: https://doi.org/10.1007/s11896-019-09323-8

Saus, E. R., Johnsen, B. H., Eid, J. (2010). Perceived learning outcome: the relationship between experience, realism and situation awareness during simulator training. *International Maritime Health, 61*(4), 258-264.

Saus, E. R., Johnsen, B. H., Eid, J., Thayer, J. F. (2012). Who benefits from simulator training: Personality and heart rate variability in relation to situation awareness during navigation training. *Computers in Human Behavior, 28*, 1262-1268. Doi: 10.1016/j.chb.2012.02.009

Simon, H. A. (1956). Rational Choice and the Structure of the Environment. *Psychological Review, 63*(2), 129-138.

Sneddon, A., Mearns, K., Flin, R. (2013). Tress, fatigue, situation awareness and safety in offshore drilling crews. *Safety Science, 56,* 80-88.

Sommerfelt-Pettersen, J. (2018). *Forord*. In J. Eid & B. H. Johnsen, (Eds.) Operativ Psykologi (3rd ed.). Bergen: Fagbokforlaget.

Tatler, B. W., Kirtley, C., MacDonald, R. G., Mitchell, K. M. A., Savage, S. W. (2014). The Active Eye: Perspectives on Eye Movement Research. In M. Horsley, M. Eliot, B. A. Knight & R. Reilly (Eds.) *Current Trends in Eye Tracking Research*. New York: Springer International Publishing.

Taylor, R. M. (1990). Situational Awareness Rating Technique (SART): The development of a tool for aircraft system design. *Situational Awareness in Aerospace Operations (AGARD-CP-478)*, 3/1-3/17. Neuilly Sur Seine, France: NATO-AGARD.

Tobii Technology. (2012). Tobii Glasses Eye Tracker: User manual (https://guidessimo.com/document/704779/tobii-glasses-3d-glasses-70.html)

Van Winsen, R., Henriqson, E., Schuler, B., Dekker, S. W. A. (2015). Situation awareness: some conditions of possibility. *Theoretical Issues in Ergonomics Science, 16*(1). 53 -68.

Waag, W. L., Houck, M. R. (1994). Tools for assessing situational awareness in an operational fighter environment. *Aviation Space and Environmental Medicine, 65*(5). A13-9

Weiss, R. S. (1994). *Learning from strangers, the art and method of qualitative interview studies*. New York, NY: Free Press.

Weltman, G., Smith, J. E., Egstrom, G. H. (1971). Perceptual narrowing during simulated pressure-chamber exposure. *Human factors, 13*, 99-107.

Wright, S. M., Fallacaro, M. D. (2011). Predictors of situation awareness in student registered nurse anaesthetics. *Journal of the American Association of Nurse Anaesthetics, 79*(6). 484-490.

Zhang, Y., Jin, S., Garner, M., Mosaly, P., Kaber, D. (2009). *The effects of aging and cognitive stress disposition on driver situation awareness and performance in hazardous conditions*, proceedings of the IEA 2009 17th World Congress in Ergonomics, Beijing, China.