Cognitive Factors Mediating Situation Awareness

Alastair McGowan

Thesis submitted to
Cardiff University for the degree of
Doctor of Philosophy

September 2006

Index of Figures

Index of Tables

Publication

Much of the work included in Experiment 3 also appears in the following book chapter:

McGowan, A.M. and Banbury, S.P. (2004) Evaluating Interruption-Based Techniques using Embedded Measures of Driver Anticipation. In S. P. Banbury, & S. Tremblay, *A Cognitive Approach to Situation Awareness: Theory and Application* (pp. 154-173).

Acknowledgements

I would like to thank Clare, Sam and Grace, and my parents for their support without which I would not have completed this work. I especially thank Drs Simon Banbury and Chris Miles for their guidance and constructive comments throughout this learning process. The research in this thesis was supported by the provision of time and expertise by the South Wales Police road traffic section and the Cardiff group of the Institute of Advanced Motorists. And thank you to Q and the Flying Carpet percussion circle for the implicit processing therapy.

Summary

The six experiments reported in this thesis tested Endsley's (1995) three level theory of perception, comprehension and projection in SA alongside relevant cognitive theories, using a driving hazard perception test (HPT) of both hazard recognition and hazard anticipation as the main dependent measures. Experiments 1 and 2 tested the effects of training and expertise on the HPT, revealing a positive association between hazard anticipation and both SA training and expertise. Experiment 2 revealed a potentially complex relationship between expertise and recognition and anticipation. The effects of training Endsley's (1995) perception and comprehension levels of SA did not appear to be additive as predicted by that theory.

Experiments 3 and 4 tested the effects of concurrent freeze probes and real time SA probes on the HPT. This revealed a negative effect of interruption and a positive effect of reorientation (revealing the rationale of the task) associated with the use of a freeze probe, in terms of hazard anticipation. Furthermore, these two effects appear to be mutually cancelling. It was also found that notification of forthcoming online probes does not ameliorate the negative effects of those interruptions in terms of hazard anticipation.

Experiments 5 and 6 tested the effects of working memory interference in terms of visual-spatial, phonological and episodic buffer processing on the two HPT measures. This revealed more deleterious effects on hazard recognition associated with visual-spatial and episodic buffer interference than with phonological interference. It is argued that in terms of SA related visual processing during driving Milner and Goodale's (1992) dual pathway theory appears to have more explanatory power than Endsley's (1995) theory of SA.

CHAPTER 1
INTRODUCTION

> Although cognitive processes are an integral part of many frameworks of SA, their description is at a rather superficial level. Indeed, models of SA refer to cognitive processes in general terms, but do not specify exactly what processes are involved and to what extent. Instead, the major impetus of research has been on the development of techniques to measure SA, at the expense of a more rigorous understanding of *why* SA varies under certain psychological and environmental conditions.
> (Banbury and Tremblay, 2004, p. xiii)

General

The above statement illustrates the different ways that cognitive psychology and human factors engineering approach situation awareness (SA) research. In the field of human factors SA has been investigated with a focus on measurement, while cognitive psychology is concerned with how SA develops.

The aim of this thesis is to inform both SA application and theory from a cognitive theory perspective, in order to contribute to a psychological understanding of why SA varies. The goal is to measure SA performance using an embedded measure of recognition and anticipation based on a driving hazard perception test in order to investigate cognitive processes and task-related factors, thereby contributing to cognitive SA theory and applied task-focused frameworks of SA. Specifically, two areas of study, situation awareness in general, and driving hazard perception which is focused on awareness during driving, give complementary accounts of what appears to be the same phenomenon. Situation awareness research tends to be focused around the general concept of a scale or levels of awareness, and theory built around levels of SA (e.g. Endsley, 1995). Hazard perception is an aspect of driving where SA is important. Hazard perception can be measured in terms of hazard recognition and hazard anticipation, both of which appear to contribute to skilled anticipation

(Horswill & McKenna, 2004). Both fields of study appear to be investigating the human ability to perceive and be aware of relevant factors in their environment and the use of this information in order to take anticipatory action. This is in contrast to human factors research which to date has mainly used observational and questionnaire based assessments of SA, the main dependent measure which will be used in the following thesis is a modified version of the driving hazard perception test (HPT; Pelz & Krupat, 1974; McKenna & Crick, 1991). The modified version (MHPT) measures both recognition and anticipation of hazards and in so doing it measures the perception and projection levels of Endsley's (1988; 1995) theory of SA.

A secondary objective of this thesis is to provide some initial analysis of the utility of hazard perception testing as a laboratory-based measure of SA during the hazard perception driving sub-task. Hazard perception has been described by Horswill and McKenna (2004) as SA for the driving task. The HPT provides what can be described as an overall index of SA in driving by measuring a driver's recognition or anticipation of hazards presented on video. The HPT can be embedded in driving simulation or video monitoring tasks to produce a measure that is less susceptible to the effects of interference, interruption. It also attracts less of the methodological issues inherent to most measures of SA. For example, the situation awareness global assessment technique (SAGAT; Endsley, 1995; described on p. 37) pauses a simulator to present questionnaire items. SAGAT therefore has the likely potential to alter performance by interrupting, interfering with, or exposing the rationale of the test during what would otherwise be a continuous dynamic performance. These specific issues will be explored in Chapter 3. In contrast with measures such as SAGAT the embedded HPT is likely to provide a higher resolution and more ecologically valid measure suited to laboratory research into the cognitive factors in SA. This will also be a test of the utility of an embedded HPT measure for the measurement of SA.

This thesis will be introduced firstly by defining SA, initially by the exploration of two key issues. These are the distinction between explicit and implicit awareness, and the distinction between SA as a product and SA as a process. The definitions of SA necessarily invoke aspects of SA, cognitive theory and epistemological considerations. Following the definitions the SA theory section will examine assumptions more closely, and in relation to cognitive theory. Research into hazard perception and its measurement during driving will then be discussed, followed by a description of SA and hazard perception measures which will be used in the following experiments.

Definitions

In simple terms SA is '...knowing what is going on around you' (Endsley, 2000b). However, defining SA has disproportionately occupied many researchers and has produced a wide range of different definitions around the same theme. See Dominguez (1994) and Breton and Rousseau (2001), for full reviews. It is clear from the definitions in these reviews that little progress has been made from the high-level definition above to a more detailed and explanatory definition in cognitive terms. Nor has there been much agreement with regard to the meanings of the separate terms awareness and situation. These difficulties are related to a similar lack of clarity in the development of SA theory - e.g, controversy surrounding Endsley's (1995) model of SA. The present section will describe some definitions of SA while remaining mindful of the default view held by many researchers that the construct is possibly a distraction from focusing on known cognitive processes and their relationships with factors in the world that are critical to the performance of complex dynamic tasks.

Before attempting to draw a definition of SA from the existing literature it is necessary firstly to clarify an area of contention over the degree to which a distinction is necessary between conscious awareness (explicit SA), and SA that can only be inferred from behaviour (implicit SA). Secondly, a distinction has also been made between SA as a *product* (the elusive state of 'having SA') and SA as a *process* (information structures and cognition). These distinctions are pertinent to the investigation of cognitive processing during SA and hence important for this thesis.

Implicit and explicit SA

Implicit awareness is that which is inferred from behaviour, in human factors research this is usually within the context of a task. In order to respond correctly to a given situation other than by chance a person must have had awareness of relevant information at some level, since processing of that information is a logical necessity for performance of the task. This is irrespective of whether or not the person is able to consciously articulate the relevant information content that she has been processing. Implicit awareness is therefore awareness deduced from behaviour observed in relation to a task. Explicit awareness can be defined as occurring when the operator can articulate information about a situation. These two forms of awareness may be experienced differently and they may be processed differently, although ultimately they both serve the same function of performance on a task. However, a central question arises as to the extent to which conscious explicit awareness is necessary for the performance of a task since it is accepted by Endsley (1995) as a central component of that widely cited theory.

Driving without awareness mode (DWAM; e.g., May & Gale, 1998) is a phenomenon which relates to the distinction between implicit and explicit awareness during the performance of a task. DWAM occurs when a person has driven a car for a period and cannot subsequently recall having done so. If a person can drive a car effectively or safely, a complex dynamic task which requires real-time adaptive behaviour, without conscious awareness then conscious awareness cannot be deemed entirely necessary to the performance of that task. It is therefore proposed that explicit SA (conscious awareness) is not a necessary condition for performance. This is not to say that conscious awareness cannot enhance performance in certain situations (e.g., novel situations) but that explicit SA is not always a prerequisite for adequate performance. For the purpose of this thesis a definition of awareness therefore includes both implicit awareness and explicit awareness.

The distinction between implicit and explicit awareness is conceptually related to Shiffrin and Schnieder's (1977) distinction between automatic and controlled processing. That is, implicit awareness can be said to be non-conscious in the same way that automatic processing does not require conscious control and hence conscious awareness. In contrast, explicit awareness equates with conscious awareness and relates to controlled processing. The distinction between automatic and controlled processing therefore arises in relation to definitions of SA, and more widely it is related to the theoretical basis of SA in the next section. Shiffrin and Schneider (1977) and Schneider and Shiffrin (1977) proposed a theory that automatic and controlled cognitive processing are separate. The definition of automatic processing, from Shiffrin and Scheider (1977) is that it has no capacity limitation and does not require attention. Controlled processing has a limited capacity and depends on attention. Modifying an automatic process is more difficult than modifying a controlled process, since automaticity is highly inaccessible to conscious control, whereas by definition controlled processes are accessible and controllable. Shiffrin and Schneider's theory has been criticised by Cheng (1985) because the way in which information is *structured* may account for the differences attributed to automatic and controlled processing. Subsequently, Norman and Shallice (1986) subdivided Shiffrin and Schneider's controlled component into two processes: content scheduling, which is only partly controlled and manages conflicts between automatic processes; and fully executive control, which Baddeley (1986) argues to be part of his working memory theory, namely the central executive. The distinction between automatic and controlled processing and the distinction between implicit and explicit awareness tend to run in parallel. Throughout this thesis for brevity they are often contracted, for example by using implicit-automatic when referring generally to automatic processing that is also inaccessible to conscious awareness.

Process and product

A further distinction arises, in terms of defining SA, between SA as a process and SA as a product. SA as a process refers to the way that information is processed in relation to a task. This processing can be in terms of both cognition and the way that actions and states of information in the world support information processing more widely. SA as a product refers to a state of knowledge which is often assumed to refer to explicit knowledge held by a person about a situation and which can be used consciously for decision-making and action. It has been argued (e.g., Banbury, Andre & Croft, 2000) that in defining SA no distinction need be made between SA as a process and SA as a product. These are unnecessary abstractions from focusing on performance during a task and associated cognitive theories. The combination of processes and product of SA (assessment, knowledge, decisions and actions) is irreducible due to the complex and interdependent nature of cognition and environment (c.f., Gibson, 1979). This position is also taken by Smith and Hancock (1995) in that '...SA resides through the interaction of the person with the world [...] in terms of the cognitive *processes* used to engineer it and also the continuously updating *product* of SA (Stanton, Salmon, Walker, Baber & Jenkins, 2005, p. 216). This permits progress beyond the concept of awareness to an understanding of the interdependence between situation assessment and decisions and actions in the real world, all of which are memory and information processes which should be investigated as such unless they support the development of a superordinate theory of SA. A theory of awareness is currently of peripheral importance.

However, one cognitive neuropsychological theory which has been gaining influence of late argues that conscious awareness *is necessary* for the interaction between a person and the world and for effective performance at a task. Baars' (1988; 2002) global workspace theory of consciousness argues that conscious awareness is necessary for executive mediation of automatic cognitive processes related to interaction with the world. It is a radical alternative to current conceptualisations of attention, perception and memory (Baars, 1997). The global workspace theory states that consciousness is a form of transient memory which provides access between representations of external and internal perceptual information content. This integrative and neurally global process is responsible for explicit states of knowledge and the subjective state of awareness (consciousness) that further emerges from the explicit states of knowledge. Baars (2002) uses a metaphor of a theatre stage to describe the functioning of the global workspace. The primary function of conscious awareness is to provide an executive spotlight on the information content created when sensory areas of the cortex are activated. This spotlight gives rise to internal representations of the same activation, this he argues causes subjective experiences of emotion, visual imagination and inner speech. It also works as a feedback process within cognition for holding transient states of the world active in memory for adaptation and learning; visual imagination appears to relate to visual-spatial working memory and inner speech appears to relate to phonological working memory.

Nevertheless, due to the complexity and irreducibility of the neural systems which give rise to this consciousness module the concept of process and product may be false. According to Baars, conscious awareness and explicit states of knowledge are transient states of memory superordinate to states of memory within for example phonological or visual memory processes. These transient states of memory permit communication between otherwise separate cortical memory processes arising from perception and memory. However, the *experience* of subjective 'visual' imagery and 'phonological' inner voice are also necessary for both controlling action and for developing the superordinate global temporary memory to which they are subordinate. That is, an increase in the accuracy of subjective experience supports accuracy of states of explicit knowledge and implicit decisions and actions and performance at a task. This means that there is a two-way or cyclical process occurring; a complex system which may be irreducible. The global workspace theory also argues that automatic processing is absolutely limited without conscious awareness, and that conscious awareness is an important factor in learning automatic processes. This indicates a circularity of processing between on the one hand explicit states of knowledge and subjective consciousness, and on the other implicit states of automatic processing and actions. It also indicates the possibility that the elusive state (product) called 'awareness' that we are trying to define may really exist but that it is inseparable from actions and the wider world in which information is embedded. The global workspace theory may be beginning to sound like an attempt to collapse the mind-body distinction by loosely appealing to complexity theory but Baars' research does indicate that automatic and controlled processes may be interdependent and irreducible; a distinction between the two concepts is not supported by the cognitive neuropsychological evidence. Dekker and Lützhöft (2004) make a similar point concerning the irreducibility of mind and environment. The unpredictability of environments is a continuum which appears to correspond to what we term awareness; low states of awareness are automatic and high states of awareness are controlled or conscious. Therefore we may be mistaking awareness for cognitive complexity. Inferring states of awareness from complexity is to make a conceptual leap that is not supported by evidence. Banbury and Tremblay (2004) describe SA as an epiphenomenon for the same reason.

In summary, the global workspace theory states that awareness is a combined temporary global memory *process* and critical cognitive *process* that supports automatic goal-related behaviour and gives rise to related explicit decisions and access to explicit knowledge. The awareness product is also a separate emergent property of cognition in terms of subjective consciousness. It is argued that the irreducibility of these interdependent concepts which have been conceptualised as being separate (e.g. Endsley, 1995) leads to substantial confusion over a definition of SA.

Task focused perspective

In contrast with the confusion over a definition of SA, Durso and Gronlund (1999) argue that a distinction can be made between what is and what is not SA. That is, attentional mechanisms (even during conscious waking states) that drive automatic action cannot be called SA, and that the term SA is specific to some area of conscious, executive meta-cognitive functioning which relates to unpredictable environments. This is similar to Endsley's (1995) view of SA which will shortly be described, in which SA is regarded as some discrete phenomena.

A second reason for the lack of a clear definition of SA is that the awareness must necessarily be directed at some task-related goal. That is, the situation is always within the context of a task and its goal (Patrick & James, 2004). A definition of awareness must include the task-related goal of - for example, flying an aircraft safely and effectively in order to complete a sortie. Flach (1995) describes this as the correspondence problem; the awareness must relate to some objectively defined correct state of a situation. To some extent this problem is solved by defining awareness in the context of a specific set of goals (i.e., the task). This means that the situation necessarily comprises factors which have some relationship with the task goals, and the awareness as a function must be related to the relevant factors required to perform and complete the task.

SA definitions

A taxonomy of consciousness has been developed by Chalmers (1996). There is one category of consciousness that is similar to the concept of SA, which he describes as *functional* consciousness. This is distinct from the subjective experience of awareness. Functional consciousness is awareness related to interactions with the world, or external awareness a '...state wherein we have access to some information, and can use that information in the control of behaviour' (Chalmers, 1996, p. 28). This form of consciousness can therefore be explicit and controlled or implicit and automatic. Functional consciousness could therefore be taken as a broad descriptive definition of SA. Smith and Hancock (1995) also define SA as '...*Adaptive Externally Directed Consciousness*'. However, both Chalmers' and Smith and

Hancock's definitions are merely descriptive. Even in terms of Barrs' neuro-psychological theory it is still not clear how the crucial state or process of awareness differs from cognitive processes or presence of information content. It is proposed that the synonyms of SA, functional consciousness, and externally directed adaptive consciousness are a shorthand for an emergent factor or 'epi-phenomenon' (Banbury & Tremblay, 2004) that is simply a hypothesis at this time. Banbury and Tremblay introduce their discussion of SA definitions with the following quote which illustrates this point:

> The term *situation awareness* shares a common history with several psychological concepts such as intelligence, vigilance, attention, fatigue, stress, compatibility, or workload. During decades, all these terms were poorly defined. However, each became important because they attracted attention on critical processes or mental states that were previously unknown. Ultimately, they changed the ways to study human factors problems, and they brought new benefits. (Pew, 2000, p. 33)

To avoid being drawn further into nebulous SA definitions which are unlikely to illuminate the situation any more than Chalmers' (1996) concept of functional consciousness or Baars' global workspace theory, a brief review of definitions of SA in the field of human factors follows.

Existing SA definitions

Perhaps the most widely cited definition of SA in human factors research is Endsley's (1988; 1995) three level product based '...perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future' (Endsley, 1988, p.97). This definition is related to Endsley's claim that SA is a state of knowledge distinct from the processes known as situation assessment from which that knowledge arises. Dominguez (1994; cited in Pew & Mavor, 1998) produces a similar definition to Endsley and yet implies both product and process: SA is the 'continuous extraction of environmental information, integration of this information with previous knowledge to form a coherent mental picture, and the use of that picture in directing further perception and anticipating future events'. Dominguez developed this definition from a full review of existing definitions. Subsequently, Breton and Rousseau (2001) developed a further review of definitions which they used (Rousseau, Tremblay and Breton, 2004) in a classification of approaches to SA and resulting definitions. This produced a distinction between operator oriented approaches and situation focused approaches. Operator approaches focus on SA as a characteristic of an operator, of which both Dominguez' (1994) and Endsley's (1995) definitions are examples. The situation focused approach focuses on the situational affordances and constraints that affect an operator's SA during interaction with an environment. The situation or task-based approach, along with a cognitive focus, is the basis of the current thesis in order to address Flach's (1995) correspondence problem.

A distinction between product and process oriented theories has also arisen. Key examples are Endsley's (1995) definition and theory of SA as product, and Sarter and Woods' (1995) process perspective which is illustrated by the position that 'the term situation awareness should be viewed as a label for a variety of cognitive processing activities that are critical to dynamic, event-driven, and multitask fields of practice'. Sarter and Woods (1995) argue that a fully explanatory definition of SA is more likely to be found in terms of cognitive factors related to the phenomenon.

Baars (1988; 2002) global workspace theory indicates that the task of drawing apart cognitive factors mediating SA is unlikely to progress quickly, if at all. The factors are only beginning to be investigated hence the broad range of perspectives and few precise indications of clear factors provided in Banbury and Tremblay

(2004). However, there are indications that factors such as long-term working memory (LT-WM; Snow & Reising, 2000) and both fluid and crystallised intelligence (Gugerty, Brooks & Treadaway, 2004) may be involved in mediating SA. These factors generally accord with Baars' global workspace theory which argues that a form of transient and superordinate executive memory provides an integrative cognitive function. Therefore a cognitive definition of SA would be that it is a label for executive or metacognitive processing during dynamic interaction with a task and related to effective performance.

From a cognitive viewpoint SA as a product can only be said to exist to the extent that transient states of processing exist separately from the temporal continuum of processing; SA is a momentary abstraction from ongoing processes. SA and functional consciousness are merely a convenient shorthand for the phenomenon which the researcher will be investigating in terms of the range of cognitive factors (operator approach) associated with implicit and explicit decision-making, action and states of knowledge in relation to clear task-related goals (situation approach). Therefore, from a cognitive perspective (notwithstanding Baars, 1988; 2002) no clear definition of SA is offered for use in this thesis, since to do so would be to continue the reification of the idea that SA exists separately as a product of, or distinct process within, cognition. Instead, the tendency will be to focus on cognitive processing during the performance of complex tasks, in later sections working memory and executive function, as a means to explore the elusive phenomenon generally known as functional consciousness or SA. This exploration will be carried out in terms of the correspondence between operator cognition and a clear task situation, both of which may be components of the phenomenon in question.

Theories

This section will firstly continue to focus on some of the fundamental assumptions about SA in relation to existing SA theories. Specific cognitive theories that may be related to SA will be introduced at the point in the thesis at which they become relevant. Due to the contested nature of the SA construct no firm conclusion is provided. However, it will be seen that an essential first step in investigating SA, the definition and measurement of the task *situation*, has been missed in much of the research. This will be addressed in a subsequent section on measuring SA. Secondly, anticipation and hazard perception theory will be discussed.

SA Theory

Endsley's theory of SA

In the previous section a cognitive theory of consciousness was discussed, Barrs' (1988; 2002) global workspace theory. This theory has been developed mainly in order to explain subjective states of conscious awareness. It is not specifically a theory of SA but it does contribute to an understanding of SA from a cognitive perspective.

The first attempt to develop a specific theory of SA in human factors research was that of Endsley (1988, 1995), whose model is illustrated in Figure 1. Endsley's theory of SA was developed in order to explain the behaviour of pilots in relation to experience. This theory is based on the assumption that there is a product of cognition that can be called SA, and that there is a separate assessment process (e.g Endsley, 1995b) which produces this SA. Endsley's definition of SA describes the SA product in three levels: '… the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future' (Endsley, 1988, p.97). This produces a structure of SA product in terms of perception which is level 1 SA; comprehension which is level 2 SA; and comprehension which is level 3 SA. According to the theory the three levels also '...represent ascending degrees of SA' (Endsley, 2004, p. 319). Decision-making, long-term memory and automaticity, and information processing are all considered to be assessment processes that are separate from the three levels of SA. However, the three ascending and sequential levels of SA are composed of what appear to be cognitive processes. This is the main area of confusion over Endsley's theory.

Endsley (2004) attempted to clarify the main points of the theory which are as follows: Perception, comprehension and projection levels of SA product; goals direct attention, and the relevance of perceived information also directs attention; memory directs attention; working memory demands restrict SA in novices and in novel situations; mental models support the development of meaning and relevance of information and in turn this supports mental projections of future states; and pattern matching of perceived information to schema.

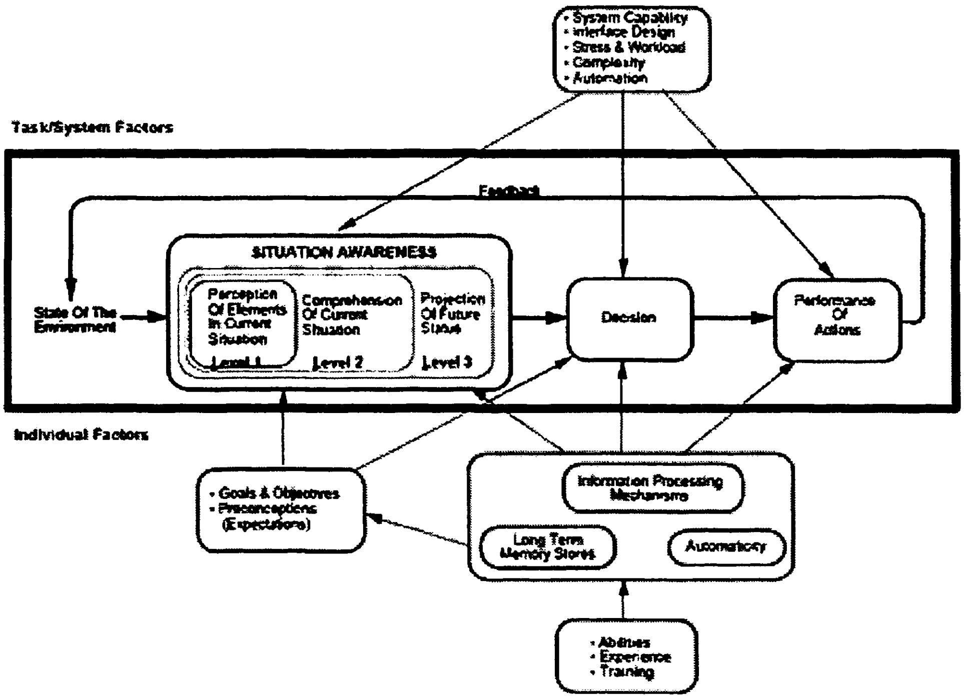

Figure 1. Endsley's model of SA, showing the perception, comprehension, and projection levels of processing or levels of SA. (Endsley, 2004, p318)

Endsley provides the following evidence to support the three level SA hypothesis. A relationship was observed between levels of SA and pilots' behaviour (Endsley & Robertson, 2000). That is, a pilot exhibiting behaviours associated with the perception level (Level 1 SA) has lower SA than a pilot exhibiting behaviours associated with the comprehension level, and a pilot exhibiting behaviours associated with the comprehension level (Level 2 SA) has lower SA than a pilot exhibiting behaviours associated with the projection level (Level 3 SA). Each level relates to a characteristic of pilot behaviour at specific levels of experience. For example, a pilot in earlier stages of experience tends only to engage in perception of information with little or no comprehension or projection behaviour. However, a degree of circular reasoning exists in this theory since Endsley does not explain whether it is the behaviour that is giving rise to the level of SA or vice versa.

Although the three levels of SA are described in Endsley (1988, 1995) as a product of assessment processes and environmental factors the levels themselves also clearly refer to perceptual and cognitive processing: perception, comprehension and mental projection. There is therefore ambiguity over whether Endsley's theory relates to SA as a product or to SA as a process (e.g. Flach, 1995). Endsley (2004) repeats this ambiguity, by stating that SA is a 'cognitive construct', i.e. a process as opposed to a data product, while repeating the theory that the '...three levels of SA, perception (level 1), comprehension (level 2) and projection (level 3), represent ascending degrees of SA' (Endsley, 2004, p. 319). However, by ascending *degrees of SA* a product rather than a process is implied.

Endsley (2004) does little to clarify her theory but instead continues to claim that SA is a product, and that situation assessment processes (both causing SA and leading to decisions and actions) are somehow separate. The theory views goals and objectives as separate from the three levels of SA, decision-making also separate from SA, and the processes of memory, information processing and automaticity as a separate situation assessment area. Furthermore, environmental (task related) conditions are described as being separate from goals and objectives. Endsley describes how the theory is in contrast with ecological theories of cognition such as Gibson (1979) in which none of the aforementioned factors are theorised to be separate. According to Gibson's theory SA would be considered to be as much a property and product of the task environment as of cognition. Gibson's theory tends to accord with Baars' (1988, 2002) theory of consciousness in terms of its claim of irreducibility within cognitive function, and between cognition and environmental factors. In contrast, Endsley's only significant evidence for her account of SA is the relationship between pilots' behaviours and experience, despite circular reasoning pervading the argument.

Endsley's (1988, 1995) theory has been criticised for its assumptions concerning the existence of SA as a mental phenomenon (Flach, 1994), and the reliance on what appears to be circular reasoning (Flach, 1995). Flach argues that there is circular reasoning in terms of the description of SA at a process level of description. The identification of situation awareness as a phenomenon merely identifies the concept, it does not explain the process. Given the lack of a clear definition of SA even this identification is far from complete. Flach's argument is that there appears merely to be a relationship between independent variables (i.e., pilots' experience) and a dependent variable (i.e., measures of SA) but this alone is insufficient to show that there is any causal relationship between levels of piloting experience and proposed levels of SA. In the case of Endsley's initial research the measure of SA was a questionnaire based assessment known as the situation awareness global assessment technique (SAGAT). This measure is based on the three level SA model. It was used despite the SA construct being the subject of the investigation. Therefore, this is one level of circularity of reasoning which confounds a clear hypothesis over the existence of SA. The evidence can only be used to state that there is a relationship between pilots' experience and their behaviour in terms of passively perceiving information, understanding information, and projection and active searching for information.

Endsley claims that there are three incrementally increasing forms of awareness, but the SA concept that Endsley identifies through association with pilot experience is then argued to be the cause of itself. That is, to state that loss of SA is the leading cause of human error or aviation mishaps is to apply causal reasoning to the concept of SA which has been deduced without the required experimental manipulation of an independent variable. This confuses the identification of a concept with the identification of a causal process, and thus relies on circular reasoning. In Flach's words 'How does one know that SA was lost? Because the human responded inappropriately. Why did the human respond inappropriately? Because SA was lost.' Flach's (1995) criticism is aimed at the existence of SA as a concept distinct from cognitive processes, showing how circular reasoning about the existence of SA as a defined product or process has led to flawed assumptions. Nevertheless, Endsley (2004) continues to make the assumption that SA exists, albeit through implication of a set of processes that combine to create the appearance of a phenomenon that looks like something which we can call SA. The circular reasoning problem undermines the basic assumption of Endsley's theory that SA supports performance. However, it will be recalled that Baars' global workspace theory states that there is a process of awareness and that this awareness is an executive function that makes all other unified cognition possible. Therefore, Endsley's theory may ultimately be found to be generally correct to the extent that SA may be found to be the cause of variance in performance, but whether the three levels of SA are cognitive processes rather than for example cognitive strategies, remains to be seen.

The majority of Endsley's (1995) theory of SA indicates the involvement of cognitive processes, as can be seen in Figure 1. Decision-making, action, goals, long-term memory, automaticity, and abilities, all involve aspects of cognition. However, the core of the model describes three distinct levels of SA (products), but there is also a linear process of perception supporting comprehension which then supports projection. Endsley's theory also includes a feedback process from the outcome of performance of action and from the new state of the environment. Endsley (2000b) argues that because pilots tend to show patterns of perception, then comprehension, and then projection, corresponding with increasing levels of experience then this pattern must describe the way that SA is acquired and maintained. It has been argued by McGuinness & Foy (2000) that we should interpret Endsley's levels more loosely as a complex network of parallel or interdependent processes[1]. The cognitive factors supporting SA have received insufficient attention in the literature, although there are indications that factors such as long-term working memory (LT-WM; Snow & Reising, 2000) and both fluid and crystallised intelligence (Gugerty *et al*, 2004) may be involved in mediating awareness. However, LT-WM and intelligence may simply be correlating with performance and therefore only indirectly related with the construct of SA.

Cognitive theories relating to SA

There are existing cognitive theories which relate to Endsley's model. Firstly, Neisser's (1976) action-perception theory of cognition also proposes a cyclic process of perceiving the environment, then comparison with the content of memory, and then active exploration from which further perception occurs. Neisser's theory gives support to Endsley's theory to the extent that the environment is perceived, the information from the environment is compared with memory content, and the performance of actions results in a new relationship with the environment which provides new information to be perceived. Therefore the performance of action controls perception and memory as much as perception and memory control the performance of action. This claim is in accord with Baars' global workspace theory (as described on page 7), and Endsley's theory, in that it would argue SA can be a

---

1 The existence of either linear, as implied by Endsley's theory, or parallel cognitive factors is approached in Chapter 2.

cause of itself. This may begin to explain the elusive nature of SA and the difficulty in defining it; SA may be an emergent property of cognition but one which nevertheless contributes to cognitive processing. Neisser's (1976) cognitive cycle permits expectations which can guide anticipatory action, thus leading to an emergent property during complex dynamic tasks; cognition is a cause of itself as a complex system and this produces a higher order effect of cognition being 'aware of itself' across time and projecting ahead in time. Such processes do not readily lend themselves to reduction to sub-processes.

Neisser's (1976) theory of cognition tends to support Endsley's account of SA as a process. However, Neisser does not support the product of SA as existing at three levels. Endsley (2004, p. 317) argues that the three levels of processing, perception, comprehension, and projection, also '...represent ascending degrees of SA'. However, there is little evidence to support the claim that perception, comprehension, and projection are separate processes that relate to ascending degrees of SA, other than Endsley and Robertson (2000) with its attendant problems of circular reasoning.

Another way that SA has been described is from the viewpoint of the Russian theory of activity (Bedny & Meister, 1997). This holds that the way in which activity occurs is a representation of the '...future results of activity, that include logically ordered systems of actions that permit one to achieve conscious goals' (Bedny & Meister, 1999, p. 64). This theory, as described by Bedny and Meister (1997) is therefore arguing that activity itself is structured according to internalised goals within a given situation. It is a 'goal-directed self-regulated system' (Bedny & Meister, 1999, p. 64). Activity is theorised to have three components or stages: orientational, executive, and evaluative.

The *orientational* concept is that a subjective model of reality is used to derive representations. Hence orientation is where the person is directed toward some aspects of the world for which they have an internal representation. The *executive* concept is that the situation can be transformed by the person by actively moving towards a goal. This process includes decision making and action. The *evaluative* concept is that the effects of the executive process in the world are perceived and adapted to. This theory is similar to Neisser's and Endsley's theories to the extent that there is a cyclic process of adaptation to the world. The theory of activity thus argues that actions occur in accordance with the rules required to achieve goals. This means that the goals of a task define the action and constrain events to certain variables aligned with the goals. Therefore SA could be said to be the extent to which behaviour is following the patterns required in order to attain task goals. It could also be said to be the extent to which (dynamic) task goals are changing, and hence the extent to which behaviour will also change (and need to change) in order to adapt to the new goal. This is the opposite of Gibson's (1979) ecological action-perception approach to cognitive processing, arguing that there is some a-priori mental goal structure for a task which provides 'affordances' for engagement with the task.

Gibson's (1979) ecological theory tends to be philosophical rather than psychological and does not present enough structure for it to explain SA, other than to support the claim that SA emerges from the interaction between an operator and a task. That is, like the theory of activity the direct perception approach argues that cognitive processing is dependent upon properties of a task and its environment. For a good performance the task and cognitive processes have a high degree of correspondence. Gibson's concept of affordance is used to explain how properties of the environment provide facilities for automatic action when schema exist that can make use of those properties in the environment. For example, a chair has properties that make it available for sitting on. This illustrates that automatic action is not so much processed cognitively as initiated from pattern-matching with the environment, what Gibson termed direct-perception. When there is a high degree of correspondence between a person's repertoire (e.g., very wide range of practice flying a passenger jet) and the task environment (e.g., flight deck of an aircraft that the pilot is familiar with), then practised routines of recognition-then-action are 'afforded'.

Smith and Hancock (1995) similarly explain the acquisition of SA as being dependent on the interaction between the agent and the environment; they are both dynamic parts of the situation. Hence their definition of SA as adaptive externally directed consciousness. This view of SA, concerning the interdependent nature of the task and human cognitive functioning, recognises the importance of Gibson's (1979) ecological approach. Patrick and James' (2004) position on SA also falls under the ecological perspective, restating Gibsons (1979) and Neisser's (1976) view that cognitive processes are an artefact of the interaction between the task-environment and the person. Patrick and James (2004, p. 61) argue that '...uncertainty still exists concerning the status of SA as a construct and whether it should be conceptualized as psychological in nature'. They also argue that if SA is accepted as a psychological phenomenon then it should also be conceptualised as a property of a task situation and its context. This extends Gibson's (1979) ecological theory by taking account of the role of context in determining automatic decisions and actions. However, it is difficult to progress further with psychological theory when a construct is considered to reside in a distributed form between a task and human cognition. None of the ecological theories above propose a clear pattern of known or hypothesised cognitive processes related with the acquisition and maintenance of SA. Working memory, long-term working memory, and intelligence have been found to be related to SA, as described above, but so have skill and experience – both of which are task dependent factors.

Cognitive streaming

A cognitive streaming account of SA, which sees SA as a working memory skill rather than a consequence of working memory capacity is proposed by Banbury, Croft, Macken and Jones (2004). They argue that the process of SA acquisition is based on the skilful probablistic calculation of information serially in working memory, particularly episodic working memory. The process of cognitive streaming is compatible with dynamic and cyclical models of SA described above. The cognitive streaming account of SA is focused particularly on the working memory processing that supports anticipation. The theory is based on the same principle as Endsley (1995) that the ontological function of SA is anticipation: A '...mechanism for projecting future states of the system based on its current state and an understanding of its dynamics' (Endsley, 1995, p. 44).

The main processing concept in cognitive streaming is of transitional probability. Transitional probability is '...the likelihood that certain types of event will occur following the occurrence of other events' (Banbury *et al*, 2004, p. 117). There is a transitional probability of an event occurring given the presence of precursors in the environment. Furthermore, cognitive streaming does not recognise short-term memory (STM) as a store. Instead it is conceptualised as a form of skilled behaviour in which transitional probabilities are used in order to predict related events. The transitional probabilities are provided by a calculation of information in memory which is recalled on perceiving related stimuli in the environment: a '...crucial mechanism for how humans become attuned to an uncertain world is their ability to learn about, and exploit, the statistical regularities of events that occur in the environment' (Banbury *et al*, 2004, p. 124). Transitional probabilities occur between events and across time. When matches are made between memory and stimuli in the world uncertainty reduces (in relation that event) and other events can be monitored. This is why STM is considered to be a skill rather than a memory capacity.

Cognitive streaming is based on the object-oriented episodic record (O-OER; Jones, Beaman & Macken, 1996). The O-OER is conceptualised as a 'blackboard' onto which information about properties of objects, and their relationships with other objects, is recorded. From this record a calculation can be made of the probability of occurrence of an event. As the cognitive stream is updated the probabilities adjust so that attention is redirected in the dynamically changing environment. The cognitive streaming account came from the claim that the serial nature of episodic memory carries a pattern of information onto which objects in memory can be connected. This provides a timeline with branches of increasing and decreasing probabilities. SA is being aware of the states of those objects and their probabilities. There are multiple cognitive streams, each for a relevant set of events and objects. This theory, like the global workspace theory, argues that a phenomenon arises in which the content of specific cognitive processes can be shared between those different processes and areas of perception.

Cognitive streaming, in contrast with other theories of SA, offers an explanation which complements the concept of working memory processing. Its basis in the O-OER model of working memory gives an alterative view on the way in which dynamic information is structured probabilistically. The theory is compatible with Baars' global workspace theory in that the processing for awareness and conscious awareness is considered under both theories to be a specific skill, one which supports other inter-related skills such as visual-spatial or serial-phonological processing. However, the cognitive streaming account of SA does not address the importance of the task context and the arguments described above in which SA is considered in varying degrees to be a property of a task as well as a property of cognition.

Alternative perspectives

There continues to be a problem of irreducibility with all the theories described so far: Baar's global workspace theory; Endsley's SA theory; cognitive streaming; Neisser's action-perception-cognition theory; Gibson's theory of ecological cognition; and the cognitive streaming account of SA. Dekker and Lützhöft (2004) take an alternative view on the source of this irreducibility. They argue that instead of a distinction between task and cognition we should view the distinction (if one exists) between mind and matter and their correspondence, as the source of the phenomenon that we call SA. That is, cognition during activity is embedded in the physical world. The physical world is indivisible, extending from the person who carries the conventionally understood 'cognitive processes' to the wider environment. Mind is the content of the information space that arises from the physical environment and human cognitive structures. The duality then is not between task and cognition but between empirical reality (physical matter) and information states of that matter. Correspondence between these factors is SA. This is to argue that SA is not a property of the task or cognition or the world but the extent to which information about the world (wherever it exists) is empirically correct. This casts the idea of irreducibility in a different context, one that edges toward the esoteric idea that awareness is a property of the whole physical world. This assists in illustrating that the main problem in SA theory is the lack of a clear psychological definition on which to focus.

However, the idea that SA may be a property of both the person and the task environment does open the investigation of SA to the measurement of cognitive performance within measurable and controlled tasks and environments. Flach, Mulder and van Paassen (2004) argue that both cognition and task environments are highly complex factors that require equal attention; the definition of a situation is itself embedded in the wider environment and brings a rich array of external organising resources (Flach, Mulder & Brooks, 2004). Patrick and James (2004) go further to argue that SA is situation specific, implying that there is no 'SA component' of cognition, it is to the situation that we should look first for theoretical development concerning SA. None of the theoretical positions described above provides a strong argument for the existence of SA as a causal factor in performance or as a product of cognition. The clearest evidence so far of the existence of SA is Baars' (1988; 2002) global workspace theory but that merely concludes that awareness is a transient emergent state that both supports and is supported by cognitive processes. Its only true distinct nature is in its manifestation as subjective consciousness. From what we currently know SA is simply '...a label for a variety of cognitive processing activities that are critical to dynamic, event-driven, and multitask fields of practice' (Sarter & Woods, 1995, p.16). Therefore, although this thesis will continue to refer to SA as though it does exist, the strong caveat is added that this is a shorthand for functional consciousness, adaptive externally directed consciousness, and other characterisations of 'being aware of a situation'. Nevertheless, Baars has shown that there are cognitive neuropsychological variables associated with states that appear to be 'awareness' and this provides strong grounds for continuing to explore the subject of cognition in SA.

This section focused on some of the fundamental assumptions about the existence of SA and how theory relates to these assumptions. Many of the arguments in the 1995(1) special edition of the journal Human Factors on SA, and in Banbury and Tremblay (2004), revolve around these fundamental points. Beyond this little else has been theorised, or should be theorised, given the lack of a clear definition, the nebulous nature of the construct, and the resulting failure to find clear relationships with robust cognitive theories.

The focus in this thesis will be on cognitive factors relating to the performance of a specific task rather than an attempt to explore the nebulous concept of SA itself as a generic process or product. This follows Patrick and James' (2004) argument that definition of a task is key to understanding SA. It does not preclude the null hypothesis of Dekker and Lützhöft (2004) and Flach *et al* (2004) that SA may be a false abstraction from an irreducible interaction between mind and the environment. An appropriate task for this thesis is driving hazard perception which has been argued by Horswill and McKenna (2004) to be SA during driving. This task has a record of previous research, and by using hazard perception tests or driving simulation it is definable and measurable with clarity. Hazard perception research will be discussed in the following section.

In terms of theories of SA, this thesis aims to test whether cognitive processes are related to product and process theories such as Endsley's model of SA, ecological theories and cognitive streaming. Specific cognitive theories (e.g, working memory, Baddeley & Hitch, 1974) will be introduced at the point in the thesis at which they become relevant to SA and driving hazard perception.

Anticipation and Hazard Perception

Driving hazard perception is described by Horswill and McKenna (2004) as SA for hazardous situations in the road environment, hence measures of hazard perception are measures of safety-related SA during driving. Hazard perception, like SA, results in the capacity to anticipate hazards. Therefore, skilled anticipation and hazard perception research, and hazard perception methodology, will now be discussed.

Anticipation

Anticipation is an important factor in the skilled performance of dynamic tasks; accurate anticipatory decisions and actions increase performance significantly over that of a passive perceiver of information. Endsley's (1995) model describes behaviours associated with SA anticipation. The three-level product (and arguably also process) is of perception supporting comprehension which enables a projection of events in space and time. This model describes a framework of behaviours leading ultimately to anticipation, and its use extends from an individual to team performance within a wide range of complex dynamic situations. The use of Endsley's (1995) model to provide assessments of SA may under some of the circumstances discussed in this thesis provide a distorted measure and hence lead to a distorted perception of the risk of failure in critical tasks.

The investigation of skilled anticipation (irrespective of conscious awareness) and how it relates to ecological and process oriented approaches is useful in understanding the cognitive factors involved in the acquisition and maintenance of

SA. Early research indicated that anticipation is related to two types of behaviour; apprehensiveness and searching (Lindner, 1938). Lindner also found that in some cases the resulting advanced action was not preceded by apprehensiveness or searching; there appeared to be a route directly to action. Engstroem, Kelso and Holroyd (1996) argue that reaction and anticipation are two parts of one pattern recognition system. In an experiment where participants tried to match finger movement to a metronome in different patterns participants displayed a transition from reaction to anticipation and from anticipation to reaction. These transitions imply that dual processes may support anticipation since anticipation and reaction are semantic attempts to characterise skilled anticipatory action.

Ericsson and Kintsch's (1995) theory of long term working memory (LT-WM) indicates that skilled anticipation performance is characterized by task related long-term memory structures made immediately available to short-term memory retrieval cues. LT-WM develops when its processes are required in order to subsume mental models and routines, implying a system which is involved in skilled performance (Chase & Ericsson, 1982). It follows that the management of changing goals and context-stimuli associations could be mediated by a process such as LT-WM. It has been argued by Altmann and Trafton (2002) that goal-directed cognition does not require support from specialized processes such as the goal stack that has been a standard model (Newell, 1990). The goal stack model states that the goals and sub-goals of a task drive the behaviour in a sequential process. Altmann and Trafton (2002) argue that this is not necessary and that a simpler explanation also describes goal-directed behaviour; the overall goal of the task is sufficient alone to invoke activation and associative priming related to the lowest level task elements. For the purposes of the present thesis this theory describes firstly how goal-directed behaviour, for which SA is necessary, can emerge from both perceived and memory information content and lead to rapid anticipatory action.

It is likely also that mental simulation is involved in anticipation, reflecting Endsley's (1995) model of SA as a process combining perception, comprehension, and projection – projection implies the development of a model of future events derived from a model of current and past events. However, this simulation may be automatic or controlled. Neuropsychological research by Chaminade, Meary, Orliaguet and Decety (2001) found that when a moving dot was used to show motion similar to mechanical movements, movements showing pointing action, and movements simulating handwriting patterns, anticipation of the mechanical motion activated the right intraparietal suculus and the left prefrontal cortex (relating to spatial awareness and prediction). However, the left frontal operculum and superior parietal lobule (motor areas) were activated when handwriting was being anticipated. This is interpreted by Chaminade *et al* as showing that anticipation of motion in skilled tasks involves a degree of simulation in the motor areas; there appears to be a motor calculation even when there is no action. This simulation may be a way in which experts can test a hypothesis about future action using their motor experience without the need for trial and error action, and such a process could be an important part of skilled spatial anticipation. This accords with Elsner and Hommel's (2001) two-phase model of action control, in which during early skill acquisition an association is learned between contingencies of motor patterns and spatial movements. Only after this first level of skill has been achieved can the associations be used for higher level learning and goal-directed, and therefore contextually-driven, associations.

It has also been suggested that context is important in anticipation. Kay and Poulton (1951) used a complex motor learning task to show how anticipation about the context in which recall will take place affects the way in which the participant integrates information about the task. Peterson, Brewer and Bertucco (1963) found that early in a paired-associate word task participants had been using the context of the previous trial to guide probabilistic responses. This suggests that context is fundamental to anticipation, a finding that supports task-related and direct perception perspectives on SA. Cavallo, Brun Dei, Laya, and Neboit (1988) found that novice drivers use a rudimentary control process while negotiating curves and experienced drivers use a more sophisticated anticipation process based on general knowledge about curves. Federico (1995) found that expert pilots depended more on contextual information in recognition of situations than did novices. These examples suggest that anticipation is a dual-level process. Expert anticipation seems to involve using contextual knowledge in order to predict and prepare for task-environment changes. This involvement of context is consistent with the ecological direct-perception approach.

Anticipation is a function of SA; a synonym for Endsley's (1995) projection level of SA. Therefore it would be useful to test a measure of anticipation for correlation with measures of SA. Although anticipation can be considered a performance variable, in the case of SA where anticipation is level three SA, a test of anticipation would also be a test of this level of SA. Furthermore, because anticipation is a macro-level product of SA it follows that a measure of anticipation should correlate with overall SA. If this is correct then cognitive factors underlying anticipation skills should resemble the same structure as the behaviours observed during differing levels of SA (e.g., Endsley, 1995). However, it is noted that although anticipation has been studied for some time from a cognitive perspective, so far there is little indication of such a pattern emerging. Skilled anticipation can be argued to be one product of SA, as is consciously accessible awareness. This presents a hypothesis that will be approached throughout this thesis; that there is a two-level process involved in the acquisition and maintenance of SA comprising automatic perception-action level awareness and controlled recognition-comprehension level awareness.

Hazard perception

There are no specific theories of driving hazard perception, or the cognitive processes which support it. However, driving hazard perception as an application of SA is related to performance, as SA is related to performance in aviation. While accident involvement is not related to overall driving ability (Horswill & McKenna, 2004), accident involvement is negatively related to hazard perception similar to the relationship observed between SA and performance in other SA critical tasks (e.g Hörmann, Soll, Banbury & Dudfield, 2003; Rowe & McKenna, 2001). The hazard perception test (HPT; Pelz & Krupat, 1974; McKenna & Crick, 1991, 1994) has been found to correlate negatively with accident involvement (Quimby & Watts, 1981; Quimby *et al*, 1986) and is related to anticipation performance (e.g Rowe & McKenna, 2001). In particular, McKenna and Crick (1991) used a non-control driving video based HPT to isolate the hazard monitoring part of the driving task. The HPT was based on recognition of road traffic hazards. The HPT measure was found to correlate negatively with accident involvement. Furthermore, McKenna and Crick (1991) found that the HPT discriminates between novices and experts and that training novices in hazard perception skills leads to an increase in hazard perception test (HPT) scores to a level similar to that of experts. Hence the HPT anticipation measure appears to be similar to other measures of SA (e.g., Endsley, 1995) which correlate with expertise and experience. There is further support for the claim that the HPT is a measure of SA from experiments on skilled anticipation in sport. Rowe and McKenna (2001) measured the task of predicting an opponent's actions in tennis performance. They found that this test also discriminated between novice and expert players in terms of their anticipation ability.

There is evidence that expert drivers are quicker to recognise relevant road situations (Groeger, 2000). This is in agreement with laboratory findings concerning expertise generally (e.g. McKenna & Farrand, 1999), and evidence such as expert chess players showing better memory for chess positions (Chase & Simon, 1973). In both cases it was argued that the effect is due to chunking of information content and automaticity. The implication is that due to automaticity cognitive processing is easier for experts, and they are less susceptible to interference from concurrent tasks due to less reliance on working memory (WM). Therefore McKenna and Farrand (1999) predicted that novice drivers who lack both expertise and experience should find hazard perception less automatic and more effortful. Hence novice drivers will experience more WM interference than experts on the hazard perception task while performing concurrent cognitive tasks. The results of McKenna and Farrand's (1999) experiment showed the opposite effect: novice drivers experienced less interference than expert drivers. This seems to contradict the generally held view that expertise is associated with automaticity. There is therefore a suggestion that in the case of hazard perception or SA the effects of automaticity may not increase linearly with expertise but that the dynamic combination of controlled and automatic processing subsumes more information content as expertise or experience increase.

Investigations of SA in pilots indicate a possible reason for the failure to find a clear link between automaticity and expertise or experience. Sohn and Doane (2004) tested novice and expert pilots on the extent to which each level of expertise depends on LT-WM and WM capacity. They found that for experts LT-WM capacity appears to be more important than WM, while the inverse is the case in novice performance. LT-WM (Ericsson & Kintsch, 1995) is a theory which integrates WM and LTM into a combined and partly controlled process. McKenna and Farrand's (1999) results make sense in this context. It appears that LT-WM processing for the maintenance of SA may place greater demands on WM memory than during simple tasks because the use of LT-WM in complex dynamic tasks is likely to require attention, action, and LTM to be mediated by WM. This would be in addition to the use of WM for controlled processing of novel events. Therefore, experts who use LT-WM processing in order to maintain SA appear to depend more on WM than novices who have less LTM resources to manage, less deliberate action to control, and therefore a lower range of event related processing to manage. For the novice in a complex dynamic task, controlled processing may appear more effortful but this may be for the following reason: without the content of expertise and experience there is simply less cognitive work for novices to do but due to that lack of content novices find the task more effortful to manage successfully. Furthermore, SA processes tend to involve a state transition as an operator enters what is known in many domains as the 'loop', the loop being a concept akin to Neisser's (1976) perceptual cycle applied at multiple iterations over time. That is, to 'be in the loop' is to have sufficient awareness to perform unaided basic control of a task. Acquisition of SA after initiating a task takes time while all the component processes are sequentially initiated and then monitored in succession. Therefore, there is likely to be a period of transition until the person enters the loop at which time sufficient SA has been acquired for continuous and unaided basic control of the dynamic complex task.

Sohn and Doane's (2004) and McKenna and Farrand's (1999) findings appear incompatible. In the former case experts depend more on LT-WM capacity and less on WM. In the latter case experts find hazard perception more effortful, an effect usually associated with WM. However, if we consider that LT-WM is likely to be more effortful due to its reliance on the integration of several processes, including WM, then the above studies support the proposition that the cognitive processes involved in the maintenance of SA are more effortful and more complex than either controlled or automatic processing alone.

This argument that SA depends on effortful LT-WM processing makes sense from a functional perspective. The maintenance of SA cannot be entirely automatic but would benefit from appropriate use of either WM or automatic LTM routines in order to process information related to either novel or common situations respectively. That is, SA involves attending to relevant events in the environment, comparison with LTM representations, conscious decision-making and automatic responding, attention to one's own actions initiated from probabilistic representations of future events and monitoring the consequences of that action. If SA is 'externally directed adaptive consciousness' as Smith and Hancock (1995) argue, then it therefore makes sense that experts' SA will depend on resource demanding, executive mediation, of multiple processes which are described by LT-WM. This matter of the cognitive factors in experts' SA will be discussed further in later chapters in relation to other cognitive theories (e.g. Baddeley, 2000; four component model of working memory) and will be addressed in Experiments 5 and 6.

Measures

This section firstly describes standard measures of SA that are currently used in applied human factors. Secondly, measures of driving hazard perception will be discussed, including the method used for developing the modified hazard perception test (MHPT), which will subsequently be used in two formats to be described later, in the experiments in this thesis.

SA Measures

In terms of individual operator SA existing SA assessment techniques fall into five categories (Stanton *et al*, 2005). These categories are SA requirements analysis, freeze probes, real-time probes, self-rating, and observer-rating techniques. SA requirements analysis is used in order to specify what factors are associated with SA in a task. Freeze-probe techniques involve pausing and hiding simulation of the task, queries are then presented to the operator from which an assessment is made of the operator's SA during the task.

One measure which falls into this category is SAGAT (Endsley, 1995) which freezes the simulator interface during a complex dynamic task and queries the operator about aspects of the task in terms of perception of current information, comprehension of the meaning of that information, and projection to future system states. The queries are developed from an SA requirements analysis and are presented at random times during the task. The responses are required in terms of the objective criteria by means of multiple choices. An example query is 'what is your altitude'. SAGAT and variants have been used in civil and military aviation and air traffic control, and in control room simulations.

There is a paucity of robust findings surrounding freeze probes such as SAGAT, despite these measures being 'intrusive to primary task performance' (Stanton *et al*, 2005, p. 216). Given the effects that interruption usually has on tasks (MacFarlane and Latorella, 2002) it is interesting that despite investigations no effects of interruption appear to have been found in relation to the use of SAGAT. This is of particular theoretical and applied relevance since SAGAT has been used in the development of Endsley's (1995) theory of SA, and it has been a widely cited measure used in assessments of operators SA in critical tasks. However, SAGAT has the potential benefit of being based on objective criteria and objective measurement, although the SA construct upon which it is based leaves it open to criticism that it is based on circular reasoning.

Real-time probe techniques are similar to freeze probe techniques in that probe queries are presented during a simulated task. However, real-time probes do not freeze the simulation or hide the scenario. This is argued by Durso, Hackworth, Truitt, Crutchfield and Manning (1998) to reduce the effects of interruption. It also makes available to the operator memory located in the task, in the same way that this would be available in the real task. An example of this form of measure is the situation present assessment method (SPAM; Durso *et al*, 1998) in which a telephone is used to query the operator, with both the response latency and the time taken to answer the telephone being used as measures of SA. This method which requires the primary task and the SPAM task to be performed simultaneously has potential to interfere with the primary task, albeit in a different manner to the finite interruption presented by SAGAT. SPAM and similar techniques such as *SA* for *SHA*PE (SASHA, a development of the European ATM SHAPE project; Jeannot, Kelly, & Thompson, 2003) also have the benefit of being based on objective criteria.

Self-rating techniques are attempts to assess operators' subjective responses to rating scales either requested during a simulation or after its completion. An example is Taylor's (1990) situational awareness rating technique (SART). As a rating scale SART has the potential to measure multiple dimensions. Ten have been compiled into the measure. Clearly, memory effects are likely to be involved in the measure and also it has been noted that if an operator has a lack of SA then there will be a bias in the responses. That is, the level of SA will be associated with the degree to which an operator has the ability to respond to any given query. However, such techniques offer useful comparisons with other measures as overall indicators of SA from the operator's subjective position.

Observer rating techniques depend significantly on subject matter experts as developers of rating scales and as assessors of operators while performing tasks. This method has been criticised (Endsley, 1995) due to the degree of construct validity that can be attained in what is an assessment of cognitive process by an observer. That is, it is a form of observational protocol and is therefore potentially less robust as a measure than a verbal protocol analysis which is elicited directly from the operator.

The various measures of SA have been developed due to specific requirements of applied human factors. One of the main factors is pragmatism due to the cost of implementation. For this reason the measures tend to have weaknesses in terms of cognitive theory. This is in addition to the difficulties of defining SA which introduces significant doubts about precisely what the techniques are attempting to measure. The measures are not based on a task analysis as such, but on an SA requirements analysis which assumes the existence of SA and its necessity in performance, thereby admitting another source of circular reasoning. This illustrates the often tenuous link between definitions of SA and the outputs of assessments. The first four kinds of SA assessment methods described above depend on their degree of validity in terms of the meaning of outputs.

The measures that have been described above will now be compared. The first SA assessment method described above, the freeze technique, is in multiple choice format and presents queries such as 'what will be your position in ten seconds' or 'which aircraft is to your left', to which a multiple choice response is required. This method, of which SAGAT is an example, demands that the operator must consider content of his memory and then respond explicitly. Endsley (1990) claim criterion validity of SAGAT based on findings that military pilots showing a SAGAT score for a target were more likely to eliminate the target in the simulation. A criticism is that the increased kill rate is due to SAGAT reorienting the pilot to certain kinds of stimuli (McGowan & Banbury, 2004). Other criticisms are firstly that queries are developed from an SA requirements analysis rather than from a task analysis. That is, while the queries are structured and appear to require simple statements about the state of the situation they are based on a subject matter expert's interpretation of the task in terms of a theory of SA rather than an objective set of criteria. Secondly, since the operator must access working memory and long-term memory before response to events which have occurred before the freeze then memory effects are likely to be present in the measure.

The real-time probe SPAM technique is less dependent memory although it still requires that the operator must verbally respond to queries that are presented verbally. However, since the queries are presented in a format which is a normal part of the task (e.g. answering a telephone call to answer questions about the task) this permits the operator to use existing strategies for managing interruptions. This is Durso *et al*'s (1998) main claim for the benefit of SPAM over freeze techniques. Durso, Bleckley and Dattel (2006) supports this with evidence showing that prediction performance using a battery of cognitive tests increased only when adding the SPAM method. S far, no interruption effects have been found for freeze techniques or SPAM techniques and yet this is contrary to the experimental findings regarding task interruption (Miyata & Norman, 1986; McFarlane, 2002) which show notable deleterious effects on cognitive and task performance. The apparent lack of negative effects due to interruption stands in clear contrast to this research. This issue will be explore in Chapter 3.

Self-rating methods such as SART can also be critiqued from the perspective of task interruption. The operator is required to analyse and the situation and translate it onto a rating scale. Furthermore, self-rating that is delayed until after the task permits confounding by memory effects, perhaps to an even greater extent than freeze techniques where a working memory trace is still likely to exist. However, rating scales have the advantage that they readily permit the assessment of multiple dimensions. Such data may then be analysed in order to control for the effects of verbal analysis on the measure.

Therefore, when comparing the above techniques it is clear that freeze techniques appear to be the most objective form of test. This is because freeze techniques depend upon multiple-choice queries derived from an objective analysis of criteria, albeit in the form of SA requirements analyses. However, the freezing of an otherwise dynamic simulation to interrupt with queries raises the problem of interruption of the normal time sequence of the main task, and hence the possibility of affecting SA. In contrast, SPAM attempts to address the problem of interruption by permitting real-time responding, although a degree of interruption is still likely. Self-rating techniques avoid the problem of interruption but admit the likelihood of memory effects, although the use of multiple dimensions raises the possibility of controlling for those memory effects.

Hazard Perception Measures

The basis of the hazard perception method is that it presents video of road scenes filmed from the driver's position of a moving vehicle. The scenes include road hazards to which the participant responds with a button press when a hazard is perceived to be developing. With a button press the response can be utilised either categorically as a hit or a miss during the time period, or as an on-screen spatial location using a computer mouse response in relation to visual spatial criteria representing the location of the hazard. Alternatively, the time of the response can be subtracted from the moment of maximum hazard to give negative latency scale data representing anticipation. This feature makes the hazard perception test ideally suited to assessing task performance in terms of anticipation and SA. McGowan (2002) combined the hit and miss recognition data with the anticipation scale to produce the modified hazard perception test (MHPT) which can provide the main dependent measures of both recognition (hit and miss data) and anticipation (time between the first spatially correct response and the later moment of maximum hazard) in this thesis.

There are two main ways in which hazard perception can be inferred. Firstly, recognition is simply perceiving a hazardous situation and reporting the hazard that has been recognised. Secondly, anticipation of hazards can be defined as the time interval before the moment of maximum hazard – maximum hazard being the moment at which the hazard begins to decrease in either severity, or in temporal or spatial proximity. Clearly, recognition is not time dependent. Anticipation however, is both time dependent and dependent on correct recognition. There is also a third measure which is not anticipation but recognition latency; the time taken to recognise a hazard that has become present visually. This measure is in effect negative anticipation. For reasons which will be discussed in the present section it is beyond the scope of this thesis to use recognition latency as a measure, the main reason being the difficult in defining what is and what is not a hazard after the moment of maximum hazard, which tends to be exceeded by many of the hazard recognition latency responses.

Early versions of the driving HPT presented a video of forward views taken from a moving car, and drivers responded by moving a lever to indicate the level of hazard portrayed in the scenes (Pelz & Krupat, 1974). There were only six hazards in this test. The score was calculated from the difference between the time at which the drivers moved the lever to the high hazard level and the time of the hazard, producing an anticipation score. This rudimentary apparatus nevertheless appears to have discriminated between histories of accident involvement and traffic convictions. Pelz and Krupat found that drivers who had no accidents or convictions moved the lever 0.5 seconds earlier than those who had a history of accidents but no convictions, and 1.2 seconds earlier than those who had a history of both accidents and convictions. These results thereby suggest that safe driving ability is associated with skilled anticipation in the form of hazard perception.

Pelz and Krupat's method was refined by Quimby and Watts (1979) and Watts and Quimby (1979), who used similar apparatus but for added ecological validity the drivers viewed the video from within a stationary car. They found a negative correlation of $r = -0.28$ between the drivers hazard perception scores and their frequency of accident involvement, further supporting the link between safe driving ability and hazard perception. It was also found that response times on the hazard perception test were positively correlated with experience (Quimby & Watts, 1981). Quimby, Maycock, Carter, Dixon, and Wall (1986) found that hazard perception test performance is related to driving record and accident involvement.

Subsequently, McKenna and Crick (1991, 1993) developed an applied HPT which differentiated between novice and expert drivers. McKenna and Crick (1994) presented video scenes of dynamic events to which participants respond when they first see the hazard developing. This method has now been implemented in the UK and Australian driving theory tests. The concept is familiar to many UK university students who have undertaken the test since late 2002. The test is based on the same principle as the earlier tests above but it is less realistic, using only a video screen, a button press response, and a computer to record the responses. Again, this HPT showed a negative correlation with HPT scores ($r = -0.11$; Horswill & McKenna, 2004). However, there is also a potential issue of validity with all the aforementioned tests since it is possible to achieve higher HPT scores by increasing the number of responses.

Recently some related research has focused on the role of skilled anticipation in sports tasks (Rowe & McKenna, 2001). A video of tennis rallies was presented to participants from the perspective of a player. The participants were set the task of anticipating the onset of the opponent's final stroke. Their experiment showed that this method, like the hazard perception test, was also able to discriminate between experts and experienced players, and between levels of expertise and levels of experience. The results from the separate tasks of sports and driving indicate that the method of using video to present a monitoring task, to which participants respond by anticipating events, is likely to be a valid method of measuring anticipation levels.

However, the hazard perception test in the formats described above may not be appropriate for measuring the full range of hazard perception exhibited by drivers. The hazard perception test typically relies on a button press response to visually presented stimuli. This differs from the driving task a driver would normally perform in relation to hazard perception. The potential exists for the effects of expertise not to be accurately represented by hazard perception tests, since responding using the button pressing format may not sample skilled automatic responding that is specific to the way that a car is controlled. Alternative methods of hazard perception responding would clearly include the use of simulated vehicle controls such as steering and braking. However, the use of full simulation is complex and costly to develop in that the simulated hazards must be embedded as adaptive data objects in equally dynamic contextual scenarios and then linked to data recording of multiple response formats, such as steering and braking. Other considerations such as combinations of responses are also problematic. Such complex methods are beyond the scope of the present thesis.

Hazard perception tests could easily be adapted to non-control formats, whereby the participant would respond by pressing a brake pedal or turning a steering wheel that had no other effects than to record a response. However, due to the non-control format this method may be unrealistic in the same way as button pressing response, although it would appear to be more ecologically valid than a button press since it supports experienced drivers' automatic responding in terms of steering or braking in order to avoid a hazard. Despite this potential, for the purpose of the present thesis the button press method is the most appropriate format. Utilising the existing format will advance a small step from the foundations of the hazard perception research described above, as a basis for future development of more ecologically valid methods.

MHPT measure

McGowan (2002) combined the hit and miss recognition data with the anticipation scale to produce the modified hazard perception test (MHPT) which is to provide the main dependent measures in this thesis. Driving hazard perception test items are based on visually identifiable stimuli. The location of the visual stimuli relating to the hazard is compared with a computer mouse response location. If the computer mouse response is in the correct location and at the correct time then the participant is deemed to have recognised the hazard. In addition, an anticipation score can be calculated for the correct recognition, by subtracting the response time from the end moment of the hazard (its moment of maximum risk). It is anticipated that the mouse response data may also be used together with eye tracking data to determine the degree of correspondence between the mouse-tracking of events and eye movements.

A further alteration to the method was to use a wide range of video collected for the purpose of Police road patrols and then to select the hazardous events using a panel of expert drivers. This, along with the avoidance of staged events, enables a critical incident approach which in this context is a combination of a task analysis and an SA requirements analysis. This ensures that the hazards relate to road driving skill criteria. In addition, the pre-selected video footage was provided by experienced

Police road patrol drivers, whose policing duty while collecting the film was to actively monitor roads for hazardous events. This process provided footage that ranged from safe clear roads to highly hazardous situations, and included both static and dynamic items. It is proposed that the high internal consistency of the hazard perception based anticipation test (Cronbach's $\alpha = 0.72$; McGowan & Banbury, 2004) is due to the rigorous selection process by the Police drivers and the expert drivers. The items appear to be measuring the same construct. Due to the careful design of the instructions to the Police and expert drivers and the facilitation and monitoring of the expert drivers while selecting the hazards, it is concluded that the criterion and construct validity are also likely to be high. Hence, the scale is likely to be a measure of the hazard perception task as practised by the Police drivers.

Development of the MHPT

The analysis of the video was performed as a group. The six expert drivers were instructed to view the video as a group but as individuals they were to say 'stop' when they saw a hazard that would require some action to be taken if they were driving a car (Horswill & McKenna, 2004; this criterion was also included in the experimental instructions in Chapters 2, 3 and 4). The action may be any manoeuvre or pre-emptive manoeuvre in response to the hazard. They viewed the video which was projected in front of them at a large size. The video ran until one of the expert drivers identified what they thought was a hazard by saying 'stop'. The video was allowed to run for a few seconds so that the others could see what the identifying expert driver was referring to, and then the video was slowly reversed through the period and then forward again repeatedly while the group discussed whether to define the events as a hazard that would require some action to be taken if they were driving a car. The role of the researcher was to facilitate by reminding the expert drivers of this criterion, and to make a decision in arbitrary cases where the group could not decide. If the group appeared to have a strong consensus but with dissenters, then the decision to accept the hazard was taken. If the dissent was strong then the event was rejected as a hazard. In most cases the group reached a decision quickly and without dissent. The group appeared to have a significant knowledge of the subject and had specific criteria of their own by which they were assessing the hazards, probably deriving from their experience as advanced driving observers. Interestingly, these criteria corresponded closely with Horswill and McKenna's (2004) criterion of hazards that would require some action to be taken during real driving.

The expert drivers provided the earliest time from which the event is predictable, and its moment of maximum hazard, in addition to identifying the visual cue related to that hazard on the screen. This defines the parameters of the anticipation latency period t –t0, where 0 is the moment of maximum hazard and a high score represents earlier (i.e. increased) anticipation, and the spatial frame x1–x2 and y1–y2. Figure 2 shows the location of these hidden spatial criteria for a correct response. The vehicle directly ahead is at its moment of maximum hazard, after decelerating significantly during the preceding three seconds. This frame follows the item as it moves within the screen. If the participant had clicked within the spatial frame during those three seconds then an anticipation score would be calculated ahead of the moment of maximum hazard. In this case the vehicle that has decelerated, and which the participant appears to be approaching rapidly, will now begin to accelerate and the distance between it and the participant's position will begin to increase, therefore the hazard diminishes. Other parts of the scene may attract a response and these responses would not be included as an anticipation score because they did not meet the experts' criteria for hazards. The events in which there is no correct response are scored as zero, thereby reducing the mean anticipation score as though a correct response had fallen on the moment of maximum hazard. In Figure 2 for example, the white vehicle to the right of the correct target is a Police patrol car and this may have attracted an incorrect response due to its association with emergencies and danger. The same applies to any response outside the correct spatial frame. This process of spatial selection improves on the basic HPT by reducing the potential for raw response rates to affect HPT scores. This is supported by the finding that the anticipation scores do not correlate with the total number of mouse clicks (McGowan & Banbury, 2004), indicating that the method is not susceptible to changes in the overall response rate and that its resolution is likely to be high.

Internal consistency. The 120 hazards in the MHPT comprising the two counterbalanced video clips indicated a homogenous scale of AMHPT scores with high internal consistency (Cronbach's $\alpha = 0.63$). The frequency distribution of hazard durations indicates the possible influence of multiple factors on hazard length (Figure 6).

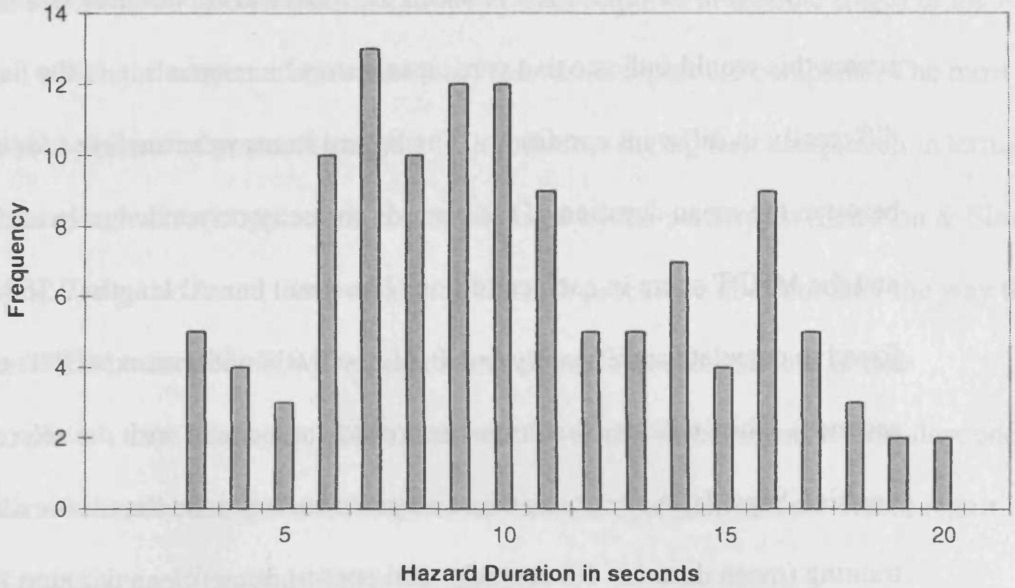

Figure 6. Hazard duration frequency indicating possible influence of multiple factors on hazard length.

Hazard duration and severity. The hazards were analysed in terms of the combination of severity and hazard duration but only for the original 103 hazards. A chi-square test (χ^2 (4) = 18.98, $p < 0.001$) confirms that the cells are significantly different from each other. Table 4 appears to show an interaction of hazard severity and hazard duration in terms of numbers of hazards; hazard severity seems to be negatively related to hazard duration. This was not expected because hazards should not differ in terms of severity (at the moment of maximum hazard) simply because of their different locations in time and space.

		Duration		
		Low	Medium	High
	Low	4.30	9.68	11.83
Severity	Medium	39.78	15.05	7.53
	High	7.53	3.23	1.08

Table 4. Chi-square frequencies (percentages) of combined duration and severity distribution.

If there is a correlation between the mean hazard duration and the MHPT scores this would indicate that participants may be responding to the hazards differently in different conditions. The hazard items were analysed for correlation between the mean duration of the hazards correctly responded to in each condition and the MHPT score in each condition. The mean hazard length (9.36 seconds) was found to correlate significantly ($r = 0.73$, $p < 0.01$) with mean MHPT score (5.53 seconds). This indicates that higher scores are associated with the selection of longer duration hazards. A t-test was then computed between the hazard durations pre-training (mean duration 8.6 seconds) and post-training (mean duration 11.29 seconds) conditions. This analysis showed a significant difference ($T(1) = 5.89$, $p < 0.01$) between the pre-training (mean 8.79 seconds) and post-training (mean 11.18) hazard durations. This indicates that participants are selecting hazards with longer lengths in the post training conditions than the pre-training conditions.

Verbal protocol analysis

The task analysis performed above is a form of verbal protocol analysis. This is because the expert drivers were instructed to describe what they would be doing if they were driving the car from which the film was recorded. This was facilitated by the researcher but not in real-time, since the video was paused and reversed when the group discussed the precise beginning and end of a hazard that they had identified. This task analysis was therefore performed as a verbal protocol analysis.

Verbal protocol analysis is a methodology for the use of verbal reports of mental representations from cognition (Ericsson, 2002). The method does not require any analysis of the mental representations by the participant who is simply asked to verbalize what is happening, since an instruction that emphasizes explanation or analysis of procedures will alter task performance (Ericsson & Simon, 1993). A task analysis provides the various operations that are required at discrete stages of the task, with which the sequence and contents of the verbal report are compared. The most valid use of verbal protocol analysis is in real-time during task completion in terms of the language of task operations commonly used by the participant (Ericsson & Simon, 1993). They also found that verbalizing the task procedure does not alter the way that the task is performed. However, there does appear to be an increase in task completion time (Ericsson, 2002). The validity of the verbal report is time dependent, with validity declining when the time between the operation and the verbal report exceeds five to 10 seconds (Ericsson, 2002).

As a form of verbal protocol analysis the expert driver panel achieved the general specification provided by Ericsson (2002) in that it did not require any analysis of the mental representations by the expert drivers who were simply asked to verbalize what was happening. That is, the criterion that was provided to the expert drivers enabled them to compare this with their observations and expertise at the task. They did not need to introspect because they were not actually performing the task. However, because they were not performing the task it could be argued that the process was not verbal protocol analysis. Instead, it could be argued to be a form of critical incident task analysis since the critical incident (the criterion provided by the researcher) need only be compared with the observed processes in the video. This may not be entirely true, because as a group the expert drivers did engage in a significant amount of discussion and analysis about their expert processes and how they related to the criterion and the events presented in the video. Other factors supporting the method as a form of verbal protocol analysis were that the process occurred in real-time, to the extent that a hazard was only analysed if one of the expert drivers initially identified it as such, thereby inviting the group to look in more detail and analyse the event against criteria. The initial recognition of the event as a hazard therefore meets the criteria for a real-time verbalisation of the process.

CHAPTER 2
EXPERTISE AND TRAINING

The two specific aims of this thesis are firstly to continue assessing the utility of hazard perception testing in the MHPT format, as a laboratory-based measure of SA. Secondly the MHPT will be used in order to explore cognitive factors involved in the acquisition and maintenance of SA. In Chapter 1 SA and hazard perception literature was described. Chapter 2 aims to begin to test predictions from both hazard perception and situation awareness evidence. Experiment 1 will investigate the measurement of SA using the MHPT, and Experiment 2 will then investigate SA theory.

The three kinds of selected definitions of SA in Chapter 1 were those which approached SA in terms of processes, those which related to its products, and those which were described as being a synthesis of a specific task and cognition. The latter definition was described as being an ecological approach. Later in the thesis this ecological approach to understanding SA will be examined in relation to ecological theories of cognition but at this stage the ecological approach refers in particular to use of the HPT method as an embedded measure of explicit and implicit task performance in order to infer situation awareness. The HPT embedded task-based measure appears to be close to being an objective assessment of SA because it appears to measure the driving sub-task of maintaining awareness in terms of both recognition and anticipation (Pelz & Fulpat, 1974; McKenna & Crick, 1991; Rowe & McKenna, 2001). This general hypothesis, Horswill and McKenna's (2004) argument that hazard perception is SA during driving, forms the context for the present chapter.

Experiment 1

The main aim of Experiment 1 will be to test the relationship between hazard perception performance and SA as proposed by Horswill and McKenna (2004). This can be achieved using the same approach as McKenna and Crick (1991) who tested the effects of expertise and hazard perception training, employing an HPT as a dependent measure. McKenna and Crick's approach can be adapted by training strategies for increasing SA instead of hazard perception, in order to assess the relationship between hazard perception and SA. A second aim of the first two experiments will be focussed on testing expertise and training effects on hazard perception (following McKenna & Crick, 1991) as a means to assessing the theorised relationship between hazard perception and SA (Horswill & McKenna, 2004), and calibrating an HPT for possible later investigation of cognitive factors. Experiment 1 will be a replication and further refinement of work carried out for the researcher's MSc dissertation (McGowan, 2002) where a modified HPT (MHPT; described on p. 29) was developed as a measure of SA. SA training materials were developed during those experiments and the MHPT was calibrated. The aim for Experiment 1 is to replicate and refine the manipulation of training in SA strategies while also testing the effects of expertise and experience on data derived from the MHPT measure. The first experiment will draw on the experience from the previous work in order to use appropriate sample sizes, in order to further calibrate the MHPT measure for use in later chapters where the aim is to manipulate cognitive factors.

As was discussed in Chapter 1, the focus in this thesis will be on cognitive factors and their relationship with a specific task (a task-focussed and ecological synthesis) rather than attempting to explore the nebulous concept of SA itself as a generic process or product. It is also stated in Chapter 1 that cognitive theory supports a range of predictions and measures, and that the specific task during which awareness arises must be definable and measurable with clarity. The definition of a situation is itself embedded in the wider environment and brings a rich array of external organising resources (Flach *et al*, 2004). This thesis is devoted to development of the methodology in terms of the task and its environment, and as such this means that despite the focus on the MHPT measure it is nevertheless very much an investigation of cognition from the ecological viewpoint. This approach is supported by Patrick and James (2004) view that defining and measuring situations (ie. Task environments) is essential to an understanding of SA, and is related to Dekker and Lützhöft's (2004) radical theory that SA is as much a product of the task and environment as it is a product of the person and her cognitive processes. However, it has been argued that what we conceptualise to be conscious awareness is simply a by-product (or epi-phenomenon; Banbury and Tremblay, 2004) of an organism's adaptive interaction with its environment. The difference between conscious awareness and non-conscious anticipation processes is that the conscious system has the ability to abstractly calculate anticipatory responses further ahead, based on LTM content as well as the current content of working memory. This by-product LT-WM fusion would be useful in order to analyse events, from both current and past representations, in order to anticipate further ahead than the current situation. However, as the literature outlined in Chapter 1 shows anticipation can occur before conscious awareness (e.g. Rowe & McKenna, 2001). Implicit awareness is inferred from behaviour which necessarily depends on the processing of some information about a situation; at some level the person must be aware of her environment whether or not she is able to explicitly describe the information content that she is processing. Explicit awareness is that where information about the situation and the action to be taken in respect of that situation is reportable by the person. Following Croft *et al* (2004) this thesis will not attempt to distinguish between explicit and implicit SA but will simply infer anticipatory action and conscious awareness from behaviour in relation to situations. The aim is to test an important hypothesis that implicit anticipatory action and explicit conscious awareness will be associated with differing patterns of recognition and anticipation scores. It is perhaps worth mentioning that during this thesis it became apparent that a full control simulation, from which samples of all control movements can be taken, is likely to have been useful in determining the level of expertise in terms of, and in relation to categories of, learned motor responses on specific control variables. This was clearly beyond the scope of the current thesis but it is an area for further study that could begin to provide evidence to test hypotheses arising in this thesis.

In Chapter 1 the implicit theory that hazard perception is a driver's application of SA (Horswill & McKenna, 2004) was discussed – notwithstanding the arguments summarised above that SA is as much a property of the task as it is a property of cognition. If Horswill and McKenna's (2004) claim holds then hazard perception scores would correlate with SA product, that is if SA product exists and is clearly definable and directly measurable. To this end the relationship between hazard perception and SA will firstly be explored in Experiment 1 by utilising hazard perception as a dependent measure, and then varying participants' SA which will be effected by training participants in SA enhancing strategies. This follows McKenna and Crick's (1991) training of hazard perception. The training uses methodology developed in consideration of Gagné's (1985) theory of training, and Hörmann *et al*'s (2003) and Hörmann, Banbury, Dudfield, Lodge and Soll's (2004) training of error management and SA enhancing strategies for pilots. Secondly, in Experiment 1 the relationship that McKenna and Crick discovered between hazard perception and driving expertise will be tested in terms of the relationship between SA and expertise which has also been observed by Endsley (2000) and by Endsley and Robertson (2000). This will be effected by experimentally manipulating expertise.

The measurement of hazard perception has been researched for several decades, as described on pages 42-45, and this measurement can be in terms of both recognition and anticipation (page 46) which are respectively related to the perception-comprehension and projection levels of SA in Endsley's (1995) theory. In Experiment 2 it is therefore intended that Endsley's SA theory will be tested in relation to recognition and anticipation dependent measures.

The research literature regarding hazard perception testing is provided in Chapter 1. The main point is that hazard perception tests have been shown to discriminate between levels of expertise and the effects of training (McKenna & Crick, 1991; 1993). Earlier methods of hazard perception failed to distinguish between responses to specific stimuli (hazardous items) and responses to the whole scene (situation recognition and hence anticipation of hazards from contextual information). Furthermore, earlier methods of hazard perception failed to address the possibly differential effects of task and awareness strategy skill on both situation recognition and anticipation; measures have either sampled situation recognition accuracy or anticipation latency. Therefore, training effects are one independent variable proposed for the first experiment, as a replication of McGowan (2002; described on p. 50) with a larger sample size and a more consistent method and procedure. The dependent measures will be provided by data derived from the MHPT relating to hazard *recognition*, and by calculating from the same data advanced responses to those stimuli as *anticipation.*

A second independent variable of expertise is proposed as a replication of McKenna and Crick (1991). In addition to replicating that experiment, another reason for manipulating expertise is that in aviation highly experienced pilots have been observed to spend more time anticipating future events and searching for information that will assist in this anticipation, whereas less experienced pilots tend to spend time passively perceiving information passively and therefore less effectively than experts (Prince & Salas, 1998). This is thought to be firstly because less experienced pilots have less memory content to draw upon, and secondly because less experienced pilots show less automaticity in their performance of sub-tasks so that mental resources cannot be freed in order to increase vigilance for changing states across the whole task system. There is only limited research reflecting a similar pattern in driving (e.g. Wikman, Niemenen & Summala, 1998). Some additional evidence is offered concerning drivers who have completed advanced driving programs (e.g, with the Institute of Advanced Drivers) who are taught to improve a skill known as defensive driving. In defensive driving actions are taken, based on hazard perception, in order to remain at a safe spatial and temporal distance from potentially hazardous situations. Recognition and anticipation of hazards are clearly major factors in experts' defensive driving technique. This is particularly so in the case of automatic perception-action anticipation in driving, given that driving hazards tend to have consequences that are only a few seconds away, in contrast with aviation where most hazards tend to be predictable and hence avoidable several minutes ahead of their consequences. Therefore, automaticity across the whole driving task is essential if the driver is to maintain awareness across all safety critical sub-tasks. Expertise, or familiarity and experience with the task and capability at the task, are therefore likely to be positively associated with SA.

It has been claimed that in aviation Endsley's (1995) levels of SA are associated with safety-related knowledge and behaviours such as attention-sharing, task-management, contingency planning, information seeking and filtering, and self-checking (Endsley & Robertson, 2000). They argue that comprehension and assessment of situations are related to the correct determination of future consequences. Schutte and Trujillo (1996) found that pilots who have strategies for dealing with distraction, interruption, and workload, which impact on the ability to manage information (i.e., strategies enabling perception of the correct information) performed better than those who did not have such strategies. Amalberti and Deblon (1992) found that experienced pilots tend to focus on anticipating future events and searching for information in order to increase accuracy and to reduce latency in anticipation. There appears to be a trend, as noted by Prince and Salas (1998), where expertise is associated with active search for information, comprehension, anticipation, and as a consequence a positive feedback cycle in terms of a more refined search for further information. Lack of expertise is associated with passive receipt of information, and hence lower SA if it can be called SA at all, constrained and controlled as novices often are to a low level of general awareness while consciously focusing on singular sub-components of a task. Assumptions about the apparent association between expertise and SA go little deeper than this superficial level in their evidential basis. In short there is a lack of research into the separate characteristics of both SA recognition and SA skilled anticipation in terms of expertise, evidence which would enable a closer investigation of SA acquisition and maintenance processes. This point will be revisited later in the thesis when theories of

SA are examined in light of both specific stimuli recognition and anticipation. However, Endsley and Robertson (2000) argue that the above relationship between expertise and SA exists in terms of active search of information, making sense of that information, and anticipation and verification of the anticipatory model through further search and comprehension. They also argue that training for such strategies increases SA. There is much support for this as follows.

Research into the training of pilots (Hörmann *et al*, 2003; Hörmann *et al*, 2004) and drivers (McKenna & Horswill, 2004) in SA enhancing strategies have shown that these non-technical skills can improve performance. In the aviation context Hörmann *et al* (2003) used a combination of training in SA, Threat Management (TM), and Situation Control (SC). Threat management is, as its name suggests, the ability to manage hazards so that their status is predicted in order to increase the probability of taking the correct avoiding or mitigating actions. TM and SA are separate but interdependent concepts (Hörmann *et al*, 2003). Situation control (Amalberti, Masson, Merritt, & Pariés, 2000) is the ability to correctly manage SA in relation to performance of the task. When these concepts were applied in a training intervention aimed at increasing SA and TM, alongside procedural skills training, it was found that the SA and TM training increased performance on a behavioural rating scale in flight simulator tasks, showing that training SA and related concepts leads to an increase in performance. Situation awareness has also been trained in the driving population in the form of hazard perception. An HPT measure was found to discriminate between levels of expertise due to hazard perception training and levels of experience (McKenna & Crick, 1991). Hence the HPT embedded anticipation measure appears to be similar to other measures of SA which correlate with expertise whether through experience or training.

If Endsley's three levels do exist independently then they are likely, as she has indicated (Endsley, 2004), to be parts of a complex system rather than a linear set of processes. The internal states of that complex system (i.e., the differing states of each of the three levels and other interconnected processes) are dependent on initial conditions and ongoing inputs. An MHPT that separately measures recognition and anticipation is a way to begin to tease apart the cognitive processes involved in the SA that arises from this system. In addition, one of the first questions that arises in relation to Endsley's model is whether recognition (perception-comprehension) and anticipation (projection) are based on the same or different cognitive processes. For example, in visual hazard perception it is likely that conscious reporting of recognised hazards may involve a different cognitive process to that supporting automatic anticipatory action. The prediction is one of many that can be derived from cognitive and neuropsychological evidence. This particular question is a major topic later in the thesis: is Endsley's (1995) model of perception supporting comprehension and projection always true? That is, do recognition (perception and comprehension) and anticipation (projection) always exist and in the same proportions, or is the product and process mediated by effects such as expertise, experience, and characteristics of the task? A hypothesis following Endsley's model, for a subsequent experiment, is that expertise will be associated with a shift from recognition to anticipation behaviour and processes. For the moment the effects of training SA strategies on recognition and anticipation in both novices and experts appear to the first step before investigating processes involved in expertise.

Therefore, Experiment 1 will not deliberately set out to test hypotheses concerning differential contributions of recognition and anticipation, this will be addressed later when investigating Endsley's model of SA and cognitive models of

SA, and when investigating cognitive processing factors during SA. Instead, since it is known that training in hazard perception and SA strategies are associated with increases in measures of those constructs, then any differential pattern in recognition and anticipation scores will be used as a test of the kinds of data which can be used to inform hypothesis in subsequent and more detailed investigations into components of expertise.

Expertise and training effects therefore come under focus for Experiment 1 as two major variables for which predictions can be drawn from Endsley's theory in relation to the MHPT measure of SA. The first step is to draw together both training and expertise and test these independent variables using MHPT data to create separate *recognition* and *anticipation* measures of SA. This will support three main areas of investigation in Experiment 1. Firstly, the prediction that hazard perception is SA in driving, secondly that training in SA strategies will be associated with an increase in SA measured with an HPT, and thirdly that expertise will also be associated with an increase in SA measured with an MHPT. The general hypothesis that hazard perception is SA in driving will be tested in two ways; firstly from the effect of training SA strategies on the MHPT; and secondly from the effects of expertise. The effects of expertise carry the assumption from Endsley's findings that expertise is associated with greater anticipatory behaviours and hence associated with greater SA. The effects of training SA strategies also carries the assumption from Endsley's findings that training is associated with greater SA. Therefore, if both expertise and SA training are associated with greater MHPT scores then it may also be inferred that hazard perception is related to SA. Any differential effects of recognition and anticipation measures in the above manipulations will be investigated for the potential of the MHPT to discriminate differences; a default hypothesis is that both measures will be correlated.

In order to ascertain whether any effects are due to SA strategies rather than a general effect of a training process it is proposed to train SA strategies and to also train a non-SA related 'sham' strategy using the same materials as an SA training condition. In the non-SA condition the textual information can be altered to change the training tasks but while retaining the images and structure of the material from the SA training condition. In order to test the above predictions, it is therefore firstly hypothesised that training in SA strategies will be associated with higher levels of hazard-specific recognition and anticipation scores, and higher levels than training in non-SA related strategies. The second hypothesis is that expert drivers will be associated with higher pre-training hazard-specific recognition and anticipation scores than novices.

A further issue is expected to arise from the second hypothesis. Assuming that hazard perception ability is constrained by the physical limits on perception of a hazard before an event - a notional point beyond which the perception or prediction of events decreases exponentially, then expert drivers are likely to be closer to an upper ceiling on their hazard perception abilities than are novices. There is a qualification to this in that anticipation appears to be associated with the perception of contextual information (see p. 32), and contextual information can be present before the hazard stimuli. This issue will be explored in later experiment in relation to the response format and automaticity.

Method

Participants

Novice drivers

Sixty-six Cardiff University undergraduate psychology students (52 female, 14 male; mean age 19 years five months; mean driving experience one year and five months) took part in partial fulfilment of a university course requirement. They were required either to have a driving licence, or to have only had driving lessons in order to increase the likelihood of them being drivers with a minimum of experience, and to have normal or corrected vision for driving.

Expert drivers

Twenty-two expert drivers participated (19 male, three female; mean age 34 years and eight months; mean driving experience 17 years and four months). They were required to have qualified as advanced drivers with the Institute of Advanced Motorists (IAM), and to have normal or corrected vision for driving. The relevant aspect of the IAM advanced driving standard is that it requires a high proportion of theoretical knowledge about defensive driving, a construct which is built around hazard perception and hazard avoidance. The goal of defensive driving is, based on high SA, to never enter hazardous situations.

Apparatus and Materials

The participants were seated at a Silicon Graphics computer workstation approximately 50 cm from a 43 cm thin film transistor (TFT) screen. The MHPT video was presented in the centre of this screen at 400 x 800 pixels (from the camera size 1:1 magnification). This had the effect of presenting no magnification or reduction of the view in relation to the scene viewed by the Police drivers. The remainder of the 1024 x 768 pixel screen showed black. The keyboard was removed and the mouse was controlled by the dominant hand.

Hazard Perception Test

The original forms of HPT (e.g., Pelz & Krupat, 1974; McKenna & Crick, 1991) were measures of the recognition of hazardous road traffic events that were staged and recorded on video. The video was presented to the participant in order to emulate the real hazard perception that occurs during driving, and therefore the hazards were preceded and followed by periods of relatively safe scenes. These interspersed periods of safe scenes were also staged as part of what was intended to be a naturalistic scenario leading up to, and following, the target hazard. During the HPT participants were required to press a response button when they perceived a hazard. If the participant pressed the button during the time period that the hazard was present then participant was deemed to have perceived the hazard.

HPT modifications (MHPT). Three modifications were made to the original form of the HPT described above in order to improve both its validity and resolution for use in thesis. A detailed description of the MHPT is provided on page 46. The first modification was to use real scenes selected by traffic Police from their archives which were then rated by a panel of expert drivers, in order to increase criterion and ecological validity. The second and third modifications were effected through computer mouse responding in order that the correct spatial and temporal points of the hazard could be determined on the video. Therefore, for the second modification, if a response occurred within the area on the screen where a hazard existed then the spatial criteria for the hazard were met. For the third modification if a response occurred during the correct time period when the hazard was present (i.e., if the spatial criteria were met) then the time criteria for the hazard were also met. If both spatial and temporal criteria were met then the response met the full criteria for the hazard. This results in a *recognition* measure. In addition the time criteria for the hazard enabled subsequent calculation of an anticipation score (after correct recognition) in hundredths of a second by subtracting the time of the response from the time that the hazard ended. That is, the *anticipation* measure is the time elapsed between the mouse button response and end of the hazard. This scale is therefore theoretically more likely to enable a higher resolution of measure than the categorical score of correct responses resulting from the original form of the HPT. The modifications were also theoretically important for this thesis in that they provide an HPT anticipation scale which corresponds at face validity both with Endsley's (1995) final product of SA (projection-anticipation) and with driving hazard perception. The two dependent measures resulting from these modifications that will be used in this thesis are will be known as *recognition* and *anticipation*.

Video. The video material used in the HPT was obtained from an archive of Police recordings filmed in April 2002 on motorways, main roads, and urban main roads in South Wales. The Police video portrayed naturally occurring events obtained from cameras mounted in Police cars in front of the driver during routine traffic patrols. The scenes varied from dry and bright to overcast and raining, and seven percent of the hazards were recorded at night in clear visibility. The video was selected for the researcher by Police officers from Police archives.

Hazard severity. The events in the video were required to represent varying severity of hazards in order to both increase realism and provide temporal separation between hazards, as would occur during real driving. Hazard severity was defined by Police drivers as either high, medium, or low severity, at the moment of maximum hazard (the end of the hazard) as follows: high severity hazards were those likely to lead to a collision resulting in death or life-threatening injuries to a person. Examples of high severity hazards used in the MHPT were a car swerving in front from right to left across motorway lanes in heavy traffic in order to reach the off-ramp, hence an increased risk of a high-speed impact; and a pedestrian stepping from an obscured location into the road immediately ahead, hence a high risk of impact and severe trauma or death. Low severity hazards were those likely to lead to a collision resulting in minor injuries to a person which do not require hospital treatment, or a collision not resulting in injuries. Examples of low severity hazards are those likely to result in a collision between a vehicle and a stationary object at less than six miles per hour such as a car braking ahead in a queue of slow moving traffic, or two vehicles on courses that are likely to result in them colliding side to side at an acute angle. Medium severity hazards were defined as those that were neither high nor low severity and which were likely to lead to hospitalisation and major chronic injuries such as amputation or paralysis.

While the hazard severity criteria are difficult to assess objectively it was expected that the Police drivers' expertise relating to the outcomes of different hazards and accidents would be sufficient to produce videos which contain a natural range of hazards in terms of their severity. The classification also enables a level of control over hazard severity. An instruction was given to the Police officers: 'To provide video containing an equal number of high severity hazards, medium severity hazards, and low severity hazards, together with periods of negligibly hazardous scenes leading up to the hazards and all within continuous footage wherever possible'. The Police officers provided one hour and twenty minutes of video under this instruction. They identified a total of 103 hazardous events. The levels of severity are shown in Table 2.

Task analysis. The video provided by the Police was analysed by a group of six advanced driving instructors (ADIs) belonging to the IAM (Institute of Advanced Motorists) who are experienced in, and continue to train in, hazard perception and other strategies related to the practice of 'defensive driving'. As a group and facilitated by the researcher, the ADIs identified hazards on the video and the temporal and spatial parameters of each hazard. The basis of this verbal protocol analysis is described on page 51. The researcher instructed the ADIs to indicate the moment at which each hazard was first perceivable and its most severe moment, thereby providing a beginning and an end time for each hazard. The researcher recorded these data where there was a consensus between the ADIs, and this informed a precise time specification for each hazard. For example, where the ADIs reached a consensus in defining a hazard as 'car overtaking will move left too close in front of us and decelerate (Figure 2), beginning at video frame *n1* and ending at video frame *n2* at which time the hazard began to diminish', the researcher recorded time points of video frames *n1* and *n2.* Time point *n1* is therefore the first moment at which the hazard may be perceived and time point *n2* is the moment at which the hazard ends. From the description of the hazard the researcher later recorded its spatial location in terms of two x and two y pixel coordinates relating to the left and right borders and top and bottom borders of a rectangle. Figure 2 provides a visual representation of the spatial criteria for a hazard, where the box labelled 'correct' is the boundary of the hazard defined by the expert drivers at its final frame before the hazard began to diminish. The ADIs identified and provided spatial and temporal criteria for 156 hazardous events, which included all the events identified by the Police drivers.

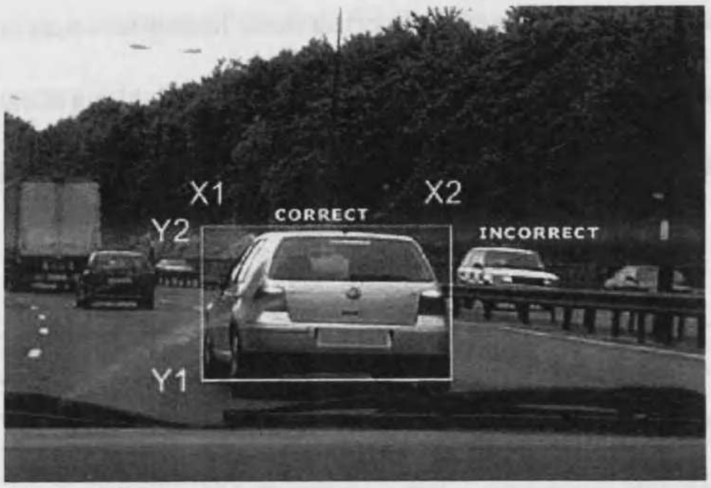

Figure 2. The white box labelled 'correct' illustrates the spatial criterion for an example hazard: 'The car overtaking will move left too close in front of me and decelerate' The hazard is at its ending video frame (time=0) as defined by the ADI group. Any visual stimuli not within the correct boundary, such as the oncoming Police vehicle, are classified as incorrect.

Video compilation. Of the 156 events identified by the ADIs 120 were selected for use in the MHPT. These 120 ADI identified events included the original 103 identified by the Police drivers. Therefore, the severity data shown in Table 2 is only a guide to the distribution of hazard severity. This is due to the omission of Police driver severity data for 17 of the subsequent ADI hazards. There were two reasons why the item set was reduced. Firstly, there was a need to maintain the continuity and original order of the scenes in the video, because the naturalistic context may provide stimuli that begin to signal the possibility of a hazard before the hazard becomes explicit. Secondly, it was necessary to organise the video material into two clips, one for each of the within-subjects pre and post-training conditions. The two clips were required to be balanced in terms of the mean hazard period (Table 1) as defined by the ADIs in order to maintain as much similarity between these split forms of the measures, which led to an imbalance in hazard severity. The balancing of hazard duration was effected by categorising the hazards into short, medium, and long duration. Short duration hazards were defined as those lasting less than six seconds, medium duration hazards between six and twelve seconds, and long duration hazards lasting more than twelve seconds. The duration and severity of the hazards is shown in Tables 1 and 2 respectively.

	Duration						
Video	Mean (seconds)	SD	Max (seconds)	Min (seconds)	Low	Medium (count)	High
A	9.12	4.49	18.18	2.83	15	25	20
B	9.11	3.65	19.39	2.68	7	36	17
Total	9.11	4.08	19.39	2.68	22	61	37

Table 1. Mean hazard duration and distribution of 120 ADI defined hazards

	Severity[2] (count)		
Video	Low	Medium	High
A	7	34	15
B	4	34	9
Total	11	68	24

Table 2. Hazard severity distribution of 103 Police driver defined hazards

The two videos were closely matched in overall length by removing small sections of footage from the beginning of the longest lead-in times before the hazards became predictable. For example, a period of eleven seconds elapsed after hazard 87 ended and before the next hazard 88, a car emerging from a side street, became predictable. Therefore, the removal of the first three of the 11 seconds was thought unlikely to affect hazard perception because the signs indicating that a junction was ahead were not visible until six seconds before the junction where the hazard (a car) became visible. Therefore the three seconds of discarded video were unlikely to interfere with hazard perception. The clipping resulted in Video A at 14 minutes and 18 seconds in total length and Video B at 14 minutes and 20 seconds in total length,

2 Severity data applies only to the original 103 hazards identified and rated by the Police drivers. Duration applies to the final 120 hazards. This means that 17 of the 120 items selected by the ADIs were not rated for severity.

each containing 60 hazards. These two clips were piloted by McGowan (2002) on 32 undergraduate psychology students and resulted in mean baseline anticipation scores of 4.9 seconds for Video A and 4.34 seconds for Video B without training.

MHPT presentation. The two clips were embedded in a Microsoft Visual Basic program which presented the required MHPT content for a trial and recorded the dependent variables. The ADIs' hazard event criteria were also embedded in a database within the program so that scores could be automatically calculated from the dependent variables. The process of developing this score reporting component was found to be beyond the researcher's time constraints and therefore the data were processed individually. The participants were to identify and respond to hazards with the computer mouse. The program was designed to compare those responses with the ADIs' criteria for the hazards and produce a *recognition* accuracy score and an *anticipation* latency score in hundredths of a second. The first correct response within the hazard criteria was selected for analysis, although all responses and the mouse positions at each frame were also recorded to enable a wider range of subsequent analyses. The final score for the whole set of hazards may therefore comprise both anticipation scores for correct responses and zero scores for incorrect responses. The mean of the hazards to which participants correctly responded is used as the final MHPT score for each trial.

SA Training Material

The SA training condition material was designed in order to train a cognitive strategy (Gagné, 1985) for maintaining SA in accordance with Gagné's (1985) theory of instruction. The material was piloted by McGowan (2002), and its use was found to be positively correlated with MHPT scores. The material was paper-based and took approximately twenty minutes to complete. The material comprised five sections.

These sections were as follows: 1 a pre-training (Gagné, 1985) or introductory section in which the context of SA and driving is explored; 2 perception of elements of information in driving; 3 comprehension or integration of perceived elements while driving; 4 projection or anticipation; 5 a summary where the context in section 1 is explored again but in relation to the material covered in sections 2 to 4. Sections 2 to 4 were designed so that each made no reference to the main concepts being trained in the other two sections. For example, section 2 made no reference to the concepts in sections 3 and 4. The material was produced this way so that in subsequent experiments the same strategies for each level of SA may be trained in separate conditions. The contextual knowledge in sections 1 and 5 was provided so that in a complete SA training condition this would assist in integrating sections 2-4. The responses provided on the forms were verifiable in terms of the completion of all the items. However, accuracy was not sampled apart from comparing the section 2 results with the expected pattern (described in results).

Section 1 pre-training. In accordance with Gagné's (1985) theory a pre-training section set the context for the following sections firstly by asking the participants to answer questions about driving and how they maintain awareness of the environment external to the vehicle. Secondly, the participants were directed to think about their driving experiences, how they had learned to drive, and whether they have continued to learn driving skills. Thirdly, SA was introduced within the context of enhancing driving skills and road safety through greater anticipation.

Section 2 perception exercise. The first level of Endsley's (1995) theory of SA is perception of information. Examples of visual scanning derived from a Police driving handbook (The Police Foundation, 1994) were provided in this section in order to transfer a habit of scanning aimed at increasing the perception and maintenance in memory of relevant information (Awh, Jonides & Reuter-Lorenz, 1998). The participants were directed to draw lines representing their visual scanning pattern and then to compare their responses with examples (Figure 3) to develop an understanding of visual scanning patterns. This section used two of these exercises. This exercise was not scored, however completion of the examples was verifiable.

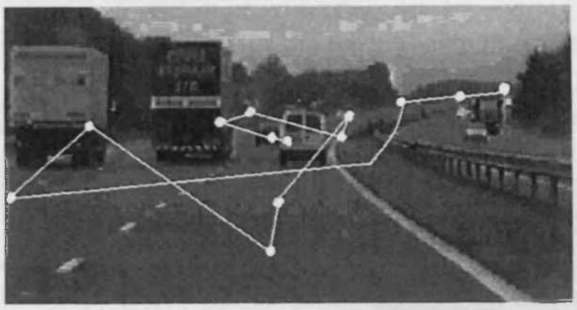

Figure 3. Visual Scanning Pattern. This pattern begins at the rear of the white van and fixates on objects progressing from near to far before repeating.

Section 3 comprehension exercise. The third section was designed to train participants to combine perceptual items in order to derive greater meaning about events in the road environment. For example, the material asked what combinations of individual items such as street lighting, road marking, and furniture might indicate about the presence of hazards. Table 3 shows two examples, participants were instructed to complete up to seven more rows in Table 3 with similar items of their own. This exercise was not scored, however completion of the examples was verifiable.

Item	Meaning to be completed by the participant
Street lighting	*(example)* This is a built up area, and there is likely to be a 30mph limit unless signposted otherwise.
Street lighting	

School-children sign	*(example)* Road designers are warning you of children or a crossing patrol on the road.

Table 3. 'Developing meaning' exercise for comprehension conditions. An item is provided on the left. A response from the participant is required on the right.

Section 4 projection exercise. The third level of Endsley's (1995) theory of SA is projection. In this training section the importance of looking and thinking ahead was explained as a way to increase anticipatory activity. The participants were then instructed to perform an exercise in anticipation (Figure 4). This exercise was not scored, however the extent to which the examples had been completed was verifiable.

Figure 4. Anticipation exercise with the instruction: What do you anticipate happening in this scene? The following list is an understanding of some items in the picture. Use these meanings, and more of your own if you can, to describe in an integrated way what likely events you anticipate.

Section 5 summary. Finally, the text of the training material revisited the context of situation awareness in driving in accordance with Gagné's (1985) theory of instruction. In this section the relationship between the three sections was explained in terms of an outline of Endsley's (1995) SA theory. The three levels of SA were shown to be interdependent, and the concept of an action-perception positive feedback cycle was described in the context of driving.

This training material was completed in 20-30 minutes, with a limit of 30 minutes imposed. All participants completed the training material within 30 minutes, although there were omissions in several cases. However, of eleven participants who omitted responses only two participants had three omissions and six had two omissions. This is unlikely to have significantly affected any training effects overall.

Non-SA training material

The non-SA training condition material used the same paper-based format as that of the training condition, and used the same video stills. The purpose of this material was to direct the participants as fully as possible to engage with the same images as those in the training condition. This involved a task that minimised references to SA and anticipation and which avoided directing participants to monitor dynamic events over time. The total of five images were the same as those in the SA training material, except the text which was replaced with a 'distance perception test' requiring the participants to estimate distances between arrows on the images laterally (an example item is shown in Figure 5.) Participants responded to each question by writing the distance in imperial or metric units beside the image. This was not scored for accuracy. However, all but three items received responses.

Figure 5. Lateral distance perception example with the instruction: Please judge the lateral distance between the white van in front of you, and the dark car to its left, shown by the two arrows.

Instructions

The printed instructions included the following statement that the task was to view the video of road scenes, and to use the computer mouse to click on hazards *as if* responding to the event in a car. Use the mouse to click on objects in the road view where you see a hazard that would cause you to take some physical action when driving a real car (Appendix A-1)

Design

This experiment used a three factor (2 x 3 x 2) mixed design. The first factor was expertise and comprised two between-participants conditions: novice drivers, and expert drivers. The first factor subsumed the second factor which comprised three between-participants conditions: no training, non-SA training, and SA training. The second factor subsumed the third factor which comprised two within-participants conditions: pre-training, and post-training.

Dependent measures

The MHPT measures anticipation after ascertaining that both the spatial criteria and the temporal criteria have been met for a hazard. However, for this experiment the number of responses meeting both the ADI's correct spatial criteria and temporal criteria, and the anticipation scores contingent upon those MHPT 'hits' were analysed separately. This produced two main dependent measures: the recognition score which sampled the sum of hits, and the anticipation measure based on those hits.

Recognition measure. This measure was the number of responses which fell within both the ADI defined spatial locations and also within the ADI defined temporal hazard periods. This measure which depends on a response somewhere within each of both sets of criteria is termed the recognition score and is a measure of recognition of a hazard. This measure recorded a score of one when at least one correct response was made in relation to a hazard.

Anticipation measure. The second dependent measure used the complete MHPT measure in which an anticipation latency score is added to the recognition measure. The anticipation latency was based on the same recognition data. However, each recognised hazard was converted into an anticipation score by subtracting the end time of the hazard from the response time. For example, if a hazard was defined by the expert drivers as first perceivable at 30 seconds from the beginning of the video and ended at 40 seconds and the response was at 33 seconds then an anticipation score of three seconds was recorded. The data used for analysis was based only on the recognition responses which were scored as either one or zero; the mean anticipation score was calculated only of the positive recognition scores. That is, the recognition score sampled the sum of hits, and the anticipation score was the mean latency of the hits before the moment of maximum hazard. The zero scores were excluded from the analysis in order that the measure would be sampling the level of anticipation only within the recognition hits. The anticipation analysis based on a mean only of the hits (scores greater than zero) permits a like for like comparison of scores from the two measures.

Data collection. The dependent variables above were measured by recording the time point and spatial location of every mouse response in x and y pixel coordinates. The resolution of this record is 0.01 second due to variance of the computer clock (theoretically 0.001 second plus or minus 0.005 second), and this is constrained to a resolution of 0.01 in the sampling of mouse responses by the Visual

Basic program with which the MHPT was constructed. Therefore the final resolution of the measure was determined to be 0.01 second and at 1 screen pixel. The resulting data were then filtered by comparison with a database of the ADI hazard criteria to select the first correct response to each hazard, or to return zero when the hazard criteria were not met for a hazard.

Demographic dependent measures. The biographic and driving history dependent measures collected were the participant's age, sex, the length of time they had held a driving license (or provisional license), the number of accidents in which the participant had been involved whether at fault or not, and the mileage the participant had driven in the preceding week. This data was appended to the MHPT measure scores and was identified by a unique number. The participant's name and consent was recorded separately.

Procedure

The experiment was presented at a computer workstation. The participants were told that the experiment comprised a hazard perception test, a short paper-based exercise, and then a second hazard perception test. Demographic data were collected from the participants by using an electronic form presented on the computer screen. Participants were then presented with the printed instructions. There was no practice trial in the present experiment, and the instructions did not place an emphasis on speed or accuracy.

The participants clicked the start button on the screen using the mouse. As the video began the main point in the instructions was reemphasised verbally: 'Use the mouse to click on objects in the road view where you see a hazard that would cause you to take some physical action when driving a real car.'(Appendix A-1)

This verbal reminder of the instruction was performed in both the pre and post training trials. There were 13 seconds to the first hazard in video A and seven seconds in video B, allowing the verbal instruction to be given before the first hazard was encountered. When the first video had ended the participants in the SA training and non-SA training conditions were given the paper-based training material. Participants in the no training condition were directed to take a short break from the screen for three minutes. However, this period was uncontrolled when the starting times of multiple participants became staggered, resulting in an unstandardised rest period, although the period was approximately between 30 seconds and 3 minutes. Following this rest the second video was presented in the same way as the first.

Results

Firstly, to discover the degree of relationship between recognition and anticipation scores a correlation was performed on the scores across all conditions between the two MHPT measures. The correlation was $r = 0.27$, $p = 0.037$. This indicates that there is a relationship between recognition and anticipation scores but that the correlation is low or weak. This supports a prediction that two separate processes are involved in the MHPT task.

Training material analysis

Scanning exercise. Participants were asked to draw two scanning patterns after being shown an example. All but three of the responses followed patterns which correctly followed the instructions. Each of the three had completed one of the items but had omitted to complete the other item. This indicates that overall the participants are likely to have carried out the exercise correctly.

Comprehension exercise. Participants were asked to develop meanings of the road scene after being shown examples. The mean number of responses was 2.3, the minimum was two and the maximum was five. This indicates that overall the participants are likely to have carried out the exercise correctly.

Projection exercise. Participants were asked to anticipate events in the road scene after being shown examples. The mean number of responses was 2.1. The minimum was one, and the maximum was four. This indicates that overall the participants had carried out the exercise correctly.

Recognition dependent measure

The first hypothesis was that training in SA strategies will be associated with higher levels of recognition (perception of correct information content) than those who have not been trained, and higher than those who have been trained in non-SA related strategies using the same materials but with the content altered in each to train SA strategies in one condition and a non-SA strategy in another condition. That is, there would be a positive effect of training in the SA training condition alone in terms of recognition scores. The descriptive statistics (Table 5) indicate that there is only likely to be a small difference between the SA training condition and no training or non-SA training in terms of pre-post training recognition scores.

Training	Expertise	Train Type	Mean recognition score	SD
Pre Train	Expert	No Train	32.16	6.63
		Non-SA	32.56	10.39
		SA Train	32.42	4.00
		Total	32.38	7.01
	Novice	No Train	34.82	6.41
		Non-SA	33.33	8.82
		SA Train	42.57	5.82
		Total	36.91	7.02
	Total		34.64	7.01
Post Train	Expert	No Train	42.63	8.46
		Non-SA	47.52	5.54
		SA Train	40.74	1.76
		Total	43.63	5.25
	Novice	No Train	33.57	9.03
		Non-SA	42.72	6.75
		SA Train	43.18	5.38
		Total	39.82	7.05
	Total		41.73	6.15

Table 5. Means and standard deviations of training, expertise, and training type recognition scores.

A three factor (2 x 3 x 2) mixed ANOVA showed no significant interaction of expertise, training type, and training in terms of correct recognition responses (F (1,2,1) = 0.58, p = 0.56). However, there was an interaction of expertise and training (F (1,1) = 12.35 p = 0.001), an interaction of training and training type (F (1,87) =

4.62, $p = 0.013$), and an interaction of expertise and training type (F (1,87) = 5.11, p = 0.008) in terms of correct recognition responses. The same ANOVA showed a significant main effect of training (F (1,87) = 35.69, $p < 0.01$) from 35.82 pre-training to 40.75 post-training correct recognition responses. There was a significant main effect of training type (F (2, 85) = 3.36, $p = 0.04$), no training mean 35.8 correct recognition responses, non-SA mean 39.04 correct responses, and SA training 39.73 correct responses. There was a non-significant main effect of expertise (F (1,86) = 0.74, $p = 0.786$) on correct HPT responses.

Having discovered significant effects of both training and training type on correct responses, the difference between pre and post training scores was computed for each case, producing a further dependent variable that was termed training difference. A one-way ANOVA was then performed between the three training type conditions in terms of the training difference scores. This ANOVA showed a non-significant effect (F (2,85) = 0.67, $p < 0.51$) of training type in terms of the difference between pre and post training recognition correct responses. The first hypothesis, that SA training would be associated with a greater effect of training than no training and non-SA training, was not supported using recognition responses as a dependent measure. In addition, a comparison of standardised differences between pre and post training in terms of Z scores shows a similar finding: z score of differences no training $z = -0.54$; non-SA training $z = 3.56$; SA training $z = 3.65$. Both these analyses concur in failing to support the first hypothesis that training SA strategies will be associated with an increase in recognition MHPT scores greater than a non-SA training strategy.

The second hypothesis was that that expert drivers would show greater recognition (a higher number of correct recognition responses) on a pre-training recognition measure than novices. I.e, expert drivers would be associated with a higher number of responses meeting the correct ADI recognition criteria than novice drivers in the pre-training conditions. A T-test ($T(86) = 0.29$, $p = 0.72$) showed no significant difference between the novices and experts in the three pre-training conditions in terms of recognition responses.

No significant correlations were found between expertise, training type, or training, and age, sex, duration of license, and number of accidents in the previous three years, and mileage driven in the preceding week. Therefore a computation of the effects of covariates on the ANOVA was not required.

Anticipation dependent measure

This analysis tested the same factors as in the spatial and time responses analyses above but using those two measures combined into the anticipation measure (mean of hazards scoring greater than zero). In this analysis, for each participant only the hazards which scored greater than zero were used when computing the mean score for a trial. The dependent measure was therefore an HPT anticipation score contingent on both temporal and spatial measures. That is, if a response was correct both spatially and temporally, then an anticipation score was generated by subtracting the time of maximum hazard from the time of the participant's response.

A second test of the first hypothesis using the anticipation dependent measure showed no effect of training. The hypothesis was that training in SA strategies will be associated with higher anticipation scores than no training, and higher than training in non-SA related strategies using the same materials but with the content altered in each to train SA strategies in one condition and a non-SA strategy in another condition. I.e, there would be a positive effect of training in the SA training condition alone in terms of anticipation scores. The descriptive statistics (Table 6) indicate that there is only likely to be a small difference between the SA training condition and no training or non-SA training in terms of pre-post training anticipation scores.

Training	Expertise	Train Type	Mean (seconds)	SD
Pre Train	Expert	No Train	6.03	1.85
		Non-SA	4.90	2.95
		SA Train	4.34	2.56
		Total	5.09	2.45
	Novice	No Train	5.09	1.31
		Non-SA	5.29	2.58
		SA Train	4.79	1.71
		Total	5.06	1.87
	Total		5.07	2.16
Post Train	Expert	No Train	6.72	0.76
		Non-SA	7.38	1.30
		SA Train	8.54	0.27
		Total	7.55	0.78
	Novice	No Train	5.24	1.87
		Non-SA	4.95	1.78
		SA Train	8.73	1.71
		Total	6.31	1.79
	Total		6.93	1.28

Table 6. Means and standard deviations of training, expertise, and training type anticipation scores.

A three factor (2 x 3 x 2) mixed ANOVA showed a non-significant interaction of expertise, training type, and training in terms of anticipation scores (F (2, 87) = 1.54, $p < 0.21$). There was no significant interaction between training and expertise (F (1,1) = 3.45, $p = 0.07$). However, there was a significant interaction between training and training type (F (1,87) = 11.5, $p < 0.01$) in terms of anticipation scores. The same ANOVA showed a significant main effect of training (F (1,87) = 30.42, $p < 0.01$) from 5.07 seconds mean pre-training to 6.93 seconds mean post-training anticipation score. There was a significant main effect of training type (F (2, 85) = 3.26, $p < 0.045$), no training mean 5.78 seconds, non-SA mean 5.66 seconds, and SA training 6.6 seconds anticipation score. There was a non-significant main effect of expertise (F (1,86) = 3.52, $p = 0.064$) on anticipation scores.

Having discovered an effect of training and training type on HPT scores, the difference between pre and post training scores was computed for each case, producing a further dependent variable that was termed training difference. A one-way ANOVA was then performed on training type in terms of the training difference scores. This ANOVA showed a significant effect (F (2,85) = 18.21, $p < 0.01$) of training type on the difference between pre and post training anticipation scores: No training 0.28 seconds; non-SA training 0.26 seconds; and SA training 4.01 seconds. A Newman-Keuls post hoc test showed that SA training was associated with a significantly greater difference in anticipation scores (mean increase 4.01 seconds) than no training (mean increase 0.28 seconds) or non-SA training (mean increase 0.26 seconds) at the 0.05 level. In addition, a comparison of standardised differences between pre and post training in terms of Z scores shows a similar finding: z score of differences no training $z = 0.26$; non-SA training $z = 0.41$; SA training $z = 3.32$. Both these analyses concur in supporting the first hypothesis that training SA strategies will be associated with an increase in MHPT scores.

The second hypothesis was repeated using the anticipation dependent measure. The hypothesis was that expert drivers would show higher scores on a pre-training anticipation measure than novices. I.e, expert drivers would be associated with a higher mean anticipation score than novice drivers in the pre-training conditions. A T-test ($t (86) = 0.117$, $p = 0.59$) showed no significant difference between the novices and experts in the three pre-training conditions in terms of anticipation scores.

Discussion

One aim of Experiment 1 was to investigate the relationship between hazard perception and SA by testing the effects of SA training and expertise in terms recognition and anticipation MHPT measures. To enable a test of the prediction that hazard perception is SA during driving SA strategies were trained and expertise was manipulated. The finding that SA training and expertise are associated with anticipation scores tends to support this hypothesis; if anticipation represents the highest level of SA then it is reasonable to use this as the most compatible measure, although in later chapters the utility of anticipation alone will be discussed and investigated further. The findings of the effects of the above manipulation on the recognition measure also support this hypothesis. Therefore, it is reasonable to assume that the anticipation measure is construct valid when used as an indicator of SA. However, the recognition measure appears not to contribute as an assessment of SA product. The pattern of the effects of training and expertise on the recognition measure did not show significant results.

A second aim of Experiment 1 was to investigate the pattern of data that would be found using the MHPT measures to repeat other findings concerning the effects of expertise and training SA strategies. In both cases, the anticipation measure showed predicted differences but the effects of expertise were difficult to explain. Nevertheless, the effects of expertise are highly significant. The effects of training are highly consistent with predictions and therefore a tentative hypothesis is that the effects of training SA strategies reveals the potential utility of the MHPT as a measure of SA, while the effect of expertise reveals what may be a more complex relationship between expertise and SA.

Also of interest is the finding that the ADI and Police drivers' selection of hazards differed. The ADI's found 156 hazards and the Police drivers found 103 hazards. The reason for this disparity is likely to have been that the ADI group reached their decisions collaboratively under the direction of the researcher and the researcher chose to accept a decision when there seemed to be a majority opinion that an event was a hazard. The Police drivers selected the hazards without the researcher being involved in the process, and furthermore the Police are experienced in decision-making concerning the use of video evidence and may have erred on the side of caution when deciding whether or not to include an item in the hazard set. For the Police if there is doubt they would exclude the item from the 'evidence'.

The pattern of results using the anticipation measure is consistent with its use as a measure of overall SA output, but the recognition measure shows a different pattern. Endsley (1995) argues that perception of information and comprehension of that information, in terms of long-term and WM representations, support anticipation ability. It would therefore be reasonable to expect to see a similar pattern of recognition scores as the pattern of anticipation scores. One explanation alluded to when introducing the present experiment, and supported by the observed low correlation between recognition and anticipation scores, is that recognition and anticipation may occur differentially in terms of the effects of expertise, and/or the effects of training, and/or cognitive processing. This explanation is given more weight in light of the finding that task expertise was not associated with a pattern of results similar to that of the effects of training in terms of anticipation scores. There are observations in the literature which suggest that task expertise and the effects of training SA are both involved in the acquisition and maintenance of SA, however, these patterns are only partially evident from Experiment 1.

The failure to support the predictions of Experiment 1 in terms of the recognition measure and expertise leads to a hypothesis that expertise may not relate to recognition in the same way as it does to anticipation. Another finding which did not concur with the prediction that experts would show higher pre-training anticipation scores is that the MHPT may not be a good measure of SA through its relationship with task expertise. There has been an assumption, conveyed in much of the SA literature including Endsley (1995), that task expertise is positively related to SA and that SA supports task expertise. However, if task expertise creates SA what we would really be interested in is the task-focused actions and how they would give rise to SA, and not the qualities of the SA itself such as how much of SA is perceptions, or comprehension, or higher order processes such as projection and anticipation. Therefore if we want to investigate processes in SA we should manipulate the task and observe how this affects certain variables that link the two. To attempt to measure representations as an indicator of SA, which from most accounts seems to be occurring at an unconscious level between actions and perceptions, admits too much uncertainty into any theory, and Endsley's theory is uncertain in several areas. For instance, the lack of clear connection with models of memory from which hypotheses can be drawn.

Of several alternatives argued by Banbury and Tremblay (2004) one is that perhaps SA is an epiphenomenon of cognition and of the cognition involved in a task. There is also research into consciousness which has tested that concept using studies of volition, to show that conscious awareness may be slightly preceded by or be simultaneous with the decision or action thought to have been permitted by awareness (Libet, Wright, Feinstein and Pearl, 1979), although alternative explanations have been made for this phenomenon (Klein, 2002) concerning the extent of conscious control over volition. However, this generally supports the argument that task-focused SA or more general consciousness are representations of the event that somehow lag behind action but nevertheless are used to predict future events and further support behaviour or learning - hence SA is simply a perception of states of internal representations that are mainly driven by automatic processing.

There are several computational models of ways in which back-propagation of real-time data (Rumelhart, Hinton, & Williams, 1986) may be associated with the emergence of factors which can then be used to predict future inputs. If SA resides in the decaying memory trace then it equates to the non-linear products of computational back-propagation. Such information content potentially supports the prediction of future events and hence it is one potential source of conscious awareness and anticipation. A cognitive mechanism by which such a system can work - in addition to Baars' (1988, 2002) global workspace - is cognitive streaming (Jones, Beaman & Macken, 1996; Banbury *et al*, 1993). Cognitive streaming resides in phonological

WM even when it carries visual-spatial calculations. A cognitive streaming account of SA supports the idea that SA relies on serial data and anticipation functions to provide a serial bridge from past events to predicting likely future events.

Failure to support the predictions of Experiment 1 in terms of the recognition measure and expertise leads to a hypothesis that expertise may not relate to recognition in the same way as it does to anticipation. Another finding which did not concur with the prediction that experts would show higher pre-training anticipation scores is that the MHPT may not be a good measure of SA through its relationship with task expertise. Generally the results of Experiment 1 were useful in that they established a degree of confidence in the MHPT as a measure of SA. The unexpected pattern of results concerning expertise and recognition also shows that the processes supporting the acquisition and maintenance of SA are complex, and in this sense the results agree with Endsley's description of SA as a complex process (Endsley, 2004). However, Endsley does not specify with any clarity the structure of the SA process which the researcher has inferred from the model. However, the findings of Experiment 1 offer some initial observations of how these complex processes may look under experimental conditions.

There is a further emerging factor related to the anomalies concerning expertise and recognition, one which arises from the literature into anticipation and context. A strong finding in research into anticipation is that it appears to be achieved as much through contextual as specific stimuli similar to the way that learning is thought to occur through these two processes (e.g., Bruner, 1966). Banbury *et al*'s (2004) cognitive streaming account of SA offers a possible probabilistic mechanism for such processes whereby the actual occurrence of events combined with long-term memory representations of related past occurrences provide a calculation of the probability of a future event occurring. Furthermore, Czerwinski, Lightfoot and Shiffrin (1992) showed that attention can automatically drive action before awareness occurs. Therefore, if anticipation is based on contextual information beyond the target item itself then the MHPT measure may not be the best way to measure experts' SA. If an expert normally takes automatic anticipatory action then expecting that expert to place a mouse cursor over a target that he may not yet be consciously aware of does not permit him to display his true task-related anticipatory ability. An anticipation measure may capture his awareness when the stimulus item becomes conscious to him, but this measure will be a sample later in time than his normal anticipatory action response. Therefore, some of the anomalies concerning expertise in Experiment 1 may also be related to the ways in which specific and contextual information usage differs with expertise.

These unexpected effects of training and expertise, and the constructs that are being measured by the MHPT will therefore be major topics of investigation later in the thesis both experimentally and theoretically. However, for the second experiment the next logical step is to continue to focus on one of the manipulations from the first experiment and in a direct investigation of Endsley's levels of SA.

Experiment 2

In Experiment 1 expertise and training effects were investigated. In order to build on this evidence it was intended to continue to use the training paradigm to manipulate components of Endsley's three-level model. That is, Experiment 2 was an opportunity to manipulate the SA strategy training terms of Endsley's theorised levels in terms of MHPT scores.

The results of Experiment 1 confirmed the predictions directly only in two areas. The effect of training SA was an increase in MHPT anticipation scores. Clearer still was the expected pattern of results between the no training, non-SA, and SA training conditions in terms of anticipation scores. Therefore, the next step in this thesis for Experiment 2, was to continue to use the training manipulation in order to investigate Endsley's levels of SA. This can therefore be effected using the MHPT anticipation measure.

Endsley (1995) argues that the cognitive aspects of SA are characterised by, and according to some accounts composed of, three main processes. These processes are perception, comprehension, and projection. Endsley further posits that all three levels of SA are necessary (Endsley, 2004). In addition, she contends that because behaviours related to the three level SA model are represented along a scale from perception, to comprehension, and then to projection, then these levels represent the levels of SA a person can report or act on (Endsley, 1995). Furthermore, when that latter concept concerning the model of SA (Endsley, 1995) and its relationship to the observed behaviours at different levels of expertise is examined, it soon becomes apparent that this is circular reasoning. It is circular in that it is based on a construct that emerged from the same data that gave rise to that construct. The levels of SA model has also been used to argue (or intriguingly has been misread as a process when it was not intended to support one; Endsley, 2004) that there are distinct processes involved in perception, comprehension, and projection. This logic appears at times to be a central tenet of Endsley's (1995) theory of SA, and at other times it is said to describe the behaviour of pilots at varying levels of expertise. Endsley (2004) states that the three levels in Endsley's (1995) behaviour-as-process theory occur neither linearly nor discretely. However, in apparent contradiction she goes on to state that level 2 SA (comprehension) builds on level 1 SA (perception) and level 3 SA (projection) builds on level 2 SA. Endsley (2004) now appears to disclaim such notions, instead saying that the processes involved in SA are complex but offering little more in terms of a model of SA processes. It is now even more unclear as to whether Endsley is referring to qualities as descriptions of civil airline pilot behaviours, or cognitive processes supporting SA. It is likely that this ambiguity stems from the potential for circular reasoning in Endsley's model (Flach, 1995). Whatever its origins the ambiguity does not assist in describing the processes that support SA. Therefore, Experiment 2 is an attempt to clarify the existence of, and precise relationship between Endsley's three theorised levels of SA.

A finding of Experiment 1 was that expertise was not associated with anticipation scores. This may be due to experts' hazard perception being an automatic anticipatory action process which is not sampled by the anticipation measure. Alternatively, SA processes may not be characterised by Endsley's three levels at all because a measure of conscious recognition (recognition) did not distinguish between SA training strategies. The recognition measure did not discriminate between the training types but the anticipation measure - including a strategy to increase recognition - did distinguish between types of training strategy. Therefore if a training strategy, of which the most effective part is to train drivers to actively search the road scene, cannot increase explicit recognition then how can the same information later increase anticipation? The two route hypothesis is one explanation. Another related alternative was that there are not three kinds of processing relating to Endsley's levels. Instead, it may be that there are two processes, one automatic and implicit and the other more conscious and controlled. Given the arguments surrounding effects seen in Experiment 1 then implicit-explicit and recognition-anticipation distinctions seems more plausible. It has been theorised by Milner & Goodale (1992; 1995) that there are two structural and processing visual systems. The ventral stream relates to representational constructivist processing and explicit recognition. The dorsal stream is related to automatic perception-action, visually guided action, and behaviourist principles.

It is striking that Milner and Goodale's theory so closely matches a two process theory of SA; skilled anticipation using the perception-action route due to the adaptive importance of rapid anticipation, and recognition as the reconstructivist and explicit perception-comprehension process in memory. This proposal is more elegant than Ericsson and Kintsch's (1995) theory of LT-WM which struggles to explain a process through which long-term and WM are mediated.

The above point, that there may be two visual systems involved in the acquisition and maintenance of SA, was introduced at this juncture because it relates to the hypotheses for Experiment 2. Experiment 1 indicated that task expertise, recognition, anticipation, and differential implicit and explicit processes may be confounding factors in the measurement of SA using the MHPT. A two-level process hypothesis appears to be in accordance with these results, and therefore a second purpose of Experiment 2 should be to test a two-level process hypothesis alongside Endsley's model of SA.

Hazard perception tests cannot measure the first two of Endsley's (1995) three levels of SA wholly independently. This is because in cognitive terms recognition (the basis of the recognition and anticipation measures) is a combination of the perception of information and comprehension of that information. Recognition involves a degree of top-down processing – a representation must exist in memory for a match to result in recognition or probabilistic recognition - and hence explicit awareness of the relationship between the perceived information and representations in long-term memory. From this perspective then, comprehension is the more top down counterpart to perception in visual recognition. Therefore an HPT recognition measure is to some extent measuring both perception and comprehension when it samples explicit responding with a cursor and mouse selection of the stimuli.

The assumption that the recognition measure samples mainly explicit responding is also supported by the pilot work on this thesis in McGowan (2002). Subsequent analysis of these data indicated that exposure to lectures on SA was positively associated with anticipation scores. The SA training material included a section on active scanning of the road scene which was therefore likely to have been the reason for the greater increase in anticipation scores in a perception only training condition. Analysis of that anticipation data, and the findings of Experiment 1 that revealed a significant effect of the non-SA control condition indicated that simply engaging with the material more attentively is associated with a large increase in anticipation scores. This suggests that a recognition process is highly dependent on information search and pickup, or perception and comprehension – both parts of that system being necessary for accurate explicit current awareness. Therefore it is proposed that for Experiment 2, due to the sensitivity of the anticipation measure to expertise and training this measure should be used again as a dependent measure in relation to SA training. This will serve to investigate both Endsley's three levels of SA (as process) and a two-process theory, both in terms of SA training effects on anticipation. The recognition measure will be omitted.

It is hypothesised that anticipation scores will be associated with training an anticipation strategy to 'think ahead'. In addition, due to potential complicating factors involving expertise seen in Experiment 1, when using the anticipation measure it is proposed that the participants in the present experiment should all be of a novice level of expertise. Novice or early-years drivers also tend to be well-represented among the undergraduate student population.

An alternative two-level process hypothesis will also be tested using the anticipation measure. The two-level hypothesis is that SA processing occurs through two processes that are to some degree separate. This hypothesis is that perception, comprehension, and anticipation scores will not be significantly different, or if significantly different only marginally so, when using the anticipation measure. This is because perception-action is likely to require active engagement with the environment in order for an appropriate level of anticipatory action to occur. The active engagement is likely to occur due to the perception training (visual scanning), the comprehension training (thinking about what you see), and the projection training strategy in which drivers are instructed to think ahead about what may be about to occur in the road scene. Each of these three strategies causes active engagement with the environment.

In addition, perception and comprehension are considered from a two-level process account to be two components of one system that are not additive in processing because they are interdependent processes. Therefore, from a two-level position a condition in which perception and comprehension are both trained will not be associated with different anticipation scores than the perception only training condition.

From a three-level perspective where perception and comprehension are additive then perception plus comprehension strategies will be associated with significantly higher anticipation scores than the perception only training condition. According to a three-level account this pattern of results will fit within an incremental pattern of HPT scores: starting with perception at the lowest score but greater than the no training condition; then comprehension; then perception plus comprehension; and the highest scores would be in the projection condition. Therefore, the two third hypotheses enable a second test of both the three-level and two-level hypotheses: a non-additive finding relating to perception and comprehension conditions will support the two-level hypothesis, and an additive finding in relation to those conditions when combined into a perception plus comprehension condition will support a three-level hypothesis.

Method

Participants

A total of 219 students (169 female and 50 male; mean age 23 years and eight months; mean driving experience five years and four months) took part in Experiment 2. These 219 Cardiff university undergraduate psychology students taking part in partial fulfilment of an undergraduate course requirement (N = 67; 55 female and 12 male; mean age 18 years 11 months; mean driving experience one year and eight months); for payment at Cardiff University (N = 35; 18 female and 17 male; mean age 20 years and two months; mean driving experience one year and 11 months); and undergraduate psychology students as volunteers at an Open University summer school (N = 117; 96 female and 21 male; mean age 32 years and one month; mean driving experience 12 years and six months).

The participants were required either to have a driving licence, or to have only had driving lessons in order to increase the likelihood of them being drivers with a minimum of experience, and to have normal or corrected vision for driving. Participants were randomly allocated to seven conditions.

Apparatus and Materials

The same anticipation dependent measure used in Experiment 1 was also used in Experiment 2. The same training material was also used for the complete SA training, non-SA training and a non-training control condition, as in Experiment 1. The same instructions (Appendix A-1) were used as in Experiment 1.

For the four further conditions representing ascending levels of SA the training material from Experiment 1 (Sections 2-4) was redesigned firstly by separating it into three discrete sections, one for each of the three component levels of Endsley's SA construct. Each of these three sections was analysed semantically to ensure that none of the content referred to any of the other two sections. For example, for the level one SA condition the training material was created only using terms relating to perception of information such as 'perceive', 'see', 'items', and 'looking'. This ensured that each of the three pieces of material trained a discrete cognitive strategy related to only one level of SA. Secondly, sections 2 and 3, representing perception and comprehension were then combined to produce a fourth section termed perception plus comprehension. The resulting four sections were then reproduced individually, each beginning and ending with an appropriate title, introduction, and summary.

Design

This experiment used a two factor (2 x 7) mixed design. The first factor was training and comprised two within-participants conditions: pre-training and post-training. The second factor was training type and comprised seven between-participants conditions: no training, non-SA training, SA training, perception, comprehension, perception plus comprehension, and projection.

Dependent measures

Anticipation measure. The MHPT anticipation measure used as the second dependent variable in Experiment 1 was used again as the dependent variable in this experiment. Each participant's mean anticipation score for a trial was computed excluding any zero scores.

Demographic dependent variables. The same demographic dependent measures collected during Experiment 1 were collected again during Experiment 2.

Procedure

The experiment was presented at a computer workstation. The participants were told that the experiment comprised a hazard perception test, a short paper-based exercise, and then a second hazard perception test. Demographic data were collected from the participants by using an electronic form presented on the computer screen. Participants were then presented with printed instructions (Appendix A-1). There was no practice trial in the present experiment, and the instructions did not place an emphasis on speed or accuracy.

The participants clicked the start button on the screen using the mouse. As the video began the main point in the instructions was reemphasised verbally: 'Use the mouse to click on objects in the road view where you see a hazard that would cause you to take some physical action when driving a real car.' (Appendix A-1) This verbal reminder of the instruction was performed in both the pre and post training trials. When the first video had ended the participants in the SA training and non-SA training conditions were given the paper-based training material. Participants in the no training condition were directed to take a short break from the screen for three minutes. This period was controlled in the present experiment by monitoring the moment that each participant finished from another computer and starting them on the post-training trial at the correct time. Following the rest the second video was presented in the same way as the first.

Results

The same 120 items on the HPT anticipation scale used in Experiment 1 were tested for differences in training effects between the seven training conditions. MHPT anticipation scores were calculated as the mean of hazards in a trial scoring greater than zero. From the means shown in Table 7 on page 103 anticipation scores in the post-training conditions appear to be greater than in the pre-training conditions.

A two factor (2 x 7) mixed ANOVA showed a significant interaction between training and training type, (F (1, 218) = 3.54, MSE = 16.62, $p < 0.01$, f = 0.31) in terms of anticipation scores. A repeated measures analysis of variance showed a significant main effect (F (1,218) = 5.762 , $p < .01$) of training on anticipation scores (Table 7) from the pre-training mean (5.31 seconds) to the post-training mean (7.46 seconds). There was a main effect of training type on anticipation scores (F (6,208) = 115.683, $p < 0.01$).

The effect of training type was further investigated by creating a new dependent variable from the difference between the pre and post training anticipation scores. This variable was tested between the seven training types with a one-way ANOVA which showed a significant difference (F (6,208) = 6.3, $p < 0.01$) between training types in terms of anticipation scores. A Newman-Keuls post hoc test showed that the no training and SA training conditions were significantly different ($p < 0.05$) from the other five conditions in terms of the difference between pre and post training anticipation scores. In addition, a comparison of standardised differences between pre and post training in terms of Z scores shows a similar finding: z score of differences no training z = -0.30; non-SA training z = -0.45; SA training z = 0.80; perceive z = 0.23; comprehend $z = 0.2$; perceive plus comprehend $z = .008$; and project $z = 0.1$. Both these analyses concur in failing to support the hypothesis that training SA strategies in accordance with Endsley's three levels will be associated with incremental increases in MHPT scores. The results repeat the finding in Experiment 1 that SA training is associated greater MHPT anticipation scores than no training.

	Train Type	Mean (seconds)	SD
Pre Training	No training	5.42	1.30
	Non-SA	5.11	2.51
	Full SA	5.34	1.05
	Perceive	5.71	2.26
	Comprehend	5.65	2.18
	Perceive plus comprehend	4.96	1.70
	Project	5.01	1.58
	Total	5.31	1.87
Post Training	No training	6.08	2.13
	Non-SA	5.21	2.37
	Full SA	9.11	1.34
	Perceive	8.25	3.85
	Comprehend	7.89	2.72
	Perceive plus comprehend	7.61	1.70
	Project	7.83	1.29
	Total	7.46	2.62

Table 7. Means and standard deviations of training and training type anticipation scores

Discussion

The hypotheses tested in Experiment 2 were derived from two alternative accounts of the processes involved in the acquisition and maintenance of SA. The first set of hypotheses related to Endsley's (1995) three-level theory of SA. If that theory is correct when applied to SA processing then three results are predicted to occur. Firstly, the MHPT anticipation measure will show the highest scores in the anticipation training condition. Secondly, there would be an additive finding in relation to the perception and comprehension conditions (when combined into a perception plus comprehension conditions).

In terms of these two hypotheses Endsley's three-level model was not supported by the results. Firstly, anticipation scores were not higher than scores in other conditions after anticipation training. Secondly, the effects of perception and comprehension did not appear to be additive – perception and comprehension training together did not show higher scores than perception alone.

The second set of hypotheses related to a two-level theory of a recognition-comprehension process and perception-action process. If that theory is correct then three results are predicted to occur. Firstly, there will be no difference in MHPT anticipation scores between the three SA training conditions. Secondly, there will be no evidence of an incremental perception, comprehension, perception plus comprehension, and projection pattern of anticipation explicit anticipation scores. Thirdly, a non-additive finding relating to perception and comprehension conditions when using the anticipation dependent measure.

In terms of this second set of hypotheses a two-level process hypothesis is fully supported by the results. The first and most important hypothesis was supported, in that there was no difference between three conditions after SA training in terms of anticipation scores. The second and third hypotheses were intended to continue to test the two and three level hypotheses if the first hypotheses was not supported. The second and third hypotheses are therefore redundant in light of the finding that there was no difference between SA training conditions in terms of anticipation scores.

The above findings support the hypothesis that a dual-visual process is involved in SA in terms of driving hazard perception, and that Endsley's levels of SA do not appear to be applicable to cognitive processes. This calls into question both Endsley's theory and the related SAGAT measure. A two-level process hypothesis would predict that as expertise increases processing gradually changes from recognition-comprehension to perception-action. This contrasts with Endsley's theory that increases in SA are supported by incremental processes of perception, comprehension, and projection that map onto the observed behaviours of pilots at different levels of expertise and experience. These behaviours might equally occur as a manifestation of a two-level process while telling us nothing about that process, in the same way that it is not necessary to tell us anything about a three level process. A more in-depth assessment of Endsley's levels of SA now follows in relation to the results of Experiment 2.

Firstly, an operator who tends to spend most of her time passively receiving information may well be engaging the ventral (recognition-comprehension) system at a low level of perception (Level 1 SA). Secondly, an operator who is understanding the information that she is viewing is actively engaging the ventral system in a top-down direction. From a two-level process perspective recognition and comprehension are interdependent processes, and if they draw on the same WM resources[3] then they are unlikely to be additive in effect on SA. More likely is that a shift in bias between less filtered perception and long-term memory filtered perception alters the quality of SA from concrete to abstract processing – from automatic perception-action to

3 It is intended to treat memory processes at a later stage in the thesis, when this issue is raised concerning the working memory processes that may be involved in recognition or anticipatory action SA. Nevertheless, a reading of Milner and Goodale's (1995) research indicates that recognition-comprehension draws on working memory processes while perception-action is likely to accord with conceptualisations of long-term memory, LT-WM and automaticity.

recognition-comprehension and subsequent decision-making. This is supported by the finding in Experiment 2 where there was no difference in overall explicit SA (as measured by the anticipation measure) between perception and perception plus comprehension SA training conditions.

Thirdly, a driver who actively searches for information to confirm hypotheses is engaging automatic recognition processes (i.e. pattern-matching; Groeger, 2000). These automatic processes are likely to be perception-action and appear to be used primarily for adaptive skilled anticipatory action, but they can also become explicit in the recognition-comprehension system when necessary. The reader is likely to recognise an effect associated with making a quick emergency response when driving a car through a complex and highly dynamic environment. During such automatic responses one can become consciously aware of having begun to perform anticipatory action after the act has been initiated. It is argued that this effect may be real – due to perception-action differing from recognition-comprehension in the greater speed of perception-action and the greater ability of to make decisions abstractly based-on multiple sources of information content distributed in time and space.

Thus three kinds of behaviour can be exhibited by a two-level process hypothesis and they may appear to have a positive and incremental relationship with expertise, but the underlying structure of the process need not follow that same pattern. Therefore it is firstly argued that in driving hazard perception a two-level hypothesis more accurately describes the cognitive processes involved in the acquisition of SA than does Endsley's three-level hypothesis. Secondly, that this is nevertheless still compatible with Endsley's observations of the relationship between expertise and SA related behaviours. The two approaches differ in terms of describing the cognitive processes supporting SA, in that in a two-level hypothesis recognition and comprehension are involved in one recognition-comprehension process. This process results in a similar level of conscious awareness as that which ultimately emerges explicitly from the perception-action process.

Another important aspect of a two-level hypothesis is that it would predict, due to the relationship between expertise and a shift from recognition-comprehension to perception-action, that as a visually perceived task becomes more automatic the processing tends to be processed by the dorsal system, freeing WM. It is reasonable to suggest that the dorsal system manages skilled anticipation in visually-based tasks by switching from a focus on the specific stimuli of an event to co-occurring contextual stimuli. The use of contextual stimuli is advantageous in skilled anticipation, because its probabilistic nature enables rapid heuristic calculations and hence earlier and less-effortful decision-making and action.

The perception involved in the perception-action process may be more involved in learning about novel stimuli (e.g., Rogers, Lee & Fisk, 1995) than recognition in the recognition-comprehension process. The suggestion is that recognition-comprehension is described by a more reconstructivist approach whereas perception-action is well by ecological theories of cognition. This point will be returned to later in the thesis when investigating WM.

Endsley (2000) argues that because pilots tend to show patterns of perception, then comprehension, and then projection, corresponding with increasing levels of experience then this pattern must describe the cognitive processes in SA. There is a paucity of previous evidence and no evidence from Experiment 2 to support the claim that perception, comprehension, and projection are separate processes that relate to ascending degrees of SA. It has been argued by McGuinness and Foy (2000) that we should interpret Endsley's levels more loosely as a complex network of parallel or interdependent processes. SA has been described as '...a label for a variety of cognitive processing activities that are critical to dynamic, event-driven, and multitask fields of practice' (Sarter & Woods, 1995, p.16). It is suggested that a model of the complex network of processes involved in the acquisition and maintenance of SA may be informed partly by the two-level process theory in relation to visually-based tasks.

CHAPTER 3
INTERRUPTION AND REORIENTATION

The first two experiments investigated the effects of training cognitive and perceptual strategies for SA on the MHPT. These experiment provided evidence to support the assumption that SA in driving is synonymous with hazard perception. The second experiment indicated that the cognitive processes that support SA may well reflect the same pattern of behaviours observed by Endsley in varying levels of pilots' expertise. However, the manifestation of behaviours is not necessarily evidence of the structure of supporting cognitive processes. The evidence as a whole tended to support an alternative two-level processing theory (i.e., Milner and Goodale, 1992, 1995) in visually-based tasks.

The next phase of this thesis has a closer focus on cognitive processes. Having made progress investigating the effects of training, expertise, and theories of SA, another area of investigation into SA processes began to emerge from the literature and in relation to other work in the research group on the effects of interruption (e.g., Hodgetts & Jones, 2003). This area of study arises from Flach's (1995) argument that freeze probing measures of SA are likely to have an effect on ongoing performance of complex dynamic tasks – and may further contribute to circular reasoning concerning either SA processes, products, or both. Endsley (2004) refutes this, claiming that no studies have so far found effects of interruption when using freeze probe measures.

However, in addition to interruption another possible effect of freeze probe methods emerged from the literature (Snow & Reising, 2000). A possibility exists for operators to be reoriented to the task – learning that the goal of the test is to measure awareness – and hence reorienting to that goal and subsequently performing better.

From the success in Experiments 1 and 2 in measuring manipulations of training and expertise, and finding apparently strong results concerning a viable theory of cognitive processing in SA, it was considered that the anticipation measure may be sufficiently unobtrusive to begin to reveal the likely effects of both interruption and reorientation.

Experiment 3

The situation awareness global assessment technique (SAGAT; Endsley, 1995) freezes the information interface, usually the system or simulator screen, during a complex dynamic task. Questions are then presented to the operator about aspects of the task in terms of perception of current information, comprehension of the meaning of that information, and projection to future system states. Given the effects that interruption usually has on tasks (described below), let alone complex dynamic tasks, it is reasonable to ask why it is that despite investigations no effects of interruption occur in relation to the use of SAGAT. This is of particular theoretical and applied relevance since SAGAT has been used in the development of Endsley's (1995) theory of SA, and it is one of the most widely cited measures used in assessments of operators' SA in critical tasks.

One parsimonious and frequent answer to the question of why no interruption effects have been found is that if another variable caused by the same intervention (SAGAT task) is having a reverse effect on the measure then both variables may mainly cancel out some of their effects in terms of the measure. The concept of reorientation arose in Sohn and Doane's (2000) experiment on WM and SA in pilots while using a questionnaire-based measure. Reorientation is likely to have reverse effects on SAGAT compared with interruption. Therefore, an investigation into the effects of, and relationship between, interruption and reorientation effects on questionnaire-based measures is a way to begin to investigate the effects of interruption. This is expected to lead, as did Experiments 1 and 2, closer to cognitive factors in a subsequent experiment in the same area.

Interruption. The ability to multi-task is vulnerable to a number of factors, including one of which is the effects of interruption. While there is no indication of interruption effects in interruption-based SA probes, (e.g. Endsley, 2000) the potential remains for interruptions to affect cognitive processing (Sarter & Woods, 1991). Interruption is known to limit performance at task resumption across a broad range of domains (Miyata & Norman, 1986; McFarlane, 2002) and the continuous and dynamic nature of SA would appear to make it vulnerable to disruption during resumption. Interruptions can cause a decrease in both accuracy and speed (Gillie & Broadbent, 1989), although as we shall see interruption can result in an increase in performance and some kinds of interruption are associated with variance in task performance. Interruptions may be of a cognitive origin due to for example, effortfully diverting processing resources to an alternate task and thereby allowing prospective memory about the primary task to fade. This would cause a failure attend to the primary task and maintain performance. Alternatively, the interruption may be driven by external events that are uncontrolled by the person. There is also variance in the strength of interruptions; some interruptions such as new but partially relevant stimuli only partially interrupt the main task. For example, a small incoming email notification window can engage attention momentarily and automatically and may allow the person to maintain some attention on the foregrounded task.

MacFarlane and Latorella (2002) noted that when a person's main task is interrupted the interruption does not always result in errors and that the similarity between the tasks appears to create complex effects. Similarity between the primary task and the interrupting task appears to lead to a reduction in performance (Czerwinski, Chrisman & Rudishill, 1991), and so does task complexity (Gillie & Broadbent, 1989). Interruption can cause effects other than simple performance decrease (e.g, Gillie & Broadbent, 1989; Kreifeldt & McCarthy, 1981), one being the Zeigarnik Effect (Van Bergan, 1968; cited in MacFarlane & Latorella, 2002) from the finding that recall of interrupted tasks was better than that of uninterrupted tasks. This particular finding is similar to the concept of reorientation.

The effects of interruption can continue even after task resumption has been completed – an effect that appears to be related to the degree of control over goal activation (Monk, 2004). However, there appears to be some resilience to the effects of interruption. When the person has a degree of control over the interruption, for example in terms of strategies for managing interruptions or the opportunity to be flexible in task performance then the effects of interruption can be ameliorated (Chapanis & Overbey, 1974; Lee, 1992). However, in some circumstances the effects of interruption can be positive. For example, tasks which make low cognitive workload demands permit the spare capacity to be affected by internal and external distracting stimuli but when such tasks are interrupted this appears to lead to less distraction effects on the primary task (Speier, Valacich, & Vessey, 1997). However, Endsley (2000) argues that there are no effects of interruption using the SAGAT method.

Reorientation. Snow and Reising (2000) also found no effects of interruption using a SAGAT method, but they noted that '…pilot comments also indicated that their attentional or cognitive behaviour may have been altered by the fact that the questions were asked.' (Snow & Reising, 2000, p.52). This raises the issue of whether asking the questions can have an effect on task performance by highlighting the relevancy of the questions to the task. This may not be due to the content of the questions, but simply due to interruption of the task enhancing memory content related to the task, as in the Zeigarnik Effect (Van Bergan, 1968). Such reorientation of the task could positively affect task performance and help to offset the effects of the interruption, thereby providing an explanation for the apparent lack of negative effects due to SAGAT. The opportunity therefore exists to test the effects of both interruption and reorientation using the SAGAT method. This can be achieved by manipulating firstly interruption versus non-interruption of the task, and secondly by manipulating the degree of reorientation by presenting differing levels of relationship between the SAGAT queries and the task.

An applied area of research on SA that has investigated the amelioration of obtrusive effects on the measurement of SA concerns Durso, Hackworth, Truitt, Crutchfield and Manning's (1998) measure of SA known as the situation present assessment technique (SPAM). SPAM is similar to Endsley's (1995) SAGAT method except that during the system-pause data about the situation and its context continue to be presented in order to ameliorate several known effects of interruption (see MacFarlane and Latorella, 2002). The on-screen presentation of contextual or specific data about the situation that the participant can refer to while answering SAGAT queries is thought to reduce the effects of interruption (Durso & Dattel, 2004). SPAM therefore provides an opportunity to test the effects of making the situation present against the effects of hiding the situation (using the SAGAT method). There is also an opportunity to test the cognitive correlates of interruption by manipulating the SPAM method in terms of the way in which the contextual and specific data is presented. It was necessary to investigate the effects of interruption on the SAGAT method before using the SPAM method to manipulate the way in which data is presented, since SAGAT will provide data about the effects of simple interruption.

For Experiment 3 it was therefore predicted that because hazard perception is a measure of SA then there would be a correlation between anticipation scores and SAGAT scores relating to the same hazards. Secondly, interruption effects are common in cognitive tasks, and because the anticipation task is complex and dynamic there would be a negative effect of query interruption on anticipation scores. Finally, as it is likely that the queries may reorient the operator to the task, it was predicted that relevance of the content of the SAGAT queries to the task would be positively correlated with anticipation scores. For Experiment 3 it is proposed that the anticipation measure be used first to test the above hypotheses, since this measure appeared to make clearer discriminations in Experiments 1 and 2 than did the recognition measure.

Method

Participants

One hundred and fifty two Cardiff University undergraduate psychology students (117 female, 35 male; mean age 19 years eight months; mean driving experience one year and five months) took part in partial fulfilment of a university course requirement. They were required either to have a driving licence, or to have only had driving lessons in order to increase the likelihood of them being drivers with a minimum of experience, and to have normal or corrected vision for driving.

Apparatus and Materials

HPT Measure

It was necessary to reduce the MHPT hazard item set from 120 to 70 hazards in order that four SAGAT query-type independent conditions could be presented with the interruptions. This was for two reasons. Firstly, the hazard items were required to be greater than three seconds in duration so that there was an opportunity to perceive the overt development of a situation before an interruption occurred. Secondly, the addition of interruptions increased the duration of the trial and therefore less video and less hazards were required for a given time period using SAGAT than in uninterrupted use of the video, e.g. continuous use of video in Experiment 1. The resulting video contained 70 hazard items with a mean duration of 8.91 seconds and standard deviation of 3.17. The 70 item MHPT without interruptions was 19 minutes and thirty-nine seconds in duration, and with interruptions it was 25 minutes and 59 seconds in duration.

SAGAT measure

SAGAT is intended to measure SA during a display screen simulated task. During a SAGAT assessment the simulator system is stopped at what Endsley (1995) describes as random times, and the operator is then queried using question items relating to aspects of SA derived from a task analysis. Figure 7 shows an example of an aviation SAGAT query. This query requires the pilot to state the altitude of another aircraft shown as 'T3' in the simulated circular radar screen, by moving the altitude scale (designed to be similar to a strip altitude instrument) on the left. This example shows the application of the principle that freeze techniques should employ interfaces and cognitive activities that are as similar as possible to the real task. The queries can be presented on the screen in graphical or text form during the period in which the system is stopped. That is, until the operator responds to the question, at which time the simulator then resumes. The dependent SAGAT measure is the accuracy of participants' responses. In this experiment the accuracy was scored as a multiple choice from five options, of which one was correct and scored one point and the other four scored zero. The SAGAT method was embedded within the MHPT by adapting the software so that it paused the MHPT video and presented a SAGAT query (Figure 7) within the ADI defined hazard periods. See Figure 10 for an illustration of the interruption (within participants) and reorientation (between participants) conditions.

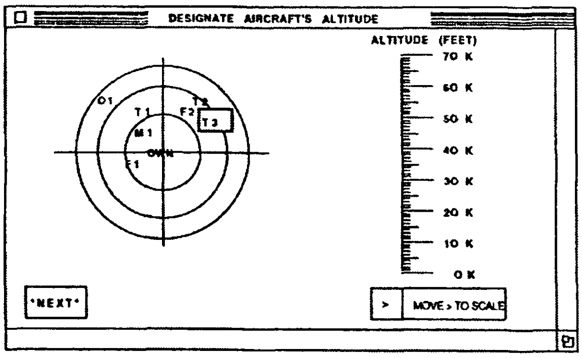

Figure 7. Example aviation SAGAT query (Endsley, 2000) showing simulated radar screen on the left and strip altitude instrument response format on the right.

Combined presentation

The combined HPT and SAGAT measures were presented using a Visual Basic program. The query presentation was at four standardised time points: Either 0.25, or 0.5, or 0.75 times the duration of the hazard, or at the end of the hazard, i.e. the full hazard duration. The four interruption options were randomised between 32 of the hazards (Figure 10: black interrupt bars) in each query type condition, so that eight hazards were interrupted at each of the four time points in the three conditions being interrupted. The remaining 38 hazard items were also uninterrupted. The no query condition was not interrupted. This design therefore provided a like-for-like comparison within and between participants (Figure 10), and between the SAGAT and MHPT dependent measures at the moment of interruption.

The Visual Basic program initiated an automated interruption-based SA probe which paused the video during the time period of each of the 32 interrupted events and presented a query. The length of the SAGAT query interruption was 10 seconds, irrespective of when a response was made. If the participant responded in less than ten seconds a counter displayed the time until the video was again displayed on the screen and continued. If the full ten seconds elapsed without a response being made then the video returned and continued automatically, therefore this was a forced choice or ipsative measure.

Query-type factor

The Query-type independent variable comprised four between-participants conditions No Query, SA Query, Orienting Query, and Irrelevant Query. Only the SA Query condition content (SAGAT) was developed in line with Endsley's (1995) prescription of a task analysis.

No Query. This condition required the participants to perform the 70 item MHPT without any SAGAT queries being presented. Therefore this video ran continuously.

SA Query. In this condition the contents of the 32 interrupted queries were developed to the specifications for SAGAT provided in Endsley (1995). The material was designed to function as SAGAT would do, although there were limitations due to the lack of information about the design of the proprietary SAGAT product. The material consisted of multiple choice queries based on perception, comprehension, and projection. For example, for projection 'where will the car in the right hand lane be in five seconds?' The question and response options were presented in one of two ways depending on the content of the query.

Firstly, for spatially important queries such as the one given above a graphical form was used, within which the response was required. On each query there were five response option radio buttons, shown in Figure 8 as black circles on a square white background. In the actual materials these buttons had the same grey background as the road surface and they were hidden from view by the participant. The smaller, grey buttons represent the positions of other vehicles at the moment of interruption. Their purpose was to mark all car locations for the researcher's reference and to enable analyses of mouse cursor location in relation to vehicles if this had become necessary.

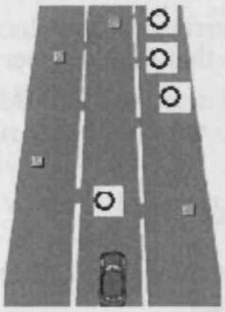

Figure 8. Graphical query showing response buttons (black circles)

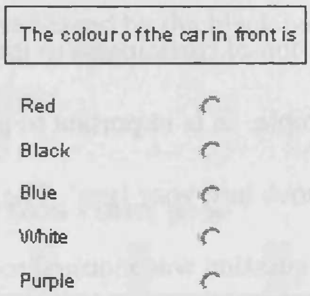

Figure 9. Standard multiple choice text-based query showing multiple choice response buttons.

In this case the radio buttons were distributed along the right hand lane, one of which was correct and the others were spread out in directions which decreased in probability based on the context. These alternate locations were mostly unoccupied space but they were occasionally located where a vehicle was. This pattern of reducing probability was determined by the researcher finding viable positions which either increased in distance from the correct location, based on the changes in the hazard over time, and by referring to the Police driving handbook (The Police Foundation, 2000) for rules relating to hazard perception.

Secondly, for queries which could not be presented graphically a form was presented using a standard five option multiple choice response format as shown in Figure 9. The order of the questions in this kind of query was randomised by the Visual Basic program. The selection of any button in Figures 8 or 9 causes the response to be logged, the query is removed and the process continues.

Orienting Query. In this condition the contents of the queries were designed to direct the participants' attention towards the same objects used in the SA Query condition, thereby giving an indication of the extent to which the redirecting of the participants' attention has an effect on MHPT scores. The responses were not sampled for accuracy. Therefore this was not a test of SA but instead the questions oriented the attention of participants to the same objects that the SA Query was probing. For example, 'it is important to pay attention to vehicles in the right-hand lane that might move into your lane'. The response was required in terms of a likert scale. The above question was required to be answered as Very Important, Important, Unsure, Unimportant, Very unimportant. This scale enabled analysis of the extent to which the participant was aware of the importance of the target item, as well as performing the function of directing attention to the item (reorientation).

Irrelevant Query. In this condition the contents of the queries were questions that were unrelated to the road environment. These questions were unrelated to road hazards but congruent with the video, e.g., 'would you buy a car based on its colour?' This condition was designed to interrupt and query the participant but with no content relative to hazard perception. The responses were not sampled for accuracy.

Design

This experiment used a two factor (2 x 4) mixed design. Figure 10 illustrates the locations at which the MHPT and SAGAT dependent measures were sampled within the two factors. The first factor (horizontal axis) was interruption and comprised two within-participants conditions: interrupt and non-interrupt, represented by the black and grey bars respectively. The second factor (vertical axis) was query type and this comprised four between-participants conditions: no query (non-interrupted throughout), SA query, orienting query, and irrelevant query. This enabled a comparison between interruption and non-interruption across both factors. The MHPT measure was sampled at all of the 70 (interrupted and non-interrupted) locations in all four query conditions. The SAGAT dependent measure was sampled only at the interrupted points represented by the black bars, for comparison with the MHPT data from the same locations.

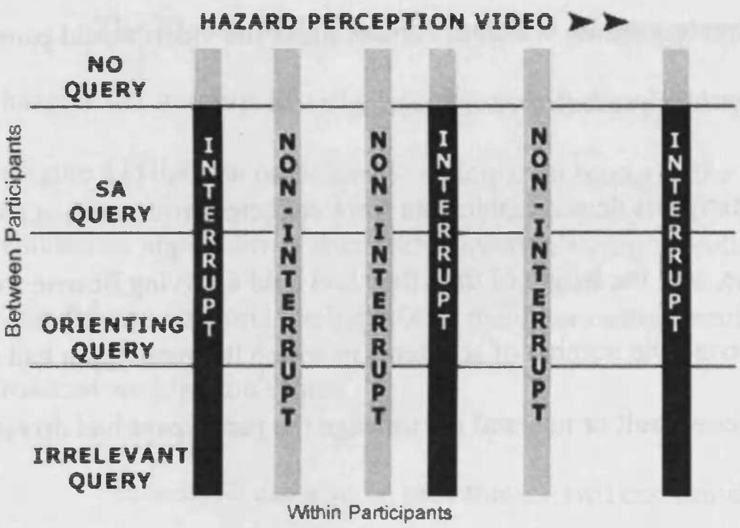

Figure 10. Interruption design: An example of the hazard events is depicted by the vertical bars, some of which (black bars) are interrupted by one of three query types. It can be seen that the between participants No Query condition is not interrupted, permitting analysis of the effects of interruption both between and within participants (McGowan & Banbury, 2004).

Dependent measures

Anticipation measure. The MHPT anticipation measure used as the second dependent variable in Experiment 1 was used again as the dependent variable in this experiment. Each participant's mean anticipation score for a trial was again computed excluding any zero scores. In this experiment Videos A and B were joined to create one 70 hazard item measure. These 70 items were sampled at the 32 interrupted and 38 non-interrupted events in all four conditions. MHPT scores were calculated as the mean of hazards scoring greater than zero.

SAGAT measure. The SAGAT measure was sampled only in the SA query condition at the 32 interrupted hazards. This measure provided accuracy scores.

Procedure

The experiment was presented at a computer workstation. For the three interruption conditions the participants were told that the experiment comprised a hazard perception test during which at various times the video would pause and be replaced by questions presented on the screen.

The participants demographic data were collected from an on-screen form: the participant's age, sex, the length of time they had held a driving license (or provisional license), the number of accidents in which the participant had been involved whether at fault or not, and the mileage the participant had driven in the preceding week.

Participants were then presented with a screen showing the instructions (Appendix A-1). After reading the instructions the participants were then directed to click a proceed button which would begin a 52 second trial which included for the Orienting Query and Irrelevant Query conditions a generic example text query which asked the participant to click option number five. During the trial the instruction 'Use the mouse to click on objects in the road view where you see a hazard that would cause you to take some physical action when driving a real car.' (Appendix A-1) was reemphasised verbally. At the end of this trial a button was presented alongside the instruction to click it when ready to begin the main trial which would last approximately 26 minutes.

Results

There was a significant positive correlation ($r = 0.58$, $p < 0.01$) between the SAGAT accuracy scores in the SAGAT queries condition (mean SAGAT score 3.59, SD 0.71) and anticipation scores (mean 6.67 seconds, SD 0.89). This finding supported the hypothesis that because anticipation is a product of SA there would be a correlation between hazard perception test scores and SAGAT queries.

Anticipation dependent measure

The 70 item MHPT was measured in all four query-type conditions at all 70 hazards (32 interrupted and 38 uninterrupted). Firstly, it can be seen from the results (Figure 11) that the participants' anticipation scores in the three interrupted conditions appear lower than each condition's corresponding non-interrupted trials. This finding supports the hypothesis that interruption would be associated with reduced anticipation scores.

Secondly, it can also be seen that the two conditions in which there were reorientation (SA query and orienting query) both show higher anticipation scores than the no query control condition, supporting the hypothesis that reorientation leads to increased MHPT scores. It can also be seen that the non-interrupted conditions within the SA query and orienting query conditions are also higher than the no query conditions. This was not expected because the non-interrupted conditions do not reorient the participants. A possible reason for this result is that because the non-interrupted and interrupted conditions are interleaved within the trial, then the reorientation effect is carrying over from interrupted to non-interrupted hazards in the same way that residual interruption effects (Monk, 2004) carry-over to non-interrupted performance. This would also explain why the non-interrupted conditions within the SA query and reorienting query conditions show higher anticipation scores than the corresponding interrupted conditions. That is, the reorientation effect is unrestrained by interruption in those two conditions. However, the same effect occurs in the irrelevant condition. This may be explained as an effect of the lack of reorientation across both the irrelevant query condition's interrupted and non-interrupted conditions. That is, there is no effect of the irrelevant reorientation condition because the query content does not reorient to hazard perception, but there is an effect of interruption in the irrelevant query interrupted condition which reduces the anticipation score below the no query baseline anticipation score.

The above pattern of data was confirmed by performing a two-factor (2x2) ANOVA which showed a significant interaction between interruption and query type (F (1,151) = 11.99, $p < 0.01$) in terms of anticipation scores. There was a significant main effect of interruption (F (1,151) = 34.29, $p < 0.01$), mean non-interrupted anticipation score 3.09 seconds and mean interrupted MHPT score 3.89 seconds. There was a main effect of query type (F (3,150) = 313.53 $p < 0.01$), no query mean MHPT score 4.34 seconds, SA query 6.46 seconds, orienting query 5.81 seconds, and irrelevant query 3.49 seconds. A Newman-Keuls post hoc test showed that all the query conditions differed in terms of anticipation score ($p < 0.05$).

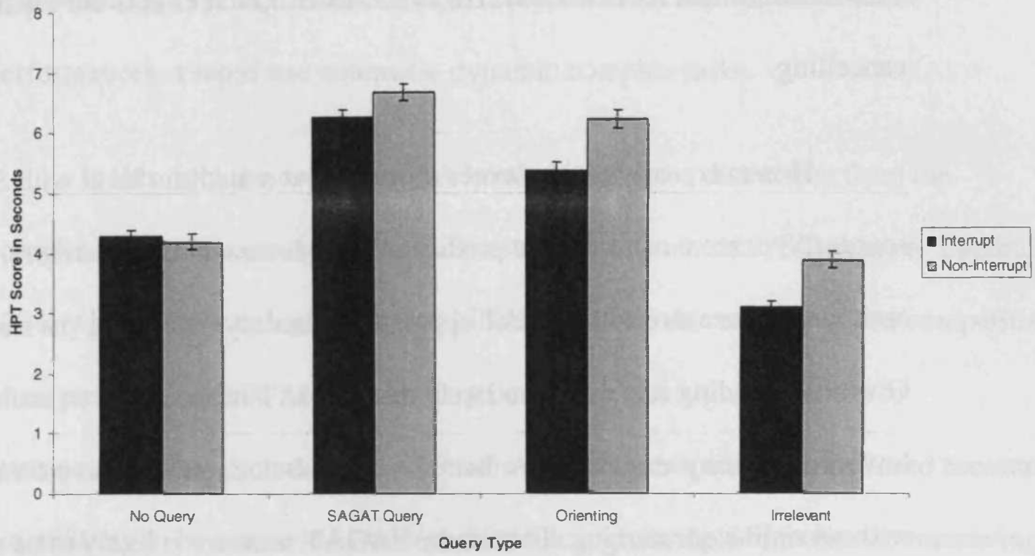

Figure 11 Means of interruption and query type on anticipation scores showing effects of interruption and reorientation.

Discussion

There were four main findings from Experiment 3 which are explained in detail on page 123 in relation to the data. The first two of these were that there appears to be a negative effect of interruption which is associated with questionnaire-based measures of SA. Secondly, there also appears to be a positive effect of reorientation due to the probe in addition to any reorientation simply due to the interruption. Together, these two effects support the hypothesis that interruption and reorientation have similar but opposite effects on a SAGAT measure.

The second two findings were firstly that residual reorientation effects tend to carry over from interrupted to non-interrupted tasks. This indicates that interruption has enduring deleterious effects on the performance of complex dynamics tasks over time. Secondly, it was also observed that where the reorientation effect was not present in the irrelevant query condition, and where interruptions occurred they caused the score to be reduced below the baseline anticipation score for no-query.

This indicates that reorientation effects and interruption effects are capable of mutual cancelling.

However, anticipation scores correlated at a medium level with SAGAT scores (58 percent of scores are predicted by both measures). Therefore, the anticipation measure and SAGAT appear to be mainly measuring the same construct. Given this finding it is therefore likely that SAGAT is providing an accurate but lower resolution measure of SA than if it were obtaining the same information without online questioning. That is, the SAGAT measure is likely to be subject to variance caused by interruption and reorientation factors and any other secondary factors related to them.

In addition, differential effects due to interruption and reorientation on cognitive processing are likely, particularly in light of evidence from Experiment 1 and 2 concerning possible multiple cognitive processes that have different functions. That is to say, interruption-based probes such as SAGAT, which tend to use a simulated situation and which require an explicit response, may not be properly sampling perception-action skilled anticipation ability. The potential therefore arises in applied human factors for experts' real skilled anticipatory SA to be under-rated, and also potentially the inverse that novices' SA will be overrated due to its comparison with experts' scores. This argument runs counter-intuitively, in that one would assume if interruption and reorientation cancel each other then the resulting score will be similar to true performance. However as introduced above, if experts and novices use different processes in the way they use awareness then there is the danger that novices' SA may be overrated by questionnaire-based methods. The risk of underrating experts' SA also fails to credit them for their often faultless performances at rapid and automatic dynamic complex tasks.

In this light the concept of SA levels is perhaps a distraction from the complexity and apparently ecologically embedded nature of SA (Dekker & Lützhöft , 2004) which pilots and other highly skilled operators, drivers, and sports competitors often present as a fluid and automatic process which is highly resistant to introspection. It is proposed that without modifications questionnaire-based measures are unlikely to be sensitive to such performance, given their limitation to measuring only the explicit part of responding. An alternative is to use methods similar to HPT's but using real task responses instead of button presses, in other words embedding HPT tests within full simulations. Another alternative is to continue to use questions but to modify their presentation such that they are as unobtrusive as possible.

An applied area of research on SA that has investigated the amelioration of obtrusive effects on the measurement of SA concerns Durso *et al*'s (1998) measure of SA known as the situation present assessment technique (SPAM). SPAM is similar to Endsley's (1995; described on p.88) SAGAT method except that during the system-pause data about the situation and its context continue to be presented in order to ameliorate expected effects of interruption. The use of on-going presentation of contextual or specific data about the situation that the participant can refer to while answering SAGAT queries is likely to reduce the effects of interruption. SPAM therefore provides an opportunity to test the effects of making the situation present against the effects of hiding the situation (using the SAGAT method). There is also an opportunity to test the cognitive correlates of interruption during SA by manipulating the SPAM method in terms of the way in which the contextual and specific data is presented.

Experiment 4

In Experiment 3 the content of SAGAT queries was manipulated in order to investigate obtrusiveness in the use of the SAGAT measure. It was found that the SAGAT method appears to interfere with anticipation performance due to both interruption and reorientation to the task. More importantly for this thesis Experiment 3 aimed to test the sensitivity of the anticipation method to these effects prior to investigating any effects of the way in which the interruption is presented in question based methods. Hence the next step in this thesis is to build on Experiment 3 by manipulating the way in which the interruption occurred while using the anticipation scores as a measure of SA, in order to investigate ways in which obtrusiveness can be reduced. It is therefore planned to use a SPAM type measure and the anticipation measure to test hypotheses concerning processing during interruption and other interferences in order to move closer to cognitive factors associated with SA.

The effects of interruption are known generally to be negative, Miyata and Norman (1986). The contributory factors include similarity between interrupting and interrupted tasks (Broadbent & Gillie, 1989), task complexity (Broadbent & Gillie, 1989), and a lack of control over the interruption and task flexibility (Chapanis and Overbey, 1974; Lee, 1992). The first two of these factors are aspects of the task itself, whereas the third relates to the way in which the interruption occurs. Therefore, it is proposed that Experiment 4 should continue to use the anticipation measure to investigate of the way in which interruptions occur, in order to explore cognitive factors in SA that are affected by interruption.

SPAM (Durso *et al*, 1998) presents operators with '...queries about the situation while the situation remains present and while they continue to perform the primary task' (Durso & Dattel, 2004, p. 134). This measure assesses SA both by sampling accuracy of the response to the query and by sampling the time taken to respond to the query, in contrast to SAGAT which samples accuracy alone. SPAM offers a comparison with SAGAT by sampling accuracy, while at the same time manipulating the way in which the interruption occurs.

The effects of interruption can be both negative and positive, depending on the nature of the interruption. In the present chapter some of these effects will be investigated in the context of the SPAM SA assessment method which aims to reduce the effects of interruption by maintaining the presentation of information about the task, and allowing the operator to continue to perform the task while querying the operator. The aim is not to test SPAM itself, because the SPAM is designed to be an online measure during 'uninterrupted' performance of the task, but to test the kinds of interruption and reorientation effects which may occur when questioning a participant while controlled elements of the SPAM method occur. Three kinds of information presentation were considered to be relevant. The presentation types and reasons for using them will now be described.

The first way to present information in order to ameliorate interruption is to give a notification that an interruption will occur and to allow the participant to accept the query when he decides to do so. This affords a level of control over the interruption. Clearly, permitting a participant to control the interruption itself admits a significant experimental problem. In addition to allowing the participant to control the trade-off between memory decay and main task demands it was not considered feasible to permit participants to continue performing the hazard perception task during SPAM queries due to the need to control the interruption externally. This is because measuring the extent to which each participant allocated attention between the SPAM query and the anticipation measure would have been beyond practical capability. Furthermore, to allow participants control in terms of variance in SPAM query response times would reduce experimental control. However, this did not preclude investigation of the effect of participants' control over the interruption to the extent that control is afforded with by the provision of a warning. For instance, a warning gives an operator an opportunity to rehearse critical representations of the task before moving attention to the interruption task. A notification is also likely to reduce any stress that may occur due to an unexpected interruption of the task. The extent to which the notification of a forthcoming interruption affects anticipation performance would be a measure of the level of control provided by notification, and more importantly this level of control can be compared to the effects of both a fully interrupted (SAGAT) condition and a non-interrupted condition in terms of anticipation scores. Therefore the first hypothesis for Experiment 4 is that notification of a forthcoming SAGAT query will reduce the effects of interruption: SAGAT scores will be greater when participants are notified of a forthcoming interruption.

However, if participants are notified of an interruption for a SAGAT query this is also likely to increase the reorientation effect seen in Experiment 3. The reason for this reorientation is that participants are likely to attend more to relevant features of the main task because they will know that they are about to be questioned on those features, thereby increasing SAGAT scores. Here there is a problem for experimental design in that a decrease in SAGAT scores due to an interruption is likely to be accompanied by an increase in reorientation due to the notification. The decrease due to interruption will be less than that in Experiment 3 and the increase in reorientation will be greater than that in Experiment 3, therefore leading to higher SAGAT scores overall than were seen when using the SAGAT measure in Experiment 3. At the same time anticipation scores will follow a similar pattern. Experimentally, this presents a problem because the SAGAT accuracy measure is likely to be of insufficient resolution required in order to discriminate between such small differences (Experiment 3 SAGAT measure standard deviation). A solution however is to infer this difference from anticipation scores post interruption in a SAGAT notification condition.

Durso and Dattel (2004) argue that it is possible to assess SA using ‘online’ measures such as SPAM by creating an interruption which is an acceptable normal secondary task. For example, in air traffic control a telephone call could be used to query an operator about aircraft positions and she would not be aware that an SA assessment is in place. The air traffic controller would have time to complete critical components of the main task and to continue to process relevant stimuli, and to intersperse processing of those stimuli while responding to the telephone call. This form of assessment is effectively an embedded task-based measure. It is nevertheless still a partial interruption of the main task which would have continued uninterrupted if the call had not occurred, and hence it is not a measure of the SA that would have occurred if there had not been a telephone call. SPAM assessment tests and operator’s SA while maintaining performance on the main task during an interruption to answer a question which although not explicitly under test conditions is nevertheless instructing the operator to engage with the current available information that he may not have engaged with if there had been no telephone call. Reorientation is therefore still likely.

The second way in which the situation may be kept present 'online' while a SAGAT query is presented is to present the system state at the moment of interruption only. That is, by interrupting the task to present a query as SAGAT does, but continuing to present the scene at the moment of interruption as an image alongside the query. This method also prevents the operator from searching ongoing states of the system to confirm what were merely hypotheses at the moment of interruption.

This prevention of hypothesis confirmation therefore provides a second hypothesis for Experiment 4, that a static image of the system at the moment of interruption presented during the interruption of the SA query will lead to both greater SAGAT scores and greater anticipation scores than in a standard SAGAT condition due to an increased reorientation effect and a reduced interruption effect. The second hypothesis is therefore that presenting an image of the moment of interruption alongside the SAGAT query be associated with higher anticipation scores in the static condition than in the SAGAT condition.

The temporal dimensions of SA are dealt with more extensively in a later chapter in the context of a cognitive streaming account of SA, leading to claims for a possible locus of some serial SA processes in phonological WM. The potential for an ongoing presentation of system states during an interruption to confirm previous hypotheses about those system states relates to the way in which SA processes are spread over time. Given that system states would have progressed up to the moment of interruption, a static image of the system at the moment of interruption is a means of confirming hypotheses about those system states. However, there are likely to be levels of cognitive hypothesis testing which span varying periods of time. Therefore it is proposed that a third method using the situation present approach is to present the scene at the moment of interruption dynamically, in order to present an approximation of the full context of the situation. One way to effect this would be to present a repeating video section of the time period leading up to the moment of interruption so that the driver can develop an awareness over time of what has happened and what may happen next. This would provide a comparison of the extent to which contextual information across time is used to infer current and future states of the system. Therefore, a third condition for Experiment 4 is proposed in which the last few seconds of the MHPT video are repeated during the interruption. It is hypothesised that repetition of these scenes will lead to greater SAGAT query scores and greater anticipation scores than in the static situation condition.

Each of the above conditions - notification, static, and replay - can be presented as modifications of the SAGAT method used in Experiment 3. As a control a fourth condition should therefore be a SAGAT condition as used in Experiment 3. Therefore, first hypothesis for Experiment 4 is that notification of a forthcoming SAGAT query will reduce the effects of interruption: SAGAT condition anticipation scores will be less than scores in the static SPAM condition.

A second hypothesis for Experiment 4 is that anticipation scores would be greater when using SPAM method of SA measurement than when using the SAGAT method; presenting an image of the moment of interruption alongside the SAGAT query be associated with higher anticipation scores in the static condition than in the SAGAT condition. In a replay condition the last few seconds of the MHPT video are repeated during the interruption. A third hypothesis is that repetition of these scenes will lead to greater SAGAT query scores and greater anticipation scores than in the static situation condition. In Experiments 3 it was observed that SAGAT accuracy scores were correlated with anticipation scores. Therefore, in Experiment 4 a fourth hypothesis was also that because hazard perception and SPAM are measures of SA there would be a correlation between anticipation scores and SAGAT scores (under conditions similar to SPAM) relating to the same hazards. A fifth hypothesis also acted as a comparison with Experiment 3. It was predicted that in Experiment 4 interrupted events would again be associated with lower anticipation scores than non-interrupted events.

A sixth hypothesis was then added that the scope of the SAGAT query content (the range of exemplars) and the number of query items (also increasing the range of exemplars viewed) would both be positively related with anticipation scores, due to their potential to increase reorientation as seen in Experiment 3. Query number refers to the number of SA queries rather than the number of times interrupted. Scope refers to the range of questions. A broader scope provides a greater range of data from which to deduce the aim of the test. It should be noted that during the implementation of SPAM in Experiment 4 the drivers will only be passively engaged in the anticipation task, in contrast to SPAM as described by Durso and Dattel(2004) which is described as being performed under more ecologically valid active tasks.

Method

Participants

One hundred and thirty four Cardiff University undergraduate psychology students (111 female, 23 male; mean age 19 years four months; mean driving experience one year and four months) took part in partial fulfilment of a university course requirement. They were required either to have a driving licence, or to have only had driving lessons in order to increase the likelihood of them being drivers with a minimum of experience, and to have normal or corrected vision for driving.

Apparatus and Materials

SPAM types. This experiment was based on the same materials as those used in the Experiment 3 SA query condition, with modifications for SPAM. The within participants SPAM type variable was required to vary content of the SA queries across four conditions. The SA queries were therefore modified as follows. For the static condition the last video frame before the interruption was presented alongside the SA query at half (400 x 200 pixels) its MHPT video resolution. For the replay condition, the three seconds of video prior to the interruption was replayed continuously alongside the SA query, in the same configuration as the static image described above. For the notify condition, as a signal to prepare for a query the background of the monitor around the video changed from black to white four seconds before the interruption. The SAGAT condition material was the same as that used for the SA query condition in Experiment 3. Therefore, there were eight of each of the four types of query.

Interruption. Interruption was manipulated in the same way as in Experiment 3. That is, the same 32 hazards were interrupted by the SPAM method and the same 38 hazards were uninterrupted.

Number. The number of queries was manipulated by replacing four of the eight queries in each SPAM condition with four irrelevant queries as described in Experiment 3.

Scope. The breadth of scope was manipulated by using questions relating to the full SA for half of the queries (broad) and by using only perception related queries for the other half (narrow)

Design

This experiment used a four factor (2 x 2 x 2 x 4) mixed design. The first factor was query scope and comprised two conditions: 1 broad scope where the range of content in the query items was greater than in the; 2 narrow scope condition. The second factor was query number and comprised two conditions: high number of queries and low number of queries. The third factor was interruption and comprised two conditions: interruption and non-interruption. The fourth factor was query type and comprised four conditions: static SPAM, replay SPAM, notify SPAM, and SA query (SAGAT).

Dependent measures

Anticipation measure. The MHPT anticipation measure used as the second dependent variable in Experiment 1 was used again as the dependent variable in this experiment. Each participant's mean anticipation score for a trial was again computed excluding any zero scores. In this experiment Videos A and B were joined to create one 70 hazard item measure. These 70 items were sampled at the 32 interrupted and 38 non-interrupted events in all four conditions. MHPT scores were calculated as the mean of hazards scoring greater than zero.

SAGAT measure. The SAGAT measure was sampled only in the SA query condition at the 32 interrupted hazards. This measure provided accuracy scores.

Procedure

The experiment was presented at a computer workstation. For all four conditions the participants were told that the experiment comprised a hazard perception test during which at various times the video would pause and be replaced by questions presented on the screen.

The participants demographic data were collected from an on-screen form: the participant's age, sex, the length of time they had held a driving license (or provisional license), the number of accidents in which the participant had been involved whether at fault or not, and the mileage the participant had driven in the preceding week.

Participants were then presented with a screen showing the instructions (Appendix A-1). After reading the instructions the participants were then directed to click a proceed button which would begin a 52 second trial which included for the orienting query and irrelevant query conditions a generic example text query which asked the participant to click option number five. During the trial the instruction 'Use the mouse to click on objects in the road view where you see a hazard that would cause you to take some physical action when driving a real car' was reemphasised verbally. At the end of this trial a button was presented alongside the instruction to click it when ready to begin the main trial which would last approximately 26 minutes.

Results

There was a significant positive correlation ($r = 0.43$, $p < 0.018$) between the SAGAT accuracy scores in the SPAM queries condition and anticipation scores. This finding supported the hypothesis that because anticipation is a product of SA there would be a correlation between hazard perception scores and SPAM queries. This correlation is similar to the same comparison in Experiment 3 but at a lower value.

Anticipation dependent measure

The results (Figure 12) do not support the first hypothesis, in that notification of a forthcoming SAGAT query does not appear to be associated with greater MHPT scores than the SAGAT condition. The second hypothesis was that anticipation scores would be greater when using the SPAM static method than when using the SAGAT method. This result was not supported by the results. There appears to be no difference between the static and notification conditions. The third hypothesis was that repetition of scenes will lead to greater SAGAT query scores and greater anticipation scores than in the static situation condition. This prediction was not supported by the results. The fourth hypothesis was that there would be a correlation between anticipation scores and SAGAT scores (under conditions similar to SPAM) relating to the same hazards, this correlation is apparent in the results. A fifth hypothesis also acted as a comparison with Experiment 3. For the fifth hypothesis it was predicted interrupted events would again be associated with lower anticipation scores than non-interrupted events. This result is evident in the large difference between interrupted and non-interrupted anticipation scores. The sixth hypothesis was that the scope of the SAGAT query content and the number of query items would both be positively related with anticipation scores. Figure 13 indicates that breadth of query type is negatively related with anticipation scores. Figure 14 does not indicate a positive association with number of queries.

The results of this experiment were tested using a (2 x 2 x 2 x 4) mixed ANOVA. There were several interactions. There was a highly significant interaction (F (3,133) = 9.63, $p < 0.01$) of interruption, SPAM query type, query scope, and query number, in terms of anticipation scores. There was a highly significant interaction (F (2,133) = 60.72, p < 0.01) of interruption, SPAM query type, and query number, in terms of anticipation scores. There was a highly significant interaction (F (2,133) = 20.94, p < 0.01) of interruption, SPAM query type, and query scope, in terms of anticipation scores. There was a highly significant interaction (F (1,133) = 12.15, p < 0.01) of interruption and SPAM query type in terms of anticipation scores. There was a highly significant interaction (F (2,133) = 9.40, p < 0.01) of SPAM query type, query scope, and query number, in terms of anticipation scores. There was a highly significant interaction (F (1,133) = 41.07, p < 0.01) of SPAM query type and query number, in terms of anticipation scores.

There was a highly significant main effect (F (1,133) = 119.42, p < 0.01) of interruption on anticipation scores, mean non-interrupted anticipation score 5.44 seconds SD 1.08, mean interrupted MHPT score 4.69 SD 0.90. There was no significant main effect (F (3,130) = 0.82, p < 0.53) of SPAM query type on anticipation scores.

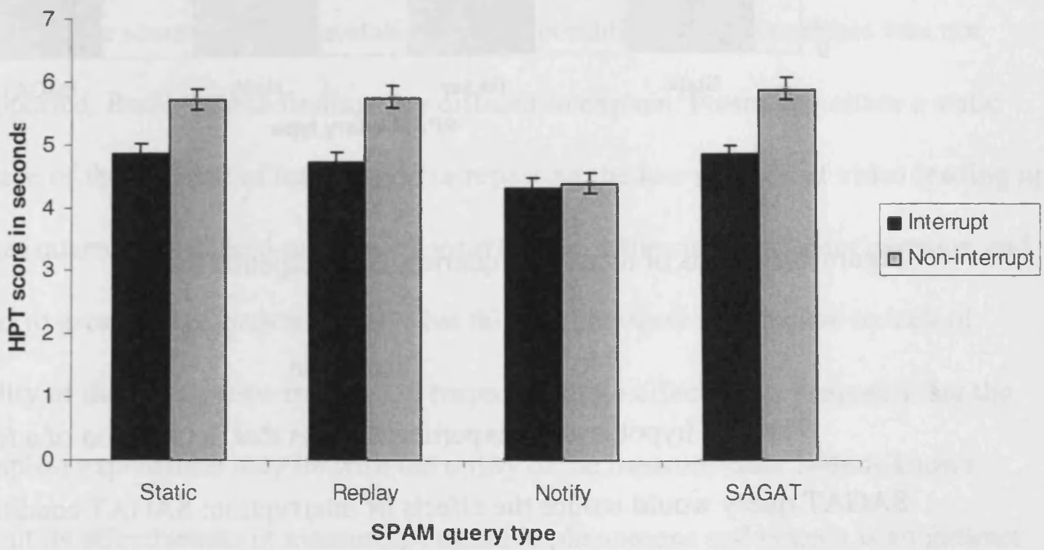

Figure 12. Means of interruption on anticipation scores

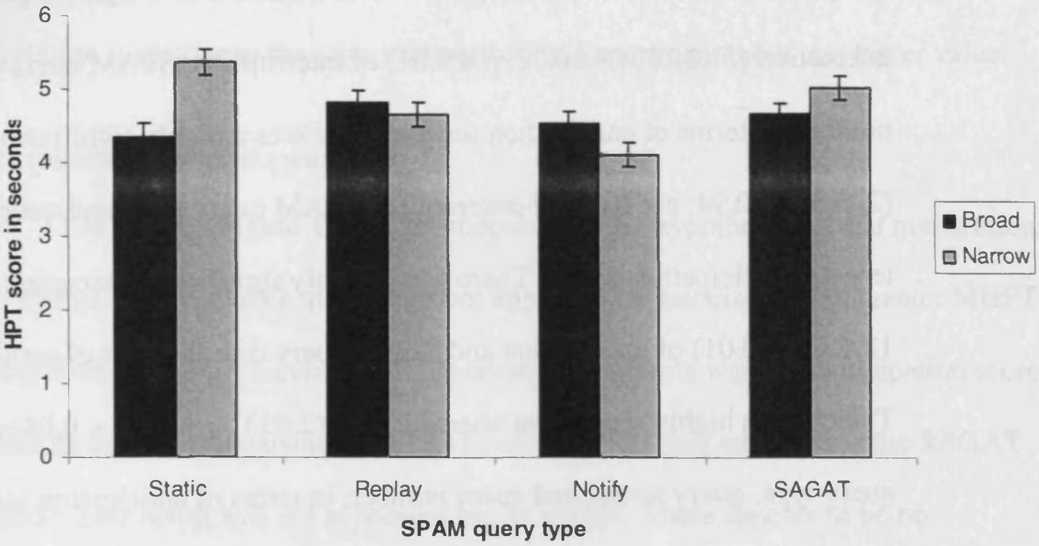

Figure 13. Means of query type breadth on anticipation scores

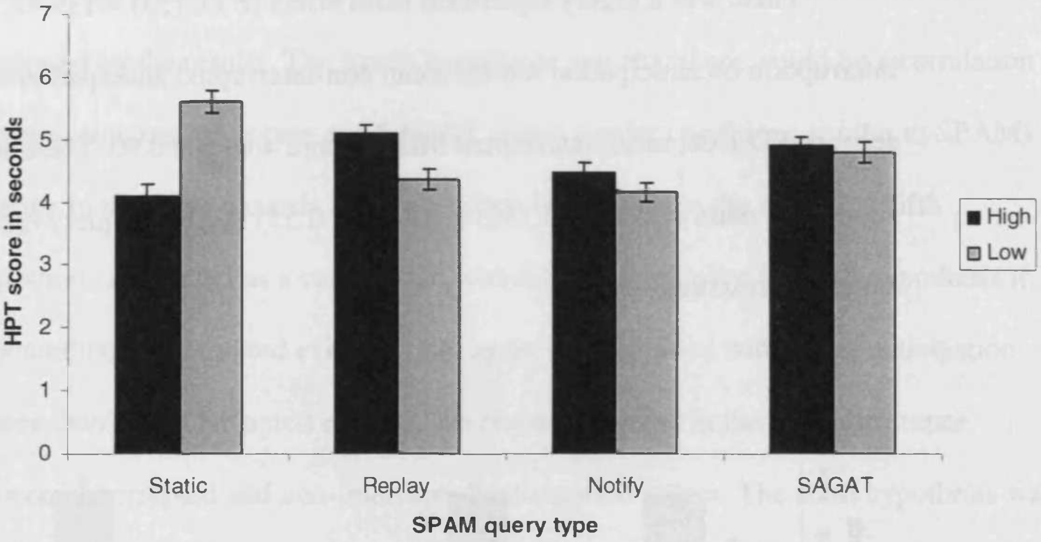

Figure 14. Means of number of queries on anticipation scores

Discussion

The first hypothesis in Experiment 4 was that notification of a forthcoming SAGAT query would reduce the effects of interruption: SAGAT condition anticipation scores would be less than scores in the static SPAM condition. This was not supported in the results which showed that the notification condition non-interrupted events anticipation scores were similar to the roughly equal scores in all the interrupted conditions. This means that notifying drivers of forthcoming interruptions does not reduce the effects of those interruptions, and that notification may be involved in some general effect on functioning. Some possible explanations follow. Notification of an interruption in driving is a rare occurrence, if it occurs at all. Stress and the internal disruption that may be caused by signalling a forthcoming interruption and the need to prepare for it could potentially be as strong as those effects due to the interruption itself. Therefore, an explanation of this unexpected result is that notification causes other effects that interfere with the anticipation measure over time to an extent which is greater than any benefit from the provision of an advanced warning.

The second hypothesis was that anticipation scores would be greater when using SPAM method of SA measurement than when using the SAGAT method. This prediction was not supported by the results. The third hypothesis was that repetition of video scenes leading up to the interruption would be associated with greater anticipation scores than in the static situation condition. This hypothesis was not supported. Both of these findings are difficult to explain. Presenting either a static image of the moment of interruption or repeating the few seconds of video leading up to an interruption should provide information about the situation under question and lead to greater anticipation scores. That this did not occur may be due to lack of utility of the anticipation measure in respect of these effects. It is proposed that the simplest explanation may lie with the utility of the measure since little is known about its effectiveness in measuring cognitive phenomena and since it is an indirect measure of skilled anticipatory action.

Rowe and McKenna (2001) used a test of anticipatory action – a button press when predicting tennis volleys – but like the MHPT it is also an indirect test of anticipatory action. With other variables potentially mediating the button response it would be unwise to put too much faith in the MHPT as a direct measure of specific factors. There is a potential for a number of effects when presenting images, and from the dynamic nature of the presentations such as MHPT responding, then a notification or image or video, and the knowledge that the MHPT task will return. Any of these uncontrolled potential effects may have been involved in moderating the manipulated effect.

However, despite the possibility that the MHPT measure is relatively untested, two results occurred which confirmed that at a broad level the MHPT seems robust. A fourth hypothesis was that because hazard perception and SPAM are measures of SA there would be a correlation between anticipation scores and SAGAT scores (under conditions similar to SPAM) relating to the same hazards. This result confirmed the same correlation in Experiment 3 and showed that SPAM methods do not appear to affect anticipation scores differentially to SAGAT. A fifth hypothesis was that interrupted events would be associated with lower anticipation scores than non-interrupted events. This result is evident in the large difference between interrupted and non-interrupted anticipation scores. Both this result and the correlation above indicate that the SPAM measure is behaving similarly to the SAGAT measure in Experiment 3. This indicates that the anticipation measure is probably robust as a measure. At the level of discriminating the effects of the way in which information is presented the anticipation measure appears to be less powerful.

The results of SPAM manipulations showed little useful data. Experiment 4 did not show a difference between the anticipation scores of the SPAM interrupted conditions and the SAGAT condition; the three SPAM variants appear not to be associated with greater anticipation scores than the SAGAT condition. This suggests that continuing to present information about an interrupted event may not ameliorate the effects of an interruption. Given the argument introduced in the discussion of Experiment 3 that even deeply embedded and unobtrusive measures are nevertheless still sampling performance at a moment of task switching (interruption to the rest of the task) to perform an action, then the above finding is less unexpected. It may be that the SPAM approach limits obtrusive effects but it has reached a lower threshold in what it can achieve as a measure. That is, a measure which uses a response only to some actions and not others is only measuring what is occurring at those measured events. What occurs at other times is not being sampled. A corollary to this effect is that even an embedded simulator measure of SA is only sampling performance at moments when some action is being taken – but never when action is *not being taken*. Therefore, there is a lower threshold to the measure due to the limitations in sampling more subtle variations between other events: a resolution problem. To sample real SA it would be necessary to sample recognition and anticipation under all circumstances, not only when engaged in some action thought to be associated with SA. Thus all measures of SA are to some extent biased towards sampling awareness at certain moments and not others.

It is proposed that the reason SPAM and SAGAT scores were not differentiated by the anticipation measure is that the threshold for discriminating those effects appears to lie somewhere between the effect sizes of interruption and reorientation, and the effects of varying the way information is presented during an interruption. Residual interruption and reorientation effects observed in Experiment 3, and possibly in Experiment 4 due to notification, do appear to be clearly discriminated using the anticipation measure. Therefore it appears that either the anticipation measure is more sensitive to residual effects or these effects are stronger than those of varying the way information is presented during an interruption. Effect size estimates reproduced in the results confirm this pattern. A consequence of these findings is that continuing to make a situation present could have limited utility in contrast with the potential for reducing the more powerful effects of interruption and reorientation. This could be achieved by adopting more embedded and hence less obtrusive measures.

The findings in relation to the sixth hypothesis are unsurprising when considering the likely effect size of number of queries and variety of queries – expected to be more subtle ways in which the presentation of information during queries may be varied. The sixth hypothesis was that the scope of the SAGAT query content and the number of query items would both be positively related with anticipation scores, due to their effect of enhancing reorientation. Query number refers to the number of SA queries rather than the number of times interrupted. Scope refers to the range of questions. A broader scope provides a greater range of data from which to deduce the aim of the test. While there were interactions between query scope and other variables, no main effect of query scope or number was observed. It therefore appears that the resolution of the present version of the anticipation measure is sufficient to test the effects of interruption and reorientation, and residual interruption effects but not more subtle effects of the way that information content is presented during an interruption.

It had been expected that Experiment 4 would uncover possible cognitive variables due to manipulations of the way information is presented during interruptions. However, due to the failure to find useful data in that respect, this chapter ends with a focus on the measurement of cognitive factors in SA using the MHPT, which was one of the main aims of the thesis. The knowledge gained about the measure so far in this thesis is likely to be useful in applying the measure directly to cognitive theory in Chapter 4.

CHAPTER 4
WORKING MEMORY SUPPRESSION

The first two experiments in this thesis focused on measuring SA unobtrusively using the MHPT, as a test of the effects of the related concepts of training and expertise, and as a test of Endsley's three-level theory of SA in the context of drivers' SA for dangerous road situations. The third and fourth experiments tested SA using the MHPT in the context of interruption and reorientation. General arguments emerged from those four experiments primarily in relation to SA theory and alternative accounts of SA, but also including a dual visual systems account of visual processing which illustrated the abstract nature of the MHPT as a measure of recognition-comprehension and perception-action.

Existing SA theory will be revisited in the general discussion but preceding that the remaining experiments will now approach cognitive theory in terms of memory while using the MHPT as a measure of SA, both in terms of recognition and anticipation. The aspects of memory which will be investigated include both short-term and long-term memory, and the potentially different memory processes involved in perception-action and recognition-comprehension that were raised in earlier chapters. In addition a specific perspective on memory processes in SA, that of cognitive streaming, will be approached.

Experiment 5

In Experiment 4 emerged as in Experiment 3, in which interruption and reorientation effects appeared to be mutually cancelling. That is, maintaining the presence of a situation seems also to interfere with SA. These underlying and more complex factors were detected by using the MHPT as an alternative and unobtrusive measure of SA while manipulating interruption, reorientation, and the way in which information is presented during SPAM. It is therefore proposed that further investigation of the presentation of information should now be carried out in order to link the factors in Experiments 3 and 4 to cognitive theory. In Experiment 4 the manipulations were of different kinds of information presentation – static image, replayed video, and notifications of forthcoming queries – all information which potentially interferes with SA processing and SAGAT responding. Experiment 5 should therefore aim to investigate the cognitive factors involved in WM interference during the MHPT task in order to attempt to expose the mechanisms which SPAM may be affecting. In addition, such a manipulation of WM aims to provide data which will further inform the findings of all the previous experiments in this thesis.

One starting point for the investigation of memory processes in SA is to investigate short-term memory during SA acquisition. The most salient aspect of short-term memory in terms of visual hazard perception is memory for visual-spatial information, particularly that associated with Baddeley and Hitch's (1974) visual-spatial slave store. Firstly however, it is necessary clearly to define the hazard perception task and the relationship between the hazard perception task and visual-spatial WM. That is, the hazard perception task performed without real or simulated movement control (as is the MHPT and all previous forms) is not necessarily the same task as the real hazard perception task performed while driving. The MHPT requires a singular response via a computer mouse-screen interface and so it is different from the steering, brake, accelerator and other responses required during real driving. This situation is related to the difference between automatic perception-action and recognition-comprehension, in that explicit recognition-comprehension may be compatible with a button press but automatic perception-action requires the response to be the same as in normal control of the car or system.

Milner and Goodale's (1992) theory of dual visual processing routes or channels relates to the way the MHPT measures recognition and anticipation. Milner and Goodale's theory is that there are two partly distinct visual systems. These two systems relate closely to ecological (direct perception) and reconstructivist (indirect perception) accounts of cognition. Milner and Goodale (1995) argue that a perception-action visual system supports direct perception theories and a recognition-comprehension system supports indirect perception theories. Neurologically the system for perception-action is associated with the dorsal visual stream and the system for recognition-comprehension appears to be associated with the ventral visual stream. The extent to which these systems are related to short-term memory is unknown but it is reasonable to assume that information in the direct perception-action route is less explicitly accessible than that in the indirect recognition route, and that recognition will be associated with short and long-term memory representations.

There are three reasons for describing the dual visual systems theory. They are firstly that the dual theory relates to the MHPT method of measuring recognition and anticipation in a non-control simulation that uses a different interface from real driving control. That is, caution must be applied when making assumptions about particular short-term memory processes in anticipatory performance. If the neurological pathways that support recognition-comprehension and perception-anticipatory action differ then the cognitive processes supporting recognition and anticipation may also differ. For Endsley (1995) to assume that recognition supports anticipation is to neglect existing theories of differences between implicit and explicit performance and theories such as Goodale and Milner's which argue that the process and functional architecture differ between perception-action and recognition-comprehension.

Explicit SA is that which can result in conscious reporting of task-relevant aspects of a situation, and implicit SA results in similar levels of appropriate action but with a lack of declarative conscious awareness of the situation. The former may be accessed via query based instruments such as SAGAT while the latter tends to be accessible through measures of anticipatory performance such as the MHPT and Rowe and McKenna's (2001) measure of anticipatory performance in tennis. The dual visual process theory permits an analysis based on both these kinds of data, and may enable an understanding of the contributions of factors involved in explicit and implicit SA processing.

A finding of Experiment 1 was that expertise was not associated with anticipation scores. This may be due to experts' hazard perception being an automatic anticipatory action process which is not sampled by the MHPT. It is also argued that SA *processes* may not be characterised by Endsley's three levels because a measure of conscious recognition (MHPT recognition measure) did not distinguish the effects of SA training strategies. The measure of recognition could not achieve this but the anticipation measure did distinguish between types of training strategy. Thus, conscious perception-recognition as describe by Endsley's model may not be a necessary process in SA, particularly during expert performance in which conscious awareness of all the stimuli relating to the task tends not to relate to implicit SA which is measured through performance.

Implicit SA which leads directly to performance but which is non-declarative tends to have been neglected in SA research. Many researchers (this includes Endsley to some extent) have taken a literal approach to the measurement of SA, in that they have almost exclusively focused on measuring explicit and conscious information content that the operator can articulate; Banbury *et al*, 2000, p. 94)

Therefore the second reason for invoking the dual visual process theory is the failure of many researchers to recognise the distinction between explicit and implicit SA. It is argued that further examination of implicit performance during SA measurement are more likely to assist in uncovering the full spectrum of processes that support SA than for example Endsley's (1995) theory. Endsley's theory of SA assumes short and long-term memory processes and interactions between them underpin each of the three levels of SA in varying degrees and combinations. If some forms of SA lead directly and automatically to action, then what of comprehension? This is especially difficult to overcome due to Endsley's (2004) position that all three levels of SA are necessary. This position is predicated on the conclusion that SA is processed at these three levels. In turn this conclusion was drawn upon evidence that behaviours related to the three level SA model are represented along a scale progressively from perception, to comprehension, and then to projection, and the subsequent claim that these levels represent the levels of SA a person can report or act upon. This argument admits circular reasoning in that it is based on a construct that emerged from the same data that gave rise to the construct. The argument also mistakes an association between levels of expertise and behaviours related to the three theorised levels of SA.

The three levels of SA model has been used to argue that there are distinct processes involved in perception, comprehension, and projection. Endsley (2004) disclaims such notions, instead stating that the processes involved in SA (situation assessment) are complex, but little further is offered in terms of a model of process, other than as shown in Figure 1. What can be said of Endsley's model is that it describes the behaviours that appear to be related to expertise and SA.

A third reason for invoking the dual visual process theory is that it may assist in developing predictions for the intended WM experiments in the present chapter. An opportunity now arises in relation a the finding in Experiment 1 which appeared to show that perception and anticipation are more important than is comprehension. It is proposed that an experiment may be constructed in which the recognition and anticipation components of the MHPT are tested separately as in Experiment 1, while manipulating WM processes. Therefore, WM will now be discussed in relation to SA, hazard perception, dual visual process theory, and finally in relation to an alternative account of cognitive processing in SA known as cognitive streaming (see Banbury *et al*, 2004).

According to Baddeley and Hitch (1974) short-term memory is a WM process comprising a central executive store which is not specific to the characteristics of the information that it carries. The central executive subsumes a visual-spatial sketchpad store and a phonological loop store. A visual-spatial sketchpad holds representations relating to visual and spatial information and a phonological loop holds auditory and serial information. According Baddeley and Hitch's account of short-term memory the visual-spatial information content in the MHPT task is processed by the visual-spatial sketchpad.

An experiment performed by Robbins, Anderson, Barker, Bradley, Fearneyhough, Henson, Hudson and Baddeley (1996), using chess as a primary task, showed that a secondary task which suppressed the visual-spatial sketchpad significantly reduced the quality of chess moves compared with a secondary task which suppressed the phonological loop. Therefore, this provides a clear experimental hypothesis in that MHPT performance should also be negatively affected more by a visual-spatial sketchpad suppression secondary task than by a phonological loop suppression secondary task.

However as discussed above, there may be two routes through which visual information is used: one for conscious recognition of a situation, and another which is an automatic route to action. The recognition-comprehension route is related to the concept of explicit SA, and the automatic route to anticipatory action is related to implicit SA. That is, SA can be considered to be present if correct actions are shown to have been taken. Therefore, both measures embedded in the MHPT task, recognition and anticipation, should be used as dependent measures. This can be implemented through Robbins *et al*'s (1996) suppression paradigm as recognition and anticipation closely relate to recognition-comprehension and perception-action respectively.

Predictions for the effects of visual-spatial suppression on both recognition and anticipation are firstly that conscious recognition in terms of the recognition measure, which requires a degree of conscious processing in the short-term memory visual-spatial sketchpad store, will be reduced more by a visual-spatial suppression secondary task than by phonological WM suppression.

Secondly, recognition scores will be reduced more by a visual-spatial suppression task than will anticipation scores which permit automaticity through a direct perception-action route. It is likely that automatic anticipation in HPT performance does not have as strong an effect as real-world responding in a car. However, performance on the MHPT anticipation measure is still likely to be less negatively affected by visual sketchpad suppression than recognition, all else being equal. That is, explicit recognition-comprehension is more likely to draw on visual-sketchpad resources than anticipation which uses at least some automatic resources.

Baddeley and Hitch's model of WM is only one of several that are relevant to cognitive processes in SA. Ericsson and Kintsch (1995) proposed a theory which combines long-term memory with WM as long-term working memory (LT-WM). According to this theory, in expert performance WM actively accesses long-term memory representations, thereby supporting decisions and actions by using WM representations of perception combined with long-term memory representations of experience. This theory has been invoked by Durso and Gronlund (1999) as an account of cognitive processes in SA.

Another account of cognitive processes in SA is that of cognitive streaming. Cognitive streaming argues that skilled anticipation is based in serial phonological WM where probabilistic relationships between long-term memory representations and current WM information content inform predictions of future events. That is, based on past experience in terms of WM, events can be tracked, and a probabilistic representation of future events can be developed. Cognitive streaming is compatible with LT-WM in that it also argues that there is an interaction between short-term and long-term memory. However, in contrast with LT-WM cognitive streaming provides a clear explanation for the structure of anticipation processing.

Predictions are difficult to derive from LT-WM theory due to a lack of specificity about precisely how WM and long-term memory interact in relation to expertise and lack of expertise. In contrast cognitive streaming offers clearer hypotheses. A central aspect of cognitive streaming is that it is based on transitional probabilities. Transitional probability refers to the possibility of an event occurring given the presence of events that have been related with that event in the past. That is, representations in short-term memory are compared with the strength and number of relationships between similar representations in long-term memory in order to anticipate future events. Another relevant part of the theory is that the processes are carried serially in phonological WM. According to cognitive streaming WM for serial information is the critical process for the maintenance of SA whether the memory content is visual or phonological in nature. Therefore, a third and alternative hypothesis to the first is that phonological loop suppression will negatively affect both recognition and anticipation scores more than will visual-spatial suppression.

The hypotheses for the present experiment are therefore firstly that recognition scores will be more negatively affected by visual-spatial suppression than by phonological suppression. Secondly, recognition scores will be more negatively affected by visual-spatial suppression than will anticipation scores relative to phonological suppression. Thirdly, an alternative hypothesis to the first is that phonological loop suppression will negatively affect both recognition and anticipation scores more than will visual-spatial suppression. In addition to testing these hypotheses it is proposed to use suppression of Baddeley and Hitch's (1974) theorised central executive process as a comparison with visual-spatial and phonological suppression in terms of their effects on recognition and anticipation dependent measures. It is predicted that central executive suppression will have a more negative effect on recognition scores than on anticipation scores because of the likelihood of automatic responding being involved in anticipation responding. That is, the central executive – although not clearly specified in terms of process – is considered to be an aspect of controlled rather than automatic processing.

Method

Participants

Sixty-eight Cardiff University undergraduate psychology students (49 female, 19 male; mean age 19 years eight months; mean driving experience one year and five months) took part in partial fulfilment of a university course requirement. They were required either to have a driving licence, or to have only had driving lessons in order to increase the likelihood of them being drivers with minimal experience, and to have normal or corrected vision for driving.

Apparatus and Materials

Visual-spatial scratchpad condition. The apparatus used for the visual-spatial concurrent task was the computer keyboard integral to the computer on which the MHPT was presented. The relevant part of the keyboard was the numeric pad on the right side of the keyboard. Therefore, the keyboard was turned through 90 degrees so that the right hand end of it was facing the participant's non-dominant hand; in the orientation shown in Figure 15. A cardboard box with an open end facing the participant was placed over the participant's non-dominant hand where it rested on the numeric pad of the keyboard. A piece of cloth was fixed to the open end of the box to hang over the participant's wrist so that the participant could not see where she was placing her fingers on the keyboard. The number lock key (top right of Figure 15) was fixed so that it could not be pressed. The keyboard was connected to the computer. The Visual Basic MHPT program which presented the video was adapted also to record the identity of every key press and the time of each key press.

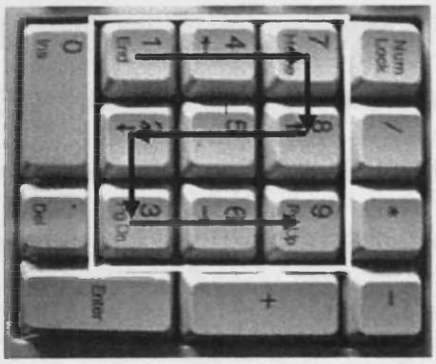

Figure 15. Numeric portion of keyboard (within white border) showing the first half of the key pressing sequence superimposed. The second part of the sequence was 9,8,7,4,5,6,3,2,1, after which the first part of the pattern repeated.

Phonological loop condition. The apparatus for the phonological loop concurrent task was a recording of sound containing irrelevant meaning that was been made by N.Perham, Human Factors Research Group, Cardiff University in 2003. The sound on the recording was of the kinds of noise usually encountered in an office environment such as telephones ringing, bleeping from machines as they are being operated, rustling of paper, and including human voices distorted so as to make the meaning of the words indiscernible. The recording was nine minutes and 49.5 seconds in length (half the length of the MHPT presentation). There were two versions of this material for counterbalancing. The first version was nine minutes and 49.5 seconds in length, and the second version presented no sound for the first nine minutes and then the irrelevant sound was presented in the second half that was also nine minutes and 49.5 seconds. The sound was presented to the participants using a mini-disc player which was connected wirelessly to six pairs of headphones.

Central executive condition. The apparatus for the central executive condition concurrent task was a second computer located behind the participant onto which the researcher entered the digits the participant verbally on a spreadsheet in real-time. A pair of headphones were worn by the participant to reduce interference by the researcher entering the digits on the second computer.

Design

This experiment used a two factor (2 x 4) mixed design. The first factor was the suppression variable and comprised two within-participants conditions: suppression and non-suppression. The second factor was suppression type and this comprised four between-participants conditions: no suppression, central executive suppression, visual-spatial suppression, and phonological loop suppression.

Dependent measures

Recognition dependent measure. The dependent anticipation MHPT measure used a 47 item abbreviated version of the 70 item hazard set (Videos A plus B) used in Experiments 3 and 4, reduced to 19 minutes and 39 seconds in order to bring its duration within 20 minutes for use in a prolonged dual task experiment. This video was clipped mainly at the beginning in order to select the maximum number of items within 20 minutes. This video clip was therefore continuous and was 19 minutes and 39 seconds in length.

Anticipation dependent measure. The MHPT anticipation measure also used the 47 item hazard set as described above.

Analysis of responding. The task of generating regular random digits and the visual spatial task of producing a pattern on the numeric portion of the keyboard were sampled. Randomness, response frequency and accuracy were thus calculable from this data.

Procedure

The experiment was presented at a computer workstation. For the three suppression (interference) conditions the participants were told that the experiment involved performing two concurrent tasks. They were told that these two tasks were firstly an HPT and secondly the concurrent task in the condition they had been allocated to was described. The concurrent tasks were described as listening to a recording (phonological loop), pressing keyboard keys (visual-spatial scratchpad), and verbally generating random digits (central-executive). The central executive condition was performed by only one participant at a time.

The participants' demographic data were collected from an on-screen form: the participant's age, sex, the length of time they had held a driving license (or provisional license), the number of accidents in which the participant had been involved whether at fault or not, and the mileage the participant had driven in the preceding week. Participants were then presented with a screen showing the MHPT instructions, and the instructions for the concurrent task they were to perform. For the participants who were to perform the concurrent task in the second half of the trial (counterbalanced design) the instruction was modified to tell the participants to be prepared for the concurrent task to start at approximately 10 minutes after the MHPT starts. To signal the beginning of the concurrent task the participants who would perform the central executive suppression task would begin when the researcher tapped the participant on the shoulder. To signal the beginning of the visual-spatial concurrent task the researcher would tap the box housing the keyboard. The phonological loop concurrent task would start automatically approximately ten minutes after the MHPT started. The instructions for the concurrent tasks were as follows.

Central executive suppression condition. The instruction (Appendix A-2.1) was given to generate a randomised sequence of the digits 1-9 verbally at the rate of approximately one digit per second, and that this speech would be recorded by the researcher who would enter it on another keyboard. They were then directed to perform this task as a trial (without the MHPT running) and were corrected until they performed it to the specification.

Visual-spatial suppression condition. The instruction was given to press the numeric keys in a particular sequence using the non-dominant hand (Figure 15): 1,4,7,8,5,2,3,6,9,8,7,4,5,6,3,2,1. They were then directed to perform this task as a trial (without the MHPT running) and were corrected until they performed it to the specification.

Phonological loop suppression condition. Participants were told that there would be sound presented through the headphones which they should ignore.

Non-suppression condition. This condition was performed in the same way as in the previous experiments.

After reading the instructions (Appendix A-1) the participants were then directed to begin the 52 second MHPT trial. During the trial the instruction 'Use the mouse to click on objects in the road view where you see a hazard that would cause you to take some physical action when driving a real car.' was reemphasised verbally. At the end of this trial a button was presented alongside the instruction to click it when ready to begin the main trial which would last approximately 20 minutes.

Results

Recognition Dependent Measure

There appears to be a decrease in terms of mean recognition scores from 33.32 in the within-participants non-suppression condition to 28.04 in the suppression condition (Table 8)

There also appear to be differences in the between participants suppression type conditions in terms of mean recognition scores. The three suppression type conditions appear to decrease in comparison with the no-suppression condition (mean recognition score 34.03) in the following order: phonological (mean recognition score 29.16); central executive (mean recognition score 27.48); and visual-spatial (mean recognition score 22.98).

A 2 x 4 mixed ANOVA was computed on within-participants suppression and suppression type in terms of recognition scores. This ANOVA showed a significant interaction between suppression and suppression type (F (1,67) = 7.51, $p < 0.01$). Effect size estimate (η^2) = 0.15. There was a significant main effect of suppression (F (1,67) = 10.09, $p < 0.01$, $\eta^2 = 0.2$). There was a significant main effect of suppression type (F (3,64) = 62.14, $p < 0.01$, $\eta^2 = 0.33$). A Newman-Keuls post hoc test revealed that the recognition scores in the non-suppression condition (mean recognition score 33.29) where significantly greater ($p < 0.05$) than in both the phonological suppression condition (mean recognition score 30.77) and the central executive suppression condition (mean recognition score 30.79). The recognition scores in both the phonological and the central executive suppression conditions were also significantly greater ($p < 0.05$) than in the visual-spatial suppression condition (mean recognition score 27.88).This indicates that visual spatial suppression has a more deleterious effect on MHPT scores than either central executive or phonological loop suppression, and that central executive or phonological loop suppression have a more deleterious effect than non-suppression conditions.

	Suppression condition	Mean Recognition	SD
Unsuppressed	Visual-Spatial	32.77	2.28
	Central Executive	34.06	2.31
	Phonological	32.43	3.06
	No Suppression	34.03	3.14
	Total	33.32	2.70
Suppressed	Visual-Spatial	22.98	5.65
	Central Executive	27.48	2.54
	Phonological	29.16	6.84
	No Suppression	32.54	2.64
	Total	28.04	4.42

Table 8. Means of suppression and suppression type on recognition scores

Anticipation Dependent Measure

In contrast to the recognition scores there appears to be no distinct pattern of anticipation scores, nor is there any pattern resembling the experimental hypothesis (Table 9). However, the no-suppression condition appears to be associated with the highest anticipation scores (overall mean 5.05 seconds), as would be expected.

A 2 x 4 mixed ANOVA was performed on within-participants suppression and between participants suppression type in terms of anticipation scores. This ANOVA showed no significant interaction between suppression and suppression type (F (1,67) = 1.06, $p < 0.37$, $\eta^2 = 0.024$). Therefore there is no clear indication of differences between suppression conditions and suppression-type conditions. This result indicates that the anticipation measure may not be sensitive to the effects of WM in direct comparison with the recognition measure.

	Suppression condition	Mean (seconds)	SD
Unsuppressed	Visual-Spatial	4.89	1.25
	Central Executive	4.98	1.64
	Phonological	4.7	1.49
	No Suppression	5.18	1.34
	Total	4.94	1.43
Suppressed	Visual-Spatial	4.28	1.54
	Central Executive	4.35	0.9
	Phonological	3.25	1.59
	No Suppression	4.91	1.67
	Total	4.2	1.42

Table 9. Means of suppression and suppression type on anticipation scores

Analysis of responding

The central executive concurrent task of generating regular random digits and the visual spatial task of producing a pattern on the numeric portion of the keyboard were, notwithstanding the control imposed by the instructions, under the control of the participant. These responses were tested below for the degree to which the participants followed the instructions in these two tasks.

Central executive suppression task, randomness

The participants were instructed to generate a series of random digits at a regular pace of one per second. Therefore the resulting series of random digits was tested for the degree to which participants followed these instructions and the possible effects on the MHPT due to non-compliance. The possibility of using Maurer's test of randomness, which would have determined the locations and boundaries of patterns (and hence enabling the identification of any relationship between random digit generation and hazard locations) was precluded by its technical complexity. Therefore a chi square test of frequencies was computed on the all the participants' digit generation series in the central executive task condition. The chi square test showed no significant difference between the digits (one to nine) in terms of their frequencies ($\chi^2 = 1.74$, $p = 0.226$).

Central executive suppression task response frequency

The total number of random digit responses was examined for its descriptive statistics. The participants were instructed to generate the random digits at a steady pace of approximately one digit per second. Therefore a mean of 590 digits may have been expected during the nine minutes 49.5 seconds of the central executive condition. The mean number of digits generated was 369.4 (SD = 76.34), a lower number than was expected. To discover the degree of uniformity in number generating rates within the participants' series across the hazard periods and non-hazard periods, the number of digits which were generated within the hazard periods was compared with the number between the hazard periods. The number was adjusted for the difference in duration between these two kinds of period by reducing the between hazard number of digits by the ratio of hazard duration (28.7 percent) to between hazard duration. Within the hazard periods the adjusted mean number of digits was 45.91 (SD = 13.27) and between the hazard periods the adjusted mean number was 92.55 (SD = 6.81). This indicates that the participants were producing fewer digits during the hazard periods.

Visual spatial suppression task errors

The visual spatial task involved pressing a specific series of numeric keys. The key presses were analysed for deviations from this pattern. There were mean 36.16 errors where the pattern of presses did not match the series, and 273.45 correct presses where the pattern was identified. Therefore there were 9.79 percent errors. To discover the degree of uniformity within the participants' series across the hazard periods and between hazard periods the percentage of errors was calculated after adjusting for the difference in duration between these two kinds of period. Within the hazard periods the percentage of errors was 6.71 percent and between the hazard periods the percentage was 8.26 percent. As with the central executive condition there were expected to be 590 presses (one per second). The mean number of presses was 217.74. To explore the uniformity in button pressing rates in the participants' series within hazard periods and between hazard periods the number of presses were calculated after adjusting for duration. The mean number of presses within hazards was 32.19 (SD = 9.27) and between hazards it was 79.6 (SD = 14.93). A paired samples T test was performed on these data and showed a significant difference ($t = 2.3$, $p = 0.022$). This indicates that participants were pressing more of the keys between the hazards than during the hazards.

Discussion

The hypotheses for the present experiment are partly supported by the results. The hypotheses that were not supported appear not to have been tested due to a lack of sensitivity of the anticipation measure. It was firstly predicted that recognition scores would be negatively affected more by visual-spatial suppression than by phonological suppression. This pattern was clearly evident in the results.

A second prediction was that recognition scores would be negatively affected more by visual-spatial suppression than anticipation scores would be, relative to phonological suppression. This prediction was not supported due to the lack of an interaction between suppression and suppression type when using the anticipation dependent measure, so that no statistical comparison could be made between those results and the results obtained when using the recognition scores as a dependent measure. There is an indication from this result that the anticipation measure may not be sensitive to the effects of WM suppression, in contrast with the recognition measure which showed a clear pattern of results in agreement with experimental hypotheses. This appears to be related to a similar pattern of results in Experiment 6, after reporting which the reasons for the lack of effect will be discussed.

The third hypothesis was an alternative to the first hypothesis. It was predicted that phonological loop suppression would negatively affect both recognition and anticipation scores more than would visual-spatial suppression. Using the recognition measure the opposite effect was observed. However, the lack of an interaction when using the anticipation measure does not permit a conclusion regarding the effect on that dependent measure. The reason for this hypothesis was that according to a cognitive streaming account of SA awareness develops over time and cognitive objects are linked serially through the phonological or articulatory loop. Hence suppression of processing by the phonological loop should affect both the recognition and anticipation measures of hazard awareness which spans time. It would appear however that in terms of at least the recognition measure the results support Baddeley and Hitch's (1974) visual-spatial sketchpad theory. Of course it remains to be seen whether, if the anticipation measure had been sensitive to the manipulations in this experiment, the phonological or the visual-spatial suppression conditions would have been associated with lowest scores. It is proposed that the lesser effect of phonological suppression should be further explored in the next experiment.

The clear pattern of results using the recognition measure indicate that in terms of recognition of hazards there is no relationship between hazard perception and the cognitive streaming account of SA. However, explicit recognition using the MHPT is likely to be the poorest measure of SA, followed by a measure of anticipation, and ultimately by fully task-embedded performance measures. Therefore the lack of utility in discriminating between Endsley's (1995) model of SA, a cognitive streaming account of SA or other accounts of memory processes is not problematic at this stage of investigation. This is to revisit the argument earlier in this thesis that the most accurate measure of SA is that which samples anticipatory controlled and automatic behaviour without influencing that behaviour. The success in measuring explicit recognition of hazards but not anticipatory responses in the present experiment may reflect the inability of the MHPT to measure fully automatic responding. This, together with Baddeley and Hitch's (1974) theory of WM being distinct from automatic and LTM processes, would explain why the recognition measure distinguished between types of WM processes but the anticipation measure did not. SA appears to be embedded in action (and hence the task), as is described by several theories in the main introduction to this thesis. It is therefore proposed that the recognition score is a measure of explicit processing whereas task performance measures (of which the anticipation measure is a simpler abstraction) are more likely to sample real controlled and automatic SA processing.

A fourth hypothesis in the present experiment was that central executive suppression would have a more negative effect on recognition scores than on anticipation scores. This is due to the likelihood that automatic responding is involved more in anticipation responding than in recognition responding. When comparing the results in terms of the two dependent measures no apparent difference is seen between their suppression and no-suppression conditions. The absence of evidence in the anticipation condition, due to the lack of an interaction, does not permit a firm conclusion.

The failure to achieve clear findings in the present experiment does not necessarily require that it be performed again. Also, to proceed along the route of attempting to investigate WM effects using the anticipation measure could narrow the focus of the present thesis. The recognition measure is sensitive to WM manipulations whereas the anticipation measure appears not to be useful data, and the pattern of scores when using the recognition dependent measure are of significant interest in partly confirming a Baddelian view of WM during SA. The results may also support the proposition made in the present chapter that HPT's and even the MHPT may not be fully sampling awareness in relation to a task. That is, the anticipation measure is probably unable to sample the automatic responding that is likely to underlie anticipatory action. The implication is that an MHPT that is fully embedded in a high fidelity simulator or real task may be necessary in order to achieve sensitivity to WM effects. It is also interesting to note that in comparison with WM effects the anticipation measure does appear to be sensitive to training, expertise, interruption, and reorientation. This comparison alone justifies the present experiment, suggesting that either WM is not associated with anticipatory responding or that the anticipation measure mouse-cursor interface is inappropriate for sampling some underlying cognitive aspects of hazard perception. In addition the potential exists for extending the investigation of WM effects by testing other manipulations such as prospective memory or episodic buffer. Such manipulations may help to shed further light on the results of the present experiment.

At this point, having earlier referred to the likelihood of a dual system of visual processing being involved in visual SA processing it is necessary to explicitly recount the evidence that relates to a dual system which has emerged in this thesis so far. Firstly, in Experiment 1 it was argued that since expertise was not associated with anticipation scores then this indicates that experts' responding may be automatic and specific to the control interface for the task, i.e. car controls. However, the explicit response recognition did show effects due to expertise. This suggests that there is an automatic route from visual perception to physical responses, and a controlled recognition route from visual perception to conscious awareness.

Secondly, in Experiment 1 it was found that perception of information and anticipation appear to be more important in SA than is comprehension. This provides more support for the position that there are two main processes involved in the development and maintenance of SA, reflecting recognition-comprehension and perception-action. The lesser influence of the comprehension training condition in Experiment 2 indicates that while there may be comprehension processing occurring this is probably less critical in terms of hazard perception than is recognition of the items. This argument suggests that training drivers in comprehension strategies has less effect simply because in the conscious visual route the important factor is to perform pattern matching, and this occurs automatically simply by scanning the environment. Any top-down influence of drawing on memory content in order to consider the meaning of a situation is probably less directly involved than is perception – hence refuting Endsley's position that comprehension is a necessary step before projection can occur.

The above evidence, in contrast with Endsley's (1995) theory which is founded mainly upon observations of pilots' behaviours, tends to suggest that a cognitive theory of SA is more likely to reflect the involvement of processes such as a dual visual system or Baddeley's three component theory of memory. There is also less evidence of the involvement of processes reflecting a cognitive streaming account of SA.

There is more cognitive theory pertaining to the development and maintenance of SA than can be dealt with in this thesis, such as the theory of LT-WM. However, the mechanisms underlying LT-WM appear to be better described by Baddeley's (2000) revised model of WM which integrates WM and LTM through the episodic buffer. As such Baddeley's model concurs with Ericsson and Kintsch's (1995) theory of LT-WM. However, due to the greater explanatory power of Baddeley's (2000) WM model aspects of LT-WM in SA will be approached by proxy through Baddeley's model. Therefore, it is proposed that the aim for Experiment 6 should be to continue to test Baddeley's WM theory in relation to Baddeley's (2000) episodic buffer component of WM, as this was associated with a distinct pattern of results when using the recognition measure.

The lack of utility when using the anticipation measure in Experiment 5 was attributed to the likelihood of anticipation being an automatic process, responses from which are unlikely to be sampled with sufficient fidelity by the MHPT mouse-cursor interface, in contrast with real-world responses to hazards when controlling a vehicle. This indicates that the anticipation measure may not be a sensitive measure of WM. However, there is an opportunity to suppress another aspect of WM which may yet show effects on the anticipation measure. Baddeley and Hitch's (1974) three component model of WM has recently been supplemented with a fourth component (Baddeley, 2000) known as the episodic buffer which connects Tulving's original (1989) episodic LTM with the central executive.

Experiment 6

Experiment 5 failed to support the hypothesis that according to a cognitive streaming account of SA, phonological WM suppression would have a greater negative effect on MHPT scores than visual-spatial WM suppression. This hypothesis may be approached again but in terms of episodic WM, in particular the episodic buffer WM component.

Until recently episodic memory for events located in time (e.g. Tulving, 1989) was theorised to exist as an LTM representation. Baddeley (2000) has proposed an adjunct of episodic LTM known as the episodic buffer. The episodic buffer is similar to the visual-spatial sketchpad and phonological loop, both of which are also forms of buffer between the central executive and LTM. However, the episodic buffer is neither specific to perceptual mode nor to WM or LTM. According to Baddeley (2000) visual and phonological processes integrate to in the episodic buffer so that the information is co-coded. The episodic buffer is proposed as an explanation of how information is transferred, managed and integrated between the LTM, the central executive, the visual-spatial sketchpad, and the phonological loop slave stores. The function of a buffer is to exchange information (both bottom-up and –down) in an efficient way. Thus the episodic buffer functions in the recall of LTM representations and in conscious control and awareness. Baddeley conceptualises it as:

> [...]a limited capacity temporary storage system that is capable of integrating information from a variety of sources. It is assumed to be controlled by the central executive, which is capable of retrieving information from the store in the form of

> conscious awareness, of reflecting on that information, and, where necessary, manipulating and modifying it. The buffer is episodic in the sense that it holds episodes whereby information is integrated across space and potentially extended across time... It can be accessed by the central executive through the medium of conscious experience. The executive can, furthermore, influence the content of the store by attending to a given source of information, whether perceptual, from other components of working memory, or from LTM (Baddeley, 2000, p.421)

Therefore the episodic buffer, argued by Baddeley to be involved in conscious control and awareness, appears to be relevant to an investigation of the cognitive processes which contribute to SA. Furthermore, it is reasonable to propose that the episodic buffer may be the process supporting SA during critical tasks such as driving hazard perception. The complexity of the episodic buffer component of the WM model as described above by Baddeley, particularly the integration of serial and spatial aspects of SA, may explain why the pattern of results in Experiment 5 did not match the predictions derived from Baddeley and Hitch's (1974) model of WM. The greater complexity may also explain why differential results were found when using the recognition and anticipation measures.

So far in this thesis visual and phonological aspects of WM have been investigated due to their close relationship with the serial and spatial nature of SA. However, if Baddeley's (2000) WM model is correct then both the serial and spatial aspects of SA may be processed in combination, thereby leading to different predictions when using the two MHPT measures of SA.

The episodic buffer account accords to some extent with the cognitive streaming account of SA described on page 130. However, cognitive streaming argues that information is coded serially, whereas Baddeley's (2000) working memory theory argues that serial and spatial information is co-coded. The results from Experiment 5 indicate that phonological-serial processing alone is not central to cognitive streaming.

Phonological working memory without rehearsal is thought to span a period of a few seconds, whereas awareness of situations can often span minutes or hours. Span is used here to describe associative processes between bits of information that are interdependent across time. For example, a driver on a motorway who is sees a road sign indicating that a junction is ahead can anticipate and prepare for the required manoeuvres many seconds or even minutes before reaching the junction. This anticipation requires a comparison of information between that provided by the sign, other information that is perceived on the subsequent approach to the junction, and LTM representations about previous similar situations.

SA is a phenomenon which is associated with information distributed across a particular and co-occurring set of spatial and temporal locations. Therefore, the serial (temporal) nature of SA is unlikely to be processed by phonological working memory alone, visual-spatial working memory processes must also be associated with the situation. The episodic buffer is a process in which spatial and temporal information are to some extent co-coded. SA may depend on phonological working memory but Baddeley's (2000) WM theory would argue that it is in combination with visual-spatial working memory and LTM that optimal SA can be achieved. It is therefore proposed that according to Baddeley (2000) suppression of the episodic buffer will be associated with a greater reduction in SA than phonological suppression. Phonological suppression may only inhibit a short span of information processing and rehearsal whereas episodic buffer suppression is likely to inhibit the overall processing of conscious awareness across the full temporal span of information required to maintain SA.

The characteristics of the episodic buffer mean that in order to suppress its function it is necessary to cause processing that involves memory information content which is related to events that are distributed over space and time. The main aspects of episodic processing are, according to Baddeley's description above, that it is controlled by the central executive, it produces information for the central executive in the form of conscious awareness. The exchange of information between the episodic buffer and the central executive is also two-way, in that the central executive can influence and modify the content of the store.

Therefore, suppression of episodic processing capacity requires a task in which information relating to past events is recalled and consciously processed by the central executive. Onslow (2003) used such a method of episodic memory suppression in which participants were required to read a story and recall details about the story after the main trail. The main trial dependent measure in that experiment was the MHPT anticipation measure, and the results indicated that the measure was sensitive to the effects of the episodic manipulation. This paradigm will permit a comparison of the effects of central executive, phonological, and episodic processing on the MHPT measures. The following experiment used this design to test the hypothesis that episodic buffer suppression will have a more deleterious effect on MHPT scores than phonological processing.

Method

Participants

Eighty-seven Cardiff University undergraduate psychology students (65 female, 22 male; mean age 19 years five months; mean driving experience one year and five months) took part in partial fulfilment of a university course requirement. They were required either to have a driving licence, or to have only had driving lessons in order to increase the likelihood of them being drivers with a minimum of experience, and to have normal or corrected vision for driving.

Apparatus

Phonological loop and central executive suppression conditions. The same apparatus was used in the phonological loop and central executive suppression conditions as those used in the same conditions in Experiment 5.

Episodic buffer suppression condition. Episodic buffer suppression was manipulated by presenting 62 images for 500 milliseconds superimposed onto the centre of the computer screen concurrent with the MHPT, the participant being instructed to try to recall an event in her life of which the image reminded her. Five hundred milliseconds was chosen as the presentation period after piloting, in order that the participant would gain a visual representation and memory trace without presenting the image for any longer than necessary, to reduce the likelihood that the participant would be distracted from watching the MHPT video. The images were presented during the hazard period (31 images) and between hazard periods (31 images) at specific times to enable analysis of the responses in relation to the MHPT structure. The times were at 0.25, or 0.5, or 0.75 of each hazard period and between hazard periods of the MHPT. These time options were randomised between all 62 periods (the total of within and between hazard periods) in order to reduce the likelihood that the relationship between the presentations and the MHPT hazards would be recognised by the participant. The images were presented in the centre of the screen immediately above the 400 x 800 pixels MHPT video. The images were 150 pixels by 150 pixels. They were obtained from a compact disc of approximately 7000 photographs produced for promotional purposes such as websites and brochures. These images were generally evocative of life in the 1980's, a period that was very early in, or before, the lifetimes of most of the participants. Sixty-two images were selected at 'random' from the complete file list which was ordered by file size. The kinds of photographs ranged from images of business activities and industry, artistic images of landscapes, family activities, sports, and other diverse events.

For the episodic buffer condition, at the end of the trial the Visual Basic MHPT program was adapted to present a screen which required responses about what they remembered during the experiment, one in a numerical field and two in a block of text fields. These text fields were arranged in two columns and 50 rows. The left hand row was titled Image, and the right hand row was titled Memories.

Design

This experiment used a two factor (2 x 3) mixed design. The first factor was suppression and comprised two within-participants conditions: suppression and non-suppression. The second factor was suppression type and this comprised three between-participants conditions: phonological loop suppression, episodic buffer suppression, and non-suppression.

Dependent measures

Recognition dependent measure. The dependent anticipation MHPT measure used a 47 item abbreviated version of the 70 item hazard set (Videos A plus B) used in Experiments 3 and 4, reduced to 19 minutes and 39 seconds in order to bring its duration within 20 minutes for use in a prolonged dual task experiment. This video was clipped mainly at the beginning in order to select the maximum number of items within 20 minutes. This video clip was therefore continuous and was 19 minutes and 39 seconds in length.

Anticipation dependent measure. The MHPT anticipation measure also used the 47 item hazard set as described above.

Analysis of responding. The task of generating regular random digits and the number of recalled memories were sampled. Randomness, response frequency and accuracy, and the extent to which participants followed the memory instructions were thus computable from this data.

Procedure

The experiment was presented at a computer workstation. For the three suppression conditions the participants were told that the experiment involved performing a main task and a secondary task. They were told that these two tasks were firstly an HPT and secondly the concurrent task in the condition they had been allocated to was described. The concurrent tasks were described as listening to a recording (phonological loop) and remembering events from the past (episodic buffer condition). At this stage the participants in the episodic buffer condition were told that during the experiment they would be asked to recall a wide range of events from the past which are at this point unknown to the researcher, some of which may be of a personal nature. At the end of the experiment they would be asked to describe the memories but that this was also subject to the above qualification. They could choose not to try to remember or report any of the events if they felt uncomfortable about doing so. They were also told that if at any time during the experiment they began to feel uncomfortable due to the evocation of the memories, then they should stop the tasks and withdraw from the experiment.

The participants' demographic data were collected from an on-screen form: the participant's age, sex, the length of time she had held a driving license (or provisional license), the number of accidents in which the participant had been involved whether at fault or not, and the mileage the participant had driven in the preceding week. Participants were then presented with a screen showing the MHPT instructions (Appendix A-1), and the instructions for the secondary task they were to perform.

General instructions. For the participants who were to perform the secondary task in the second half of the trial (counterbalanced design) the instruction was modified to tell the participants to be prepared for the secondary task to start at approximately 10 minutes after the MHPT starts. To signal the beginning of the secondary task, the participants who would perform the episodic buffer suppression task would begin when the researcher tapped the participant on the shoulder. The phonological loop secondary task would start automatically approximately ten minutes after the MHPT started. Condition-specific instructions are described below.

Episodic buffer suppression instructions. The participants were told that images would appear very briefly in the centre of the MHPT video. When an image appeared they were to recall non-verbally any events in their lives that the image reminded them of. They were not instructed to attempt to remember the image itself.

At the end of the MHPT trial in the episodic suppression condition the following instructions were given on the screen. Firstly, the participant was instructed to enter into a numerical field on the screen an estimate of the total number of memories they had recalled. Secondly, the participant was instructed to enter into text fields a very brief description or title of each image and the associated memory they had recalled which had also prompted a memory from their life. They were instructed to 'describe the images you saw in column A and describe the memory in column B' They were again advised that they were not required to describe any memories if they were uncomfortable about doing so. These instructions remained on the screen while the data were collected.

Phonological loop and central executive suppression instructions. These conditions were performed in the same ways as in Experiment 5.

Results

Recognition dependent measure

There appeared to be an increase in terms of mean recognition scores from 29.68 in the within-participants suppression condition to 32.33 in the non-suppression condition. There also appear to be differences between the three suppression type conditions in terms of mean recognition score (Table 10).

A 2 x 3 mixed ANOVA was performed on within-participants suppression and between participants suppression type in terms of recognition scores. This ANOVA showed a significant interaction between suppression and suppression type (F (1,86) = 5.74, $p < 0.01$, $\eta^2 = 0.75$).

There was also a significant main effect of suppression (F (1,86) = 13.55, $p < 0.01$, $\eta^2 = 0.64$) and a significant main effect of suppression type (F (2,84) = 3.49, p =

0.03, $\eta^2 = 0.33$). A Newman-Keuls post hoc test revealed that the recognition scores in the phonological suppression condition were significantly greater than in the episodic buffer suppression condition ($p < 0.05$). This supports the hypothesis that episodic buffer suppression has a greater effect on recognition scores than phonological processing.

	Suppression condition	Mean Recognition	SD
Unsuppressed	Central Executive	32.70	3.28
	Episodic buffer	32.44	4.8
	Phonological	31.87	5.0
	Total	32.33	4.36
Suppressed	Central Executive	29.19	3.73
	Episodic buffer	27.33	4.52
	Phonological	32.54	5.16
	Total	29.68	4.47

Table 10. Means of suppression and suppression type on recognition scores

Anticipation dependent measure

There appeared to be an increase in terms of mean anticipation scores from 3.74 in the within-participants suppression condition to 4.65 in the non-suppression condition. There also appear to be differences between the three suppression type conditions in terms of mean anticipation score (Table 11).

A 2 x 3 mixed ANOVA was performed between suppression (suppression and non-suppression conditions) and suppression type in terms of anticipation scores. This ANOVA showed a non-significant interaction between suppression and suppression type (F (1,86) = 1.97, p = 1.43, $\eta^2 = 0.23$). Therefore there is no statistical support for the hypothesis that episodic buffer suppression would have a greater negative effect on the anticipation measure than phonological suppression However, although not statistically significant there nevertheless appears to be a pattern of results in line with the hypothesis and the pattern is similar to that obtained using the recognition measure.

	Suppression condition	Mean (seconds)	SD
Unsuppressed	Central Executive	4.48	2.81
	Episodic buffer	4.7	3.28
	Phonological	4.77	3.01
	Total	4.65	3.03
Suppressed	Central Executive	3.93	1.86
	Episodic buffer	2.66	1.32
	Phonological	4.64	3.37
	Total	3.74	2.18

Table 11. Means of suppression and suppression type on anticipation scores

Episodic buffer suppression task

The secondary task in the episodic buffer suppression condition was to recall non-verbally any events in the participant's life of which the image reminded them. To assess whether there were differences between the hazard periods and non-hazard periods in terms of number of memories recalled, a measure of the specific images remembered at the end of the trial was taken. Irrespective of variance caused by primacy and recency effects this would provide some degree of inference about whether the participants had been performing the task correctly as a secondary task rather than switching to it as a main task in the non-suppression conditions and as a secondary task during the hazard periods when ADI defined hazards were present. The participants' responses to the on-screen question 'describe the images you saw…' were assessed by looking for a clear indication that she was referring to only one of the images.

The result was that in the hazard periods there was a mean of 2.61 (SD = 0.37) correctly remembered images and 4.78 (SD = 0.56) correctly remembered images. A paired samples T-test was performed on these data and showed a significant difference ($t = 1.69$, $p = 0.037$). This indicates that participants were remembering more images during the between hazard periods than during the hazard periods.

Central executive suppression task, randomness

The participants were instructed to generate a series of random digits at a regular pace of one per second. Therefore the resulting series of random digits was tested for the degree to which participants followed these instructions and the possible effects on the MHPT due to non-compliance. The possibility of using Maurer's test of randomness, which would have determined the locations and boundaries of patterns (and hence enabling the identification of any relationship between random digit generation and hazard locations) was precluded by its technical complexity. Therefore a chi square test of frequencies was computed on the all the participants' digit generation series in the central executive task condition. The chi square test showed no significant difference between the digits (one to nine) in terms of their frequencies ($\chi 2 = 1.02$, $p = 0.094$).

Central executive suppression task response frequency

The total number of random digit responses was examined for its descriptive statistics. The participants were instructed to generate the random digits at a steady pace of approximately one digit per second. Therefore a mean of 590 digits would be expected during the nine minutes 49.5 seconds of the central executive condition. The mean number of digits generated was 428.7 (SD = 53.18), a lower number than expected. To discover the degree of uniformity in number generating rates within the participants' series across the hazard periods and non-hazard periods, the number of digits which were generated within the hazard periods was compared with the number between the hazard periods. The number was adjusted for the difference in duration between these two kinds of period by reducing the between hazard number of digits by the ratio of hazard duration (28.7 percent) to between hazard duration. Within the hazard periods the adjusted mean number of digits was 37.3 (SD = 9.72) and between the hazard periods the adjusted mean number was 107.61 (SD = 11.18). This indicates that the participants were producing less digits during the hazard periods.

Discussion

The main hypothesis for Experiment 6 was that episodic buffer memory suppression would have a greater negative effect on MHPT scores than phonological processing. This hypothesis was supported when using the recognition measure and not when using the anticipation measure, similar to the result from Experiment 5 where only the recognition measure showed effects. There is an indication from this result that the anticipation measure may not be sensitive to the effects of working memory suppression, in contrast with the recognition measure which again showed a clear pattern of results in agreement with the hypothesis. This is in contrast with training and expertise effects in Experiments 1 and 2 which were found in terms of the anticipation measure.

The finding that episodic buffer suppression negatively affects the recognition measure more than phonological suppression supports Baddeley's (2000) theory that the episodic buffer is central to binding information from different sources and hence acquiring and maintaining awareness. These results again appear to indicate that awareness at the recognition level is affected by working memory suppression but that anticipation responses are not affected or affected at a lower level. The lack of anticipation effects in the present experiment and in Experiment 5 require an explanation given that Baddeley's (2000) model of working memory predicts that overall SA would be diminished due to episodic buffer suppression. The reason for this difference will now be discussed.

In Experiment 1 it was argued that recognition and anticipation could be processed differentially, particularly with regard to visual information content and Milner and Goodale's (1995) dual visual processing theory. The lack of training and expertise effects on the recognition measure was explained as being due to recognition being of lesser importance in awareness than skilled anticipation which may involve a direct perception-action process. The episodic buffer component of working memory is theorised by Baddeley (2000) to be a process in which different kinds of information content are integrated into an integrated representation of conscious experience which is then used and modified by the central executive. It is proposed that the lack of effects when using the anticipation measure in Experiments 5 and 6 was due to anticipation ability being supported by a direct perception-action process which does not always provide input to the episodic buffer that processes content for conscious experience. Alternatively, this lack of effect may be due to the insensitivity of the anticipation scores to measure automatic responding.

For Experiments 5 and 6 the central hypotheses were firstly that because visual information predominates in driving hazard perception then recognition scores would be reduced more by visual-spatial suppression than phonological suppression. This effect was found in Experiment 5. Secondly it was predicted that recognition scores would be more affected than anticipation scores, but this effect was not found. A third hypothesis was that according to a cognitive streaming account of SA phonological suppression would negatively affect both recognition and anticipation more than would visual-spatial suppression. This was not supported. Fourthly, it was predicted that central executive suppression would have a more negative effect on recognition scores than on anticipation scores, due to the likelihood of automatic responding being involved more in anticipation responding than recognition responding. However, no significant results were found. Fifth and finally, in Experiment 6 it was predicted that episodic buffer suppression would be associated with greater negative effects on the recognition and anticipation measures than phonological suppression. This hypothesis was supported in terms of recognition scores.

Conclusions that can be drawn from these effects are that firstly in terms of recognition visual processing contributes more to SA than phonological processing, supporting a Baddelian view of cognitive processing during SA and counter to a cognitive streaming account of SA. Secondly, the episodic buffer appears to be important in terms of SA recognition and conscious awareness. Thirdly, the sensitivity of the recognition measure to working memory effects and the lack of anticipation sensitivity to those effects is the inverse of the effects of training and expertise in Experiments 1 and 2 where the anticipation measure was sensitive and the recognition measure was not. This conflict and implications from the pattern of findings in Experiments 3 and 4 will be discussed in Chapter 5.

CHAPTER 5
THEORY AND DEVELOPMENT

This thesis set out partly to explore the proposition that situation awareness is described as much by aspects of a task as it is by aspects of cognition, and also to use aspects of a task as a means of measuring SA from behaviour in relation to the task. To this end a hazard perception test in its two main forms – recognition and anticipation – has shown utility in measuring the effects of training and expertise, interruption to question about aspects of SA, and working memory suppression. That only part of the results accord with predictions from cognitive and applied theory indicates that between the MHPT measure, experimental design, and theory there is a great deal more to explore.

The findings of the six main experiments in this thesis can be summarised as follows. In Chapter 2 the conclusions were that anticipation is associated with the effects of both training and expertise. The effects of training SA strategies reveals the potential utility of the MHPT anticipation measure (the complete version of the MHPT) as a measure of SA, while the effects of expertise reveals what may be a more complex relationship between expertise and SA. The effects of perception and comprehension training did not appear to be additive, hence Endsley's levels of SA do not appear to be applicable to cognitive processes. The results of Experiment 2 overall indicate that Milner and Goodale's (1992) theory of dual visual processes may have more explanatory power in terms of cognitive processes in SA than Endsley's three-level theory.

In Chapter 3 the conclusions were that there is a negative effect of interruption on anticipation scores associated with questionnaire-based measures of SA. There also appears to be a positive effect of reorientation. It was further concluded that positive reorientation effects and negative interruption effects on questionnaire based measures of SA have not been found previously due to a mutual cancelling of their effects. Furthermore, notification of forthcoming questionnaire measure interruptions does not reduce the effects of those interruptions. Notification appears to cause negative effects that interfere with the anticipation measure over time to an extent greater than any benefit from the provision of an advanced warning. This effect of notification may be due to the general residual effects of interruption. That is, if pre-interruption notification is understood to be a form of interruption then the residual negative effects of the notification are likely to continue to exist during the main interruption itself. Together the results show that interruption effects appear to have a diffused effect on performance of the main MHPT task.

In Chapter 4 the conclusions were that in terms of recognition visual-spatial processing contributes more to SA than phonological processing. Secondly, the episodic buffer appears to be important in terms of SA recognition. Thirdly, the sensitivity of the recognition measure to working memory effects and the lack of anticipation sensitivity to those effects conflicts with the effects of training and expertise where the anticipation measure was sensitive and the recognition measure was not sensitive. The issues that have arisen from the above findings will now be discussed.

The results of Experiments 1 and 2 indicate that Horswill and McKenna, (2004) are correct in stating that hazard perception is SA for drivers. The manipulations of training and expertise and their effects on the MHPT support the logical position that hazard perception is SA. Therefore, the basis of the MHPT measure used in this thesis appears robust and accords with the findings from

McKenna and Crick's (1991) manipulations of expertise, and with Hörmann *et al*'s (2004) training of SA in pilots and Endsley's (1995) and Endsley and Robertson's (2000) observations of the relationship between expertise and SA.

However, when assessing hazard perception in terms of the recognition measure a more complex picture emerged. Recognition scores were not related with expertise. A plausible explanation is that expert drivers' superior anticipation ability is based on automatic and contextual information. Contextual information is likely to be the only way a driver can anticipate a situation if the particular event is not yet visible. Expert drivers may not be able to use contextual information in order to explicitly recognise and respond to the same events when they come into view, thus resulting in their recognition scores being similar to those of novice drivers[4]. This supports the proposition that recognition may not be an appropriate measure of driving hazard perception if automatic perception-anticipation is one of the processes supporting experts' awareness ability. Nevertheless, in a different argument to that proposed in the third hypothesis in Experiment 1, the lack of expertise effects on the recognition measure indicates that expertise is related to a shift towards the use of context in anticipation. The importance of context in skilled anticipation is a long-standing observation (e.g. Lindner, 1938; Cavallo *et al*, 1988).

The low correlation between recognition and anticipation scores in Experiment 1 indicated that recognition and anticipation might occur differentially in terms of the effects of expertise, and/or the effects of training, and/or cognitive processing. There are observations in the literature which suggest that task expertise and the effects of training SA are both involved in the acquisition and maintenance of

---

4 This is also supported anecdotally in discussions the researcher has had with experienced driving instructors who believe that a significant number of qualified and approved driving instructors are failing to pass a very simple recognition-format hazard perception test.

SA. However, the lack of correlation in this first experiment appears to conflict with Endsley's (1995) theory of SA which predicts (if it is taken as a description of cognitive processes) that because perception supports anticipation there would be a positive correlation between recognition and anticipation scores. The low correlation indicates such an association but it also admits that the relationship is unlikely to be linear; there may be intermediate factors. However, an alternative to Endsley's theory, Milner and Goodale's (1992) theory of dual visual systems, argues that there is an explicit recognition-comprehension visual process and an automatic perception-action visual process. This is supported by Czerwinski *et al* (1992) who showed that there appears to be a route directly from anticipation to action without awareness. Thus it appears more plausible that the low correlation between recognition and anticipation scores is due to these two different processes; the recognition measure being sensitive to explicit recognition-comprehension and the anticipation measure being more sensitive to control-specific skilled anticipatory action, given potentially reduced suitability of the MHPT method for measuring real-world anticipatory action. Interestingly, McKenna and Crick (1991; 1994) found that when prediction is included in hazard perception there are training improvements in terms of a recognition MHPT measure. This supports the position that anticipation usually depends on automatic processes but that when the anticipation task is made more explicit it can be processed through a recognition-comprehension route and hence can be sampled using a button-press measure of SA. Croft *et al* (2004) indicate that SA measurement techniques which sample explicit SA alone will provide a distorted measure of SA since they do not take account of automatic responding and associated implicit SA. The results of Experiments 1 and 2, and Milner and Goodale's theory, are in agreement with this position.

The two and three-level approaches differ in terms of describing the cognitive processes supporting SA, in that according to a two-level hypothesis recognition and comprehension are involved in a singular recognition-comprehension process. Experiment 2 was a test of the relative explanatory power of Endsley's three-level theory (as if it were relevant to process) and Milner and Goodales's two-level visual processing theory. The results accord more with a two-level recognition and anticipation theory than a three level perception, comprehension, and projection theory. While three kinds of behaviour may be evident in behaviour (i.e. Endsley, 1995) this is compatible with a two-level process hypothesis describing the underlying cognitive processes. Three kinds of behaviour may appear to have a positive and incremental relationship with expertise, but the underlying structure of the process need not follow that same pattern.

Endsley is careful to separate the three-level SA product from the situation assessment process but makes no argument as to how process and product are related, itself an admission that process and product require separate analysis. This thesis has demonstrated that the observed levels of SA product bear little relation to the cognitive processes associated with each of them. This is relevant to measures of SA such as SAGAT which is based on the three-level product theory. By attempting to measure SA in terms of this theory SAGAT is restricting its sampling to a limited range of the explicit cognitive processes supporting SA. It does not account for effects of expertise or experience. This is an important finding since Endsley's (1995; 2004) three-level theory is the most widely cited account of SA and SAGAT is the most widely cited measure. However, it must be acknowledged that in the present thesis SA is only being examined from a visual processing perspective; other modes of input to SA measures may support alternative conclusions.

The third experiment sought to reveal the independent effects of both interruption and reorientation on the anticipation measure. This experiment appears to have demonstrated that there are negative effects of interruption when questioning a driver about aspects of SA during a task that relies upon SA. It is difficult to see how an interruption-based measure such as SAGAT (Endsley, 1995) would not interfere with the processing of a main task. This is especially true when the task is as dynamic and complex as maintaining a responsive awareness to the external environment in driving hazard perception, or for that matter during aviation SA. Interruption effects are common and often highly deleterious even after the resumption of the main task (MacFarlane & Latorella, 2002; Monk, 2004). However, Endsley (2000) reports no effects of interruption when using the SAGAT measure.

In addition to the effects of interruption and reorientation at the moment of interruption there also appears to be a residual effect of both factors on the non-interrupted anticipation measure. The results of Experiment 3 indicate that interruption has a deleterious effect on performance after the main task has been resumed, an effect that has also been observed by Monk (2004). This effect may be due to the cognitive disruption caused by processing associated with the task of answering the question. The effect may be also due to other factors associated with the questionnaire measure, such as internal mental distraction associated with the content of questions continuing to be processed in relation to the ongoing MHPT video. There is also an indication that reorientation has effects on the non-interrupted anticipation measure, since the irrelevant condition score was reduced below the baseline anticipation score for the no-query condition. This provides support for the conclusion that reorientation effects and interruption effects are capable of mutual cancelling. This finding has significant relevance to applied measurement of SA, since Endsley's (1995) interruption-based SAGAT measure that the researcher has attempted to emulate in Experiments 3 and 4 is the most widely cited measure of SA. SAGAT is used extensively in civil and military aviation as a measure of SA. The potential therefore exists for SAGAT to provide a widespread distortion of the real levels of operator SA, a measure which may not appropriately control differential effects of expertise, training, or experience. The distortion would be influenced by individual differences in the ability to manage the effects of interruption and to use the content of queries in order to deduce the rationale behind the test. Therefore operators who are able to inhibit the effects of interruption and make use of the reorientation effect would achieve higher SA scores.

Experiment 4 investigated ways in which obtrusiveness of measures such as SAGAT can be reduced by using a SPAM method. It was expected that a SPAM method of questioning would be associated overall with greater anticipation scores, since continuing to present ongoing information about a task (i.e. SPAM method) was predicted to increase significantly drivers' awareness of hazards compared with using a situation freezing SAGAT method. There was no difference between anticipation scores in the SPAM and SAGAT conditions, and there was a correlation between SPAM and SAGAT scores. Therefore, despite Durso and Gronlund's (2004) claims it appears that the SPAM method may not ameliorate the effects of interruption on SA during driving hazard perception.

Notification of forthcoming questionnaire measure interruptions was found not to reduce the effects of those interruptions. It was suggested therefore, that the notification was a clear message about how to perform the task better by rehearsing sufficient stimuli before the moment of interruption in order to answer the question and to be ready to re-enter the flow of information when returning to the main task. In order to rehearse the main task in the lead-up to an interruption processing resources are probably required. The residual effects of interruption and reorientation tend to be sustained beyond the point of task-resumption, as was observed in Experiment 3 and by Monk (2004). Thus, due to these demands a diffuse effect on processing is created around the period of the interruption. However to complicate this further, the negative effect of an interruption may be being mutually cancelled by positive effects on the anticipation measure such as the effect of a notification.

It is proposed that a likely reason for the failure to differentiate between SPAM and SAGAT methods in terms of both SAGAT scores and anticipation scores is because the interruption-based question methods disrupt cognitive processing. The effects are likely to be complex since SA in dynamic complex tasks such as driving hazard perception tends to extend across both space and time. Whether such effects are due to competing interruption and reorientation factors mutually cancelling each other or whether the lack of effect can be explained by more complex hypotheses is beyond the scope of this thesis. However, it can be concluded that interruption-based methods of SA measurement have significant shortcomings in relation to their potential to interfere with cognitive processing related to SA. The evidence in this thesis also indicates that a theory of SA must account for automatic processes direct to action as well as explicit SA.

The use of explicit SA measures such as SAGAT, and Endsley's three level theory of SA product (and its related situation assessment processes), do not appear to be helpful in understanding the contributions of implicit and explicit SA, or for explaining effects of interruption or reorientation, or effects of expertise and training, in terms of existing cognitive theory. Embedded measures of SA such as the MHPT and embedding of SA measures in full-control simulation begin to circumvent such problems, as has been demonstrated in the present thesis. Embedded measures need not interrupt or reorientate. They are not a priori dependent on any theory of SA. Embedded measures can independently sample explicit recognition-comprehension and automatic skilled anticipation.

A theory of dual-system visual processing explains how disparate conscious visual awareness and automatic action may be. Before a theory of SA is developed such processes require a deeper understanding which is more likely to be derived from performance and cognitive theory based measures such as the MHPT. The MHPT (McGowan, 2002) based on expert task analysis (ADI and Police drivers) was used in the present thesis partly as measures of SA in order to investigate cognitive processes, and partly to test its utility as an embedded measure.

The pattern of results which enabled discrimination of training, expertise, interruption, reorientation and working memory effects is evidence of the utility of the MHPT. This evidence is the basis upon which specific analysis of the technique can now be performed. Potential areas of more detailed investigation include the validity of item sets across variables such as expertise and experience; the development of a taxonomy of kinds of hazards and their effects on the measures; and the embedding of the measure in full-control simulation, in terms of the differential results found when measuring expertise, training, and working memory in this thesis.

The most widely cited theory of SA (Endsley, 1995; 2004) has been useful in particular as a straw man, and as such it has shown in the results in the present thesis, that research should focus on aspects of a task, task performance and cognitive processes during that performance rather than the development of a superordinate theory of SA. SA is both an epiphenomenon of processing in terms of conscious awareness and a part of the processing that supports task performance, '...the invariant at the core of the agent's action-perception cycle that supports skilled performance' (Smith & Hancock, 1995, p. 137). What matters is how a task is performed in accordance with a specific goal (Smith & Hancock, 1995; Patrick & James, 2004) and the cognitive processes that are associated with that performance. Definitions and characteristics of SA as a product are of secondary importance. To this end, Durso *et al*'s (1998) SPAM improves on SAGAT, not least because it is aimed at measuring an operator's ability to use information in the environment as a task-based measure of explicit SA. Nevertheless, SPAM still interrupts in order to query an operator, admitting the possibility that the measure may be influenced in some way by the intrusion, as deduced from Experiment 4.

Endsley (2004) continues to argue that perception, comprehension, and projection behaviours represent degrees of SA, despite cognitive theory which argues that the interaction between automatic and controlled processes is likely to confound such a simple linear progression. The present thesis has shown support for the argument that explicit and implicit SA and expertise are positively related to a perception-comprehension and anticipation dimension. However, in terms of visually processed SA this dimension appears to relate more to Milner and Goodale's (1992; 1995) theory of dual processes than Endsley's theory of SA product, a product which has little meaning when not defined in terms of a task and cognitive theory. Endsley (2004) repeats the claim that SA can be understood as a state of knowledge (presumably either explicit or implicit knowledge) separate from the processes used to achieve SA. The ecological approach argues that SA cannot be understood as a phenomenon distinct from the task, goals, environment, and cognitive processing[5] which may be associated with it. SA as a product is an epiphenomenon of these factors (Banbury & Tremblay, 2004; Smith & Hancock, 1995). The findings of this thesis, that both controlled and automatic processes are involved in SA, indicates that SA processing is likely to be both ecological (due to automaticity and task constraints) and cognitive. Beyond this, as Baars (2002) has indicated, SA as a product is likely to be a highly elusive construct.

Conclusions

One aim of this thesis was to test the utility of the HPT as a measure of SA. The role of automaticity tends to support this against arguments concerning the notion of mental processes, as distinct from factors external to the person such as the structure of a task. Dekker and Lutzhoft's (2004), and Patrick and James' (2004) ecologically driven approaches to SA tend to support the notion that SA describes the task (the interaction with the environment) as much as it does the mind or behaviours that the task is connected with. In this context hazard perception is a task ideally suited to the study of SA given its utility for embedding non-intrusively within tasks.

The thesis also uncovered evidence indicating that the measurement of SA is complicated by differential processes involved in controlled (explicit) SA and automatic (implicit) SA. There appears to be a central role of the episodic buffer in organising information into a coherent conscious awareness of a situation, but skilled performance of action (the more critical path) may involve other processes. Cognitive streaming and LT-WM as accounts of SA, Milner and Goodale's (1992, 1995) theory of dual visual processes, and Baddeley's (2000) model of WM including the episodic

5 To the extent that cognitive processing can be defined in a closely coupled operator-task interaction as conceptualised from an ecological perspective

buffer were invoked without particular depth due to the focus of this thesis being on testing a broad range of factors in the context of the MHPT embedded measure of SA. However, at face value those theories appear to be a more fruitful area of investigation of cognitive processes in SA than Endsley's (1995) product-based theory and its related situation assessment process. It is the combination of, and relationship between, explicit and implicit SA and the task that require investigation in terms of cognitive processes. Furthermore, measuring the subtleties of SA requires a more sophisticated approach than simply enquiring about current explicit knowledge and its use; such measures may mask more important factors than explicit awareness.

The aim of this thesis was to inform both SA application and theory from a cognitive theory perspective, in order to contribute to a psychological understanding of why SA varies. It has been argued that by embedding measures of SA no assumptions need be made about the state of knowledge being explicit or implicit. SA can simply be inferred from performance at the task, providing the task is sufficiently well defined. Embedded measures of SA offer utility in uncovering relevant cognitive processes, an endeavour that has been briefly attempted here with a broad brush approach in order to illustrate the ease with which a high-resolution measure can be used in order to test cognitive theory in the laboratory under potentially high ecological validity.

The apparently differential cognitive processes involved in explicit and implicit SA, and expertise and training, quickly emerged as factors that invite further investigation. The way that experts and novices use context and specific information also emerged as an important factor. Much of the theoretical background to costly training of SA enhancing strategies and assessment of SA in high-risk situations has been influenced by Endsley's (1995) pragmatic but contentious approach to SA. Effectiveness of training and improvements to safety would be better served by pursuing a fully cognitive approach to the theory and measurement of SA. Existing cognitive theory presents a wide range of robust models, and is supported by a wider range of evidence on SA related phenomena, than could have been dealt with in this thesis. The investigations begun in this thesis provide a contribution to cognitive theory of awareness during the performance of complex dynamic tasks.

REFERENCES

Altmann, E. and Trafton, J. (2002). Memory for goals: An activation-based model. *Cognitive Science: A Multidisciplinary Journal*, 39-83.

Amalberti R. Deblon F. (1992) Cognitive modelling of fighter aircraft's control process: A step towards intelligent onboard assistance system. *International Journal of Man-Machine Studies*. 36, 639-671.

Amalberti, R., Masson, M., Merritt, A. and Pariès, J. (2000) *Briefings: A Human Factors Course for Pilots and Aviation Professionals*. Dedale, Paris, France.

Awh, E., Jonides, J. and Reuter-Lorenz, P.A. (1998) Rehearsal in spatial working memory. *Journal of Experimental Psychology: Human Perceptual Performance*. 24(3),780-79.

Baddeley, A.D. (1986) *Working Memory*. Oxford, Clarendon Press.

Baddeley, A.D. (2000) The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*. 4(11), 417-423.

Baddeley, A.D. and Hitch, G.J. (1974) Working memory. In G.H. Bower (Ed.) *Recent Advances in Learning and Motivation* 8, New York, Academic Press.

Banbury, S., Andre, A. and Croft, D. (2000) Do We 'Know' All That We 'Know'? The Role of Implicit Knowledge in Situation Awareness. *Proceedings of 44th Annual Human Factors and Ergonomics Society Conference*. HFES, San Diego.

Banbury, S. P. and Tremblay, S. (2004). *A Cognitive Approach to Situation Awareness: Theory and Application*. Aldershot, Ashgate.

Banbury, S., Croft, D.G., Macken, W.J. and Jones, D.M. (2004) A Cognitive Streaming Account of Situation Awareness. In S. Banbury and S. Tremblay (Eds.), *A Cognitive Approach to Situation Awareness: Theory and Application*. 117-136, Aldershot, Ashgate.

Baars, B. (1988) *A Cognitive Theory of Consciousness*. New York, Cambridge University Press.

Baars, B. (1997) Consciousness versus attention, perception, and memory. *Consciousness and Cognition*. 6, 363-371.

Baars, B.J. (2002) The conscious access hypothesis: Origins and recent evidence. *Trends in Cognitive Science*. 6 (1), 47-52.

Bedny, G. and Meister, D. (1997) *The Russian Theory of Activity: Current Applications to Design and Learning*. Mahwah, NJ, Lawrence Erlbaum.

Bedny, G. and Meister, D. (1999) Theory of activity and situation awareness. *International Journal of Cognitive Ergonomics*, 3(1), 63-72.

Breton, R. and Rousseau, R. (2001) *Situation Awareness: A Review of the Concept and its Measurement.* Technical Report No. 2001-220, Defence Research and Development Canada, Valcartier.

Bruner, J.S. (1966) *Toward a Theory of Instruction.* Norton, New York.

Cavallo, V., Brun Dei, M., Laya, O. and Neboit, M. (1988) Perception and anticipation in negotiating curves: The role of driving experience. In A. Gale, *Vision in Vehicles* II (365-374). Oxford, UK, North-Holland.

Chalmers, D. (1996) *The Conscious Mind: In Search of a Fundamental Theory.* Oxford, Oxford University Press.

Chaminade, T., Meary, D., Orliaguet, J. and Decety, J. (2001) Is perceptual anticipation a motor simulation? A PET study. *Neuroreport: For Rapid Communication of Neuroscience Research.* 12(17), 3669-3674.

Chapanis, A. and Overbey, C. M. (1974) Studies in interactive communication: III. Effects of similar and dissimilar communication channels and two interchange options on team problem solving. *Perceptual and Motor Skills.* 38, 343–374.

Chase, W. and Ericsson, K. (1982) Skill and Working Memory. In G. Bower, *The Psychology of Learning and Motivation,* 16 (1-58). New York, Academic Press.

Chase, W. and Simon, H. (1973) The mind's eye in chess. In W. Chase, *Visual Information Processing* (215-281) New York, Academic Press.

Cheng, P.W. (1985) Restructuring versus automaticity: Alternative accounts of skills acquisition. *Psychological Review*, 92, 414-423.

Croft, D., Banbury, S. P., Butler, L. and Berry, D. C. (2004) The role of awareness in situation awareness. In S. Banbury and S. Tremblay (Eds.), *A Cognitive Approach to Situation Awareness: Theory and Application.* 82-103, Aldershot, Ashgate.

Czerwinski, M.P., Chrisman, S. E. and Rudisill, M. (1991) Interruptions in multitasking situations: The effects of similarity and warning (*Technical Report JSC-24757*) National Aeronautics and Space Administration, Houston, Texas.

Czerwinski, M., Lightfoot, N. and Shiffrin, R.M. (1992) Automatization and training in visual search. *American Journal of Psychology.* 105, 271-315.

Dekker, S. and Lützhöft, M. (2004) Correspondence, cognition and sensemaking: A radical empiricist view of situation awareness. In S. Banbury and S. Tremblay (Eds.), *A Cognitive Approach to Situation Awareness: Theory and Application.* 22-41, Aldershot, Ashgate.

Dominguez, C. (1994) Can SA be defined? In M. Vidulich, C. Domingues, E. Vogel, and G. McMillan, *Situation Awareness: Papers and Annotated Bibliography.* Wright-Patterson Air Force Base, Armstrong Laboratory Technical report.

Durso, F.T., Bleckley, M.K. and Dattel, A.R. (2006) Does situation awareness add to the validity of cognitive tests. *Human Factors* 48(4) 721-733.

Durso, F. and Dattel, A. (2004) SPAM: The real-time assessment of SA. In S. Banbury and S. Tremblay (Eds.), *A Cognitive Approach to Situation Awareness: Theory and Application.* 137-154, Aldershot, Ashgate.

Durso, F. and Gronlund, S. (1999) Situation awareness. In F. Durso, R. Nickerson, R. Schavaneveldt, S. Dumais, D. Linsay and M. Chi. *Handbook of Applied Cognition.* John Wiley and Sons.

Durso, F.T, Hackworth, C.A., Truitt, T., Crutchfield, J. and Manning, C.A.(1998) Situation awareness as a predictor of performance in en route air traffic controllers. *Air Traffic Quarterly*, 6, 1-20.

Elsner, B. and Hommel, B. (2001) Effect anticipation and action control. *Journal of Experimental Psychology: Human Perception and Performance*, 27(1), 229-240.

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

Endsley, M.R. (1990) Predictive utility of an objective measure of situation awareness. *Proceedings of the Human Factors Society 34th Annual Meeting* (41-45) Santa Monica, CA, The Human Factors Society.

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

Endsley, M.R. (1995b) Toward a theory of situation awareness. *Human Factors*, 37 (1), 32-64.

Endsley, M.R. (2000) Direct measurement of situation awareness: Validity and use of SAGAT, In M.R. Endsley and D.J. Garland (Eds.), *Situation Awareness Analysis and Measurement*, Erlbaum, NJ.

Endsley, M.R. (2000b) Theoretical underpinnings of situation awareness: A critical review, In M.R. Endsley and D.J. Garland (Eds.), *Situation Awareness Analysis and Measurement*, Erlbaum, NJ.

Endsley, M.R. (2004) Situation awareness: Progress and directions. In S. Banbury and S. Tremblay (Eds.), *A Cognitive Approach to Situation Awareness: Theory and Application.* (317-341) Aldershot, Ashgate.

Endsley, M. and Robertson, M. (2000) Training for situation awareness in individuals and teams. In M. Endsley and D. Garland, *Situation Awareness: Analysis and Measurement* (349-365) NJ, Erlbaum.

Engstroem, D., Kelso, J. and Holroyd, T. (1996) Reaction-anticipation transitions in human perception-action patterns. *Human Movement Science*, 15(6), 809-832.

Ericsson, K.A. (2002) Towards a procedure for eliciting verbal expression of non-verbal experience without reactivity: Interpreting the verbal overshadowing effect within the theoretical framework for protocol analysis. *Applied Cognitive Psychology* 16(8), 981-987.

Ericsson, K. and Kintsch, W. (1995) Long-term working memory. *Psychological Review,* 102, 211-245.

Ericsson K.A. and Simon H.A. (1993) *Protocol Analysis: Verbal Reports as Data* MIT Press, Cambridge, MA.

Federico, P.A. (1995) Expert and novice recognition of similar situations. *Human Factors.* 37(1), 105-122.

Flach, J. (1994) Situation awareness: The emperor's new clothes. In M. Mouloua and R. Parasuraman (Eds.) *Human Performance in Automated Systems: Current Research and Trends* (241-248), NJ. Erlbaum.

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

Flach, J., Mulder, M. and van Paassen, M.M. (2004). The concept of the situation in Psychology. In S. Banbury and S. Tremblay (Eds.), *A Cognitive Approach to Situation Awareness: Theory and Application.* 42-60, Aldershot, Ashgate

Gagné, R.M. (1985) *The Conditions of Learning and Theory of Instruction,* CBS College Publishing, New York.

Gibson, J. J. (1979) *The Ecological Approach to Visual Perception.* Houghton Mifflin. Boston.

Gillie, T. and Broadbent, D. (1989) What makes interruptions disruptive? A study of length, similarity and complexity. *Psychological Research,* 50, 243-250.

Groeger, J. (2000) *Understanding Driving: Applying Cognitive Psychology to a Complex Everyday Task.* Philadelphia, PA, Psychology Press.

Gugerty, L., Brooks, J. and Treadaway, C. (2004) Individual differences in situation awareness for transportation tasks. In S. Banbury and S. Tremblay (Eds.), *A Cognitive Approach to Situation Awareness: Theory and Application.* 193-212, Aldershot, Ashgate.

Hodgetts, H. M. and Jones, D. M. (2003) Interruptions in the Tower of London task: Can preparation minimise disruption? *Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society.* Santa Monica, HFES.

Hörmann, H.-J., Banbury, S.P., Dudfield, H., Lodge, M. and Soll, H. (2004) Effects of situation awareness training on flight crew performance. In S. P. Banbury and S. Tremblay (Eds.), *A Cognitive Approach to Situation Awareness: Theory and Application* (213-232). Aldershot, Ashgate.

Hörmann, H.-J., Soll, H., Banbury, S. and Dudfield, H. (2003) ESSAI - Training of Situation Awareness and Threat Management Techniques: Results of an Evaluation Study. *Proceedings of The Twelfth International Symposium on Aviation Psychology.*

Horswill, M. S. and McKenna, F. P. (2004) Drivers' hazard perception ability: Situation awareness on the road. In S. P. Banbury and S. Tremblay (Eds.), *A Cognitive Approach to Situation Awareness: Theory and Application* (154-173). Aldershot, Ashgate.

Jeannott, E., Kelly, C. and Thompson, D. (2003) *The Development of Situation Awareness Measures in ATM Systems*, EATMP report. HRS/HSP-005-REP-01.

Jones, D.M., Beaman, C.P. and Macken, W.J. (1996) The object-oriented episodic record model, In S. Gathercole (Eds.), *Models of Short-term Memory,* Lawrence Erlbaum Associates, London.

Kay, H. and Poulton, E. (1951) Anticipation in memorizing. *British Journal of Psychology*, 42, 34-41.

Klein, S.A. (2002) Libet's temporal anomalies: A reassessment of the data. *Consciousness and Cognition*, 11, 198-214.

Kreifeldt, J. G. and McCarthy, M. E. (1981) Interruption as a test of the user-computer interface. *Proceedings of the 17th Annual Conference on Manual Control* (655-667). Jet Propulsion Laboratory Publication 81-95, Cal Tech.

Lee, W. O. (1992) The effects of skill development and feedback on action slips. *Proceedings of the HCI 1992 Conference on People and Computers* VII.

Lindner, R. (1938) An experimental study of anticipation. *American Journal of Psychology*, 51, 253-260.

Libet B., Wright, E., Jr, Feinstein, B. and Pearl, D. K. (1979) Subjective referral of the timing for a conscious sensor experience: A functional role for the somatosensory specific projection system in man. *Brain*, 194, 191–222.

May, J. L. and Gale, A. G. (1998) How did I get here? – driving without attention mode. In M. Hanson (Ed.), *Contemporary Ergonomics.* London, Taylor and Francis.

McFarlane D. C. and Latorella K. A. (2002) The scope and importance of human interruption in human-computer interaction design, *Human-Computer Interaction,* 17 (1), 1-61.

McGowan, A.M. (2002) An event-based measure of situation awareness. *MSc dissertation*, Cardiff University.

McGowan, A.M. and Banbury, S.P. (2004) Evaluating interruption-based techniques using embedded measures of driver anticipation. In S. P. Banbury and S.

Tremblay (Eds.), *A Cognitive Approach to Situation Awareness: Theory and Application* (154-173).

McGuinness, B. and Foy, J. (2000) A subjective measure of SA: The crew awareness rating scale (CARS). *First Human Performance, Situation Awareness, and Automation Conference.*

McKenna, F. and Crick, J. L. (1991) *Hazard Perception in Drivers: A Methodology for Testing and Training,* Final Report. Behavioural Studies Unit, Transport and Road Research Laboratory, Crowthorne, UK.

McKenna, F.P. and Crick, J.L. (1993) A cognitive psychological approach to driver training: The use of video technology in developing the hazard perception skills of novice drivers. In Grayson (Ed.) *Behavioural Research in Road Safety III.* TRL report PA 3004/93, Transport and Road Research Laboratory. Crowthorne, Berkshire.

McKenna, F. and Crick, J. (1994) *Hazard Perception in Drivers. A Methodology for Testing and Training.*TRL. Crowthorne UK, TRL.

McKenna, F. and Crick, J. (1997) *Developments in Hazard Perception.* Final Report 297. Crowthorne, UK, TRL.

McKenna, F. and Farrand, P. (1999) The role of automaticity in driving. In G. Grayson (Ed.), *Behavioural Research in Road Safety IX.* Transport Research Laboratory. Crowthorne, UK.

Milner, A. D. and Goodale, M. A. (1992) Separate visual pathways for perception and action. *Trends in Neuroscience*, 15, 20-25.

Milner, A.D. and Goodale, M.A. (1995) *The Visual Brain in Action.* Oxford University Press, Oxford.

Miyata, Y. and Norman, D. (1986) Psychological issues in support of multiple activities, In D.A. Norman and S.W. Draper (Eds.), *User-Centered Design,* 265-284, Lawrence Erlbaum, NJ.

Monk, C.A. (2004) *The Residual Disruption Effect in Interrupted Task Performance.* PhD dissertation, George Mason University.

Neisser, U. (1976) *Cognition and Reality.* San Francisco, Freeman.

Norman, D.A. and Shallice, T. (1986) Attention to action: Willed and automatic control of behaviour. In R.J. Davidson, G.E. Schwartz and D. Shapiro (Eds.), *The Design of Everyday Things.* New York. Doubleday.

Onslow, L.S. (2003) Investigating the Loci of Situation Awareness: The Role of Working Memory in Expert and Novice drivers. *MSc dissertation*, Cardiff University.

Patrick, J. and James, N. (2004) A task-oriented perspective of situation awareness. In S. P. Banbury and S. Tremblay (Eds.), *A Cognitive Approach to Situation Awareness: Theory and Application* (61-81). Aldershot, Ashgate.

Pelz, D. C. and Krupat, E. (1974) Caution profile and driving record of undergraduate males. *Accident Analysis and Prevention*, 45-58.

Peterson, L., Brewer, C. and Bertucco, R. (1963) Guessing strategy with the anticipation technique. *Journal of Experimental Psychology,* 65(3), 258-264.

Pew, R.W. (1995) The state of situation awareness measurement: Circa 1995. In D. Garland and M. Endsley (Eds.), *Experimental Analysis and Measurement of Situation Awareness*, 7-16, Embry-Riddle, Daytona Beach.

Pew, W. and Mavor, S. (1998) Modeling human and organisational behavior: Application to military simulations. *Panel on Modeling Human Behavior and Command Decision-Making: Representations for Military Simulations*, 174.

Police Foundation (1994) *Roadcraft - The Essential Police Driver's Handbook.* The Stationary Office (HMSO).

Prince, C. and Salas, E. (1998) Team Situation Awareness, Errors, and Crew Resource Management. In M.R. Endsley and D.J. Garland (Eds.), *Situation Awareness Analysis and Measurement*, Erlbaum, NJ.

Pritchett, A.P. and Hansman, R.J. (2000) Use of testable responses for performance-based measurement of situation awareness, In M.R. Endsley and D.J. Garland (Eds.), *Situation Awareness Analysis and Measurement*, Erlbaum, NJ.

Quimby, A., Maycock, G., Carter, I., Dixon, R. and Wall, J. (1986) *Perceptual Abilities of Accident Involved Drivers*. TRL Report 27. Crowthorne, Berkshire.

Quimby, A. and Watts, G. (1981) *Human Factors and Driving Performance.* TRL Report 107. Crowthorne, Berkshire.

Robbins, T.W., Anderson, E.J., Barker, D.R., Bradley, A.C., Fearneyhough, C., Henson, R., Hudson, S.R. and Baddeley, A.D. (1996) Working memory in chess. *Memory and Cognition,* 24 (1), 83-93.

Rogers, W.A., Lee, M.D. and Fisk, D. (1995) Contextual effects on general learning, feature learning, and attention strengthening in visual search. *Human Factors,* 37(1), 158-172.

Rousseau, R., Tremblay, S. and Breton, R. (2004) Defining and modeling situation awareness: A critical review. In S. Banbury and S. Tremblay (Eds.), *A Cognitive Approach to Situation Awareness: Theory and Application.* 3-21. Aldershot, Ashgate.

Rowe, R. and McKenna, F. (2001) Skilled anticipation in real-world tasks: Measurement of attentional demands in the domain of tennis. *Journal of Experimental Psychology: Applied*, 7(1), 60-67.

Rumelhart, D.E., Hinton, G.E. and Williams, R.J. (1986) Learning internal representations by error backpropagation, In D. E. Rumelhart and J. L. McClelland (Eds.), *Parallel Distributed Processing* 1, 318-362, MIT Press.

Sarter, N. and Woods, D. (1995) How in the world did I ever get into that mode: Mode error and awareness in supervisory control. *Human Factors*, 37(1), 5-19.

Schneider, W. and Shiffrin, R.M. (1977) Controlled and automatic human information processing: I. detection, search and attention, *Psychological Review*, 84, 1-66.

Schutte, P.C. and Trujillo, A.C. (1996) Flight crew task management in non-normal situations. *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting*, HFES, Santa Monica.

Shiffrin, R.M. and Schneider, W. (1977) Controlled and automatic human information processing: II. perceptual learning, automatic attending, and a general theory, *Psychological Review*, 84, 127-190.

Smith, K. and Hancock, P. A. (1995) Situation awareness is adaptive, externally directed consciousness. *Human Factors*, 137-148.

Snow, M. and Reising, J. (2000) Comparison of two situation awareness metrics: SAGAT and SA-SWORD. *Proceedings of the Human Factors and Ergonomics Society Conference,* HFES, Santa Monica, CA.

Sohn, Y.W. and Doane, S.M. (2000) Predicting individual differences in situation awareness: The role of working memory capacity and memory skill, *Proceedings of the Human Performance, Situation Awareness and Automation Conference*, Savannah.

Speier, C., Valacich, J. S. and Vessey, I. (1997) The effects of task interruption and information presentation on individual decision making. *Proceedings of the 18th International Conference on Information Systems* (21–36). New York, Association for Computing Machinery.

Stanton, N.A., Salmon, P.M., Walker, G.H., Baber, C. and Jenkins, D.P. (2005) *Human Factors Methods*. Defence Technology Centre: Human Factors Integration. Ashgate, Aldershot.

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

Tulving, E. (1972) Episodic and semantic memory. In E.Tulving and W. Donaldson (Eds.), *Organization of memory*. New York, Academic Press.

Tulving, E. (1989) Memory: Performance, knowledge and experience. *European Journal of Cognitive Psychology*. 1, 3-36.

Van Bergan, A. (1968) *Task Interruption*. North Holland.

Watts, G.R. and Quimby, A.R. (1979) *Design and Validation of a Driving Simulator*, Report LR 907. TRL, Crowthorne, UK.

Wikman, A.S, Nieminen, T. and Summala, H. (1998) Driving experience and time sharing during in-car tasks on roads of different width. *Ergonomics,* 41 (3), 358-372.

Appendix A-1 HPT instructions

Watch the video of a road scene and use the mouse to click on objects in the road view where you see a hazard that would cause you to take some physical action when driving a real car.

Appendix A-2.1 Central executive suppression task instructions

Using the digits one to nine please say one digit per second and randomise your use of the digits. Randomise means to avoid making any patterns in the series of digits you say.

Appendix A-2.2 Visual spatial suppression task instructions

Press the following series of keys at the rate of one per second: 1,4,7,8,5,2,3,6,9,8,7,4,5,6,3,2,1. You will notice that this series traces an S shape which inverts and rotates midway through the sequence.

Appendix A-2.3 Episodic buffer suppression task instructions

Images will appear very briefly in the centre of the video that you are to be shown. When an image appears hold in your mind any events in your life the image of which the image have reminded you. If you do not have any memories in relation to an image try to keep your mind on past experiences anyway. Do not try to remember the image itself, only past events in your life.