

Situation Awareness During Driving: Explicit and Implicit Knowledge in Dynamic Spatial Memory

Leo J. Gugerty
Galaxy Scientific Corporation

This study investigated people's ability to monitor changing spatial information in a simulated driving task. Drivers' knowledge of the locations of traffic cars was assessed with both direct recall measures and indirect performance measures. The direct and indirect measures were positively correlated (associated), suggesting that drivers' knowledge of nearby cars is largely explicit with little contribution of implicit knowledge. When there were too many traffic cars to monitor, drivers used cues such as car location to focus attention on potentially hazardous cars. When drivers had more active control of the driving task, they remembered the locations of potentially hazardous cars more accurately than when they viewed driving scenes in a passenger mode. These findings have implications regarding how people maintain situation awareness during real-time tasks and potentially for the development of dynamic tests of driving ability.

Keeping track of a dynamic, changing situation is a key element of real-time tasks such as driving, flying, air traffic control, and emergency medicine. For example, drivers must keep track

of their route location, the position and speed of their own and other vehicles, road conditions, and the condition of their vehicles. In addition, drivers must estimate how these variables might change in the near future in order to make good decisions about navigation, maneuvering, and other driving subtasks. Both researchers and practitioners, in particular, airplane pilots, have used the term *situation awareness* to describe the complex set of information that must be maintained in real-time tasks (Adams, Tenney, & Pew, 1995; Endsley, 1995b).

The research described here investigated situation awareness in the real-time task of driving. A PC-based driving simulator was used as an experimental vehicle. In particular, the focus was on knowledge needed for the driving subtask of maneuvering, that is, how drivers monitor the locations of the vehicles around them.

In terms of cognitive processes, Endsley (1995b) stressed that the concept of situation awareness is best seen as encompassing perceptual and comprehension processes but not decision-making and response execution processes. Adams et al. (1995) described situation awareness as a dynamic mental model of the situation that has two elements: (a) explicit focus or active knowledge in working memory and (b) implicit

Leo J. Gugerty, Galaxy Scientific Corporation, San Antonio, Texas.

This research was performed at the Armstrong Laboratory, Cognitive Technologies Branch, Brooks Air Force Base. It was supported in part by the Air Force Office of Scientific Research (AFOSR) Grant PE 61102F, Work Unit 2313T147, Task 2313/BA Learning Abilities Measurement Program (LAMP), LRIR 92AL009, and a fellowship from the National Research Council. Portions of this research were presented at the International Conference on Experimental Analysis and Measurement of Situation Awareness, November 1995, Daytona Beach, Florida.

I thank William Tirre for helpful comments on all phases of the research, Scott Chaiken for comments on an earlier version of this article, Jamie Burns and Roger Dubbs for programming the driving simulator, and Brennan Underwood and Linda Robertson-Schule for help with data collection.

Correspondence concerning this article should be addressed to Leo J. Gugerty, c/o Armstrong Laboratory (HRMC), 7909 Lindbergh Drive, Brooks Air Force Base, Texas 78235. Electronic mail may be sent via Internet to gugertyl@cobra.brooks.af.mil.

focus or less active knowledge that is relevant to the current situation but more accessible than irrelevant knowledge in long-term memory (cf. Sanford & Garrod, 1981).

It is evident that some of the critical knowledge drivers use while maneuvering is explicit and conscious. However, implicit knowledge of nearby traffic might also be used. Implicit knowledge might come into play when a driver focuses on one car and still encodes information about a second car, as in divided attention studies (Johnston & Dark, 1986). The driver may not be aware of the second car during encoding or may lose awareness of the car quickly; yet information about this second car may persist in memory in an implicit form that may affect later driving actions. As an extreme example of this phenomenon, one can sometimes drive a car while being engrossed in a conversation and immediately thereafter have no recollection of the past few minutes of driving. Flinn (1965) reported instances of this phenomenon among airplane pilots. Thus, the first research question investigated was "Can drivers use implicit knowledge of the vehicles around them to help maneuver safely?"

To address this question, I developed an explicit free-recall measure of drivers' car location knowledge and also two performance-based measures that could potentially assess both implicit and explicit knowledge of car locations. Data from the recall and performance measures were compared to determine whether there was any evidence of implicit knowledge in the absence of explicit knowledge.

The second question I investigated followed from the observation that in order to track surrounding traffic, drivers must shift attention to locations in the forward visual field as well as to rear- and sideview mirrors. Thus, this task requires not only perceptual tracking but also the use of dynamic spatial working memory. Furthermore, in heavy traffic working memory is overloaded, and drivers must allocate attention to a subset of the nearby cars. Therefore, the second research question concerned how drivers' knowledge of traffic car locations is affected by changes in working memory load. The recall and performance-based measures of situation awareness were used to investigate how drivers' memory for dynamic spatial information was affected by changes in working memory load and how driv-

ers allocated attention among vehicles when their working memory capacity was overloaded. Although considerable research has been conducted on spatial working memory (Baddeley, 1990, 1992), very little has focused on memory for dynamic information. Thus, this research contributes to a little understood area in spatial cognition.

Measuring situation awareness during driving can have useful applications because research has shown that poor situation awareness is an important cause of driving accidents. Based on a detailed evaluation of 2,258 traffic accidents, Treat et al. (1979) concluded that improper lookout and inattention were the two leading causes of the accidents. Improper lookout and inattention, which are aspects of situation awareness, were cited as causes of more accidents than factors related to decision making (e.g., excessive speed) and psychomotor ability (e.g., improper driving technique). Others have found tests of hazard perception and attention switching ability to be significantly related to on-road accident rates (Elander, West, & French, 1993). The situation awareness measures used in this research included measures of how well drivers perceive hazards and allocate attention among vehicles. Thus, simulator-based measures of situation awareness ability can potentially be part of inexpensive test batteries to select people for real-time tasks (e.g., in a driver licensing battery) and to identify people who need additional training.

Measures of Explicit and Implicit Knowledge in Situation Awareness

A large body of research has been conducted that is claimed to support a distinction between explicit and implicit memory systems or processes (Dorfman & Mandler, 1994; Roediger & McDermott, 1993). Explicit memory involves "conscious recollection of previous experience," whereas implicit memory can improve performance on a task in the absence of conscious recollection (Schacter, 1987). In order to measure people's explicit and implicit knowledge, researchers have developed what have been called direct and indirect memory tasks or measures. Roediger and McDermott (1993) defined a *direct measure* as one in which "subjects are asked to recall or to recognize events from . . . their past experience" (p. 65), and an *indirect measure* as

one in which subjects are not asked to recall past events but rather to perform a task as well as possible. Participants' retention is then inferred from their performance.

A number of researchers have used direct measures of real-time knowledge (Endsley & Bolstad, 1994; Fracker & Davis, 1991; Vidulich, Stratton, Crabtree, & Wilson, 1994). A common direct measure is to have an individual participate in a real-time scenario in a simulator and then blank the simulator screen and ask the participant to recall information about the scenario such as the locations of nearby vehicles in a driving scenario. However, recall-based measures alone may provide an incomplete picture of situation awareness because many real-time tasks involve well-practiced, automatic processes that may register information in explicit, working memory only fleetingly, if at all.

Given this criticism, it seemed important to develop situation awareness measures that would not depend on explicit recall, such as indirect performance-based measures. For example, if a driver in a simulator escapes from a hazard by changing lanes and thereby crashes into a car in the blind spot, it can be inferred that the driver had poor knowledge of that car. Thus, the percentage of times a driver avoids hitting blindspot cars while maneuvering could be a useful performance-based measure of knowledge of cars in the blind spot.

It is important to stress that there is not a one-to-one relation between direct measures and explicit knowledge and between indirect measures and implicit knowledge (Debnar & Jacoby, 1994). For example, a driver who is good at avoiding crashes with blind spot cars in a simulator would be considered as showing good performance on an indirect measure of car location knowledge. However, this good performance could be based exclusively on explicit knowledge of car locations, exclusively on implicit knowledge, or on a combination of both. Whether an indirect measure taps explicit or implicit knowledge is an empirical question.

The main type of empirical evidence for implicit knowledge comes when performance on direct and indirect measures is dissociated. For example, researchers have found dissociations (i.e., zero or negative correlations) between participants' ability to perform process control tasks

and to answer explicit questions about the task domains (Berry & Broadbent, 1984, 1988; Buchner, Funke, & Berry, 1995). Another type of dissociation occurs when individuals are learning a task, and indirect measures show the growth of their knowledge much sooner than direct measures. For example, research on serial reaction time tasks has shown that when a sequence of object locations is repeatedly presented to individuals, indirect, reaction-time measures (priming) suggest increased implicit knowledge of the sequence before they become explicitly aware of the sequence (Willingham, Nissen, & Bullemer, 1989).

Distinctions between explicit and implicit memory and learning have been clouded in recent years by certain empirical findings. Buchner et al. (1995) showed that the negative correlation between performance and recall on Berry and Broadbent's (1984) process control task occurred because good performers received less practice on the types of items in the recall test than poor performers. Also, Perruchet and Amorim (1992) used a more sensitive test of explicit knowledge in serial reaction time tasks and found that individuals had considerable (although fragmentary) explicit knowledge of the sequences very early in the learning process when only implicit knowledge had been shown previously.

These findings suggest that some dissociations between direct and indirect measures may be due to experimental artifacts, such as differential practice, or insensitive direct and indirect measures and that the growth of explicit knowledge during learning may closely parallel knowledge tapped by indirect tests. Applied to the previous example of driving without recollection, the findings suggest that during the unrecalled period of driving, people have at least brief periods of explicit awareness of key driving knowledge and that this knowledge may be revealed by appropriate tests. Endsley (1995b) expressed a similar viewpoint. If the knowledge used during maneuvering is primarily explicit, it would be expected that direct recall and indirect performance-based measures of car location knowledge would be associated (positively correlated). Conversely, if implicit knowledge is important for maneuvering, the recall and performance measures would be dissociated. This would oc-

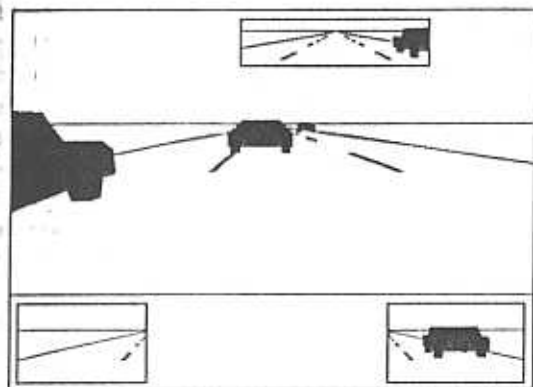


Figure 1. Three-dimensional scene from the driving simulator. The participants saw the actual scenes in color.

cur, for example, when individuals performed well but recalled poorly.

Below, the direct and indirect measures of situation awareness are described. Then, three experiments are presented that assess the extent to which explicit and implicit knowledge are used during maneuvering and the effects of working memory overloads on situation awareness.

Driving Task Description

The driving task was performed on a PC-based driving simulator. The simulator showed 3-dimensional animated driving scenes on a computer screen. Figure 1 shows a frame from one of these scenes. The participant saw the view from the driver's seat and could use the rearview, left sideview, and right sideview mirrors to track cars in the rear. Participants watched animated scenes lasting from 18 s to 35 s and were instructed to imagine that their simulated car was on "autopilot" (i.e., cruise control plus lane position control). At the end of each scene, knowledge of the locations of the traffic vehicles was probed with one or both of two methods. In the recall probes, the moving scene disappeared, and the participants indicated the locations of the traffic cars at the end of the scene on a bird's-eye view of the road by using the mouse (see the left side of Figure 2). Then they received feedback (as on the right side of Figure 2) that indicated the correct final car locations for that scene. The recall probes are similar to Endsley's (1995a) situation awareness global assessment technique (SAGAT).

In the performance probes, participants could make driving responses while viewing the moving scenes; that is, they could override the autopilot. On some trials, an incident would occur that required a driving response: for example, a car would move into the driver's lane ahead of or behind the driver while moving slowly or fast enough that it would collide with the driver. Participants could avoid these hazards by using the keyboard arrow keys to accelerate, decelerate, move to the lane on the left, or move to the lane on the right.

On each hazard trial, a response interval was defined as starting when the hazard car moved into the driver's lane, close to the driver; it ended just before the hazard car hit the driver at the end of the trial. A response interval was also defined at the end of each catch trial. For both hazard and catch trials, if participants responded during the response interval, the scene stopped and they received textual feedback concerning the correctness of their response. If subjects responded before the response interval, the computer beeped, the moving scene continued, and participants had to be prepared to respond later in the scene, if necessary.

Discrete arrow key responses were used for the performance probes instead of interactive driving; thus participants' performance could be compared across the same set of hazards, while still allowing participants to make driving responses. It was hoped that requiring participants

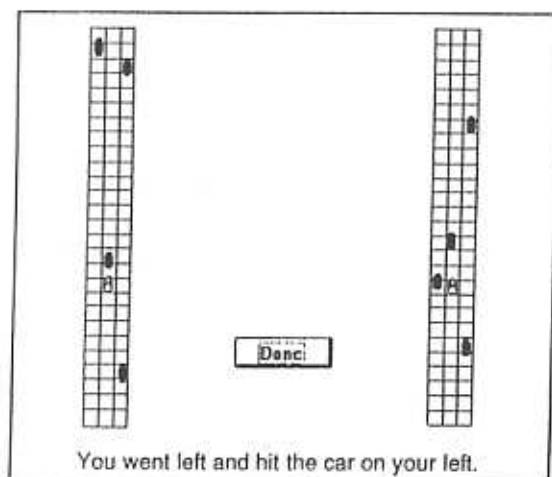


Figure 2. Recall grid (left) and feedback for a performance-recall probe.

to make driving responses would cause them to monitor the locations of traffic vehicles as if they were really driving.

In some cases, the performance probes and recall probes were combined into a third type of probe, performance-recall probes, in which participants gave driving responses, as necessary, during the moving scenes and then recalled the traffic car locations at the end of the scene. In the performance-recall probes, the textual feedback about the correctness of the driving responses was withheld until after the participants had completed the recall probe (see Figure 2).

Situation Awareness Measures

Recall Measures of Situation Awareness

To evaluate how well a participant recalled traffic car locations on a trial, it is necessary to compare the vehicle locations recalled by the participant with the actual vehicle locations. Because the participants recalled cars only by locations, using no other identifying characteristics, the first task was to compare the recalled and the actual locations and decide which recalled cars the participant intended to be matched with which actual cars. In Figure 2, the driver's car is white. It seems likely that the participant has correctly recalled the car directly in front and the two cars in the right lane and missed the car directly on the left. The car placed far ahead in the left lane by the participant is a *false alarm*.

Because matching the recalled and actual cars by hand would have involved too many subjective judgments, a computer algorithm was developed to perform the matching. This algorithm, which is described in Appendix A, compares all likely ways of matching the recalled and actual cars for a trial and chooses the match that minimizes the average error distance between the recalled and the actual cars. The algorithm was validated by comparing it with matches for 14 trials made by two human judges, one of whom knew the algorithm and one who did not. The algorithm agreed exactly with the matches of each judge on 11 of 14 trials. The discrepancies on the remaining 6 trials were minor and would not greatly affect the dependent variables calculated based on the matches.

Once the matching was completed for a trial, it

is possible to calculate quantitative estimates of the goodness of the participant's recall. These include percentage of cars recalled (75% in the above example) and recall error, which is the average error distance for cars recalled by the subject. In addition, these variables were combined into a composite measure called composite recall error. This measure weighted errors in recalling distant cars as less important than errors to close cars because this pattern fit with intuitions about how to optimally allocate attention during driving and with the participants' data.

Performance Measures of Situation Awareness

The major difficulty in developing performance measures of situation awareness is creating measures that reflect people's perceptual and comprehension processes more than they reflect decision and response-execution process, even though all of these processes must contribute to any measure of real-time performance. Using Endsley's (1995a) terms, I intended to develop imbedded task measures that reflect particular aspects of situation awareness. In contrast to imbedded task measures, global performance measures reflect a more even mix of situation awareness and decision-action processes. Global measures are also useful as criterion variables to assess how both explicit and implicit situation awareness affects overall driving performance.

The first imbedded task measure developed was hazard detection, which was calculated with the A' nonparametric signal detection measure of sensitivity (Grier, 1971). As described above, on each signal (hazard) trial, the response interval began when a car entered the driver's lane on a trajectory that would collide with the driver and it ended when it was too late for the driver to avoid the oncoming car. Following the procedure of Watson and Nichols (1976) for measuring sensitivity and bias with continuous signal detection tasks, the catch trial response intervals were set to be equal in duration to those on hazard trials. A *hit* was defined as any driving (arrow key) response, even an incorrect response, during the response interval of a hazard trial. A *false alarm* was any arrow key response during the response interval of a catch trial. For all trials, responses

before the response interval, which were infrequent, were ignored in this analysis.

When participants responded incorrectly to a hazard car, this shows that they were aware of the hazardous situation but selected and executed an inappropriate avoidance response. Therefore, because the hazard detection measures counted even incorrect responses to hazards as hits, the measure should reflect participants' ability to detect hazards (an aspect of situation awareness) more than their decision-action abilities. The hazard detection measure focused on participants' awareness of vehicles in front of and behind their cars because the hazardous cars always entered the driver's lane from a side lane and then approached the driver.

The second imbedded task measure, blocking car detection, focused on participants' awareness of blocking cars to their immediate right and left. These cars were usually completely within the participants' blind spot. Because the three-dimensional display did not show cars immediately to the right or left of the driver, participants had a larger blind spot than in real driving. On hazard trials, participants could usually only know about blocking cars by remembering that a car had entered the blind spot and had not left it. All cars defined as blocking cars were located such that the participant's car would hit them if the participant tried to avoid the hazard car by moving to the right or left at any time during the response interval.

A small percentage of participants adopted a strategy of only making accelerate or decelerate responses, probably so as to avoid having to remember cars in the blind spot. Because of this, when participants accelerated or decelerated or made no response on a hazard trial, it was difficult to determine whether their response was based on knowledge of blocking cars. Therefore, blocking car detection was estimated based only on hazard trials where the participant responded right or left during the response interval. For example, on a trial where the hazard car approached from the front and there were blocking cars to the right and left, participants were considered as detecting one of two blocking cars if they went right or left. Overall, blocking car detection was estimated by the ratio of the total number of blocking cars avoided over the total number of blocking cars. In each case, these

totals were summed over all hazard trials where the participant responded right or left. Blocking car detection was only estimated for participants' who responded right or left on more than 5 of the 42 hazard trials.

As in the hazard detection measure, scoring high on the blocking car detection measure does not depend on making a correct response in terms of global task performance. In the above example, participants would be credited with 50% blocking car detection on a trial where they crashed. Thus, blocking car detection should reflect participants' awareness of blocking cars more than their decision and action processes.

Global Measures of Driving Performance

Global performance measures were assumed to tap both situation awareness ability and decision and response-execution processes. The main global measure used was crash avoidance, the percentage of hazards safely avoided. A "safe" or correct response for this variable involved avoiding any hazard cars without hitting blocking cars on hazard trials. Because participants responded very accurately on catch trials with a false alarm rate of 2.5%, the percentage of correct responses on catch trials was not a sensitive measure of global performance; and thus was not used.

In summary, the recall measures provide a detailed picture of participants' awareness of the locations of cars some distance from the driver as well as nearby cars. The performance measures of situation awareness focus on participants' knowledge of nearby cars, both in the driver's lane for the hazard detection measure and to the right or left for blocking car detection. These measures are summarized in Table 1. In terms of tests of explicit and implicit memory, the recall measures were direct tests, and the performance measures were indirect tests, as defined by Roediger and McDermott (1993).

Experiment 1

A primary goal of Experiment 1 was to assess the extent to which drivers use explicit and implicit knowledge during maneuvering. The method for doing this was to test whether the recall and performance measures were associated

Table 1
Situation Awareness and Global Performance Measures

Type of measure and variable name	Definition
Situation awareness	
Recall	
% cars recalled	% of actual cars recalled.
Recall error	Average error distance for recalled cars.
Composite recall error	Average error distance for recalled, nonrecalled, and false-alarm cars, with errors to near cars weighted more.
Imbedded-task performance	
Hazard detection	Sensitivity at responding to hazards and not responding on catch trials.
Blocking car detection	Frequency of avoiding some blocking cars on hazard trials.
Global performance	
Crash avoidance	Frequency of avoiding all hazard and blocking cars on hazard trials.

or dissociated by using the performance-recall probes. Performance on the recall and performance measures of situation awareness was expected to be associated (positively correlated). This expectation was based on the research cited above in which explicit knowledge was found to be available very early in the learning process (Perruchet & Amorim, 1992).

The second goal of Experiment 1 was to see how participants' situation awareness for car locations would be affected by changes in working memory load. Thus, across the driving scenarios shown to participants, the number of cars to be tracked and recalled was varied from three to eight. This manipulation promised to show how memory for dynamic spatial information is affected by changes in working memory load. The data analysis concentrated on describing the effect of working memory load on the recall and performance-based measures of situation awareness and the strategies participants used to allocate attention when their working memory was overloaded.

The recall measures of situation awareness measures provide a richer picture than the performance measures of participants' knowledge of traffic at a variety of road locations. Thus, a third goal of the experiment was to explore the recall measures of situation awareness in some depth by comparing two methods of collecting recall data:

recall probes and performance-recall probes. One might expect that having participants focus completely on recalling vehicle locations, as in the recall measure, would give a more complete picture of their explicit car location knowledge than in the dual-task situation where they make driving responses and then recall locations. However, the recall-only measure puts participants in an unrealistic situation. By not requiring participants to make driving responses, it may be distorting their patterns of attending to traffic vehicles and thus distorting their situation awareness. As Endsley (1995a) pointed out, situation awareness measures should not drastically change the participants' situation awareness. Thus the experiment gave information about how the driving task affected recall.

It was expected that when participants had more active control over their simulated vehicle, as with the performance-recall probes, they would attend to the cars around them differently than when they had no control, as with the recall probes. Prior research has suggested specific hypotheses concerning this issue. For example, Triggs and Drummond (1993) found that participants who were actively driving in a simulator were quicker to respond to a light flashed on the simulator screen, which was unrelated to the driving task, than participants who were passively watching the same driving scenes. This

suggests that active control of driving leads to an overall heightened awareness of visual events in the driving field, even ones unrelated to driving. Using a flight simulator, Larish and Anderson (1991) found that active controllers were better at recalling flight variables they were actually controlling (pitch and roll), as compared with passive observers, but were no better at recalling variables not being controlled. Thus, it was hypothesized that with active control (performance-recall probes), participants would be more attentive overall, and thus recall more accurately, or that they would focus attention more on the vehicles most relevant to their driving decisions.

Method

Participants

The 35 participants were hired from temporary employment agencies and ranged in age from 18 to 30 years. All knew how to drive. They took part in the experiment as part of a week-long study involving a variety of cognitive abilities tests. One participant was dropped for lack of effort; he recalled no cars on 45% of the trials, even though each trial had between three and eight cars to be recalled.

Design

A within-subjects design was used with 18 participants (8 men and 10 women) receiving the recall probes on Day 1 and the performance-recall probes on Day 2 and 16 participants (7 men and 9 women) receiving the reverse order.

Materials

The driving simulator software was run on personal computers (486 IBM compatible). The animated scenarios were shown in a window (15.5 cm wide $\times$ 11.4 cm high) centered on the computer screen. To minimize the size of the blind spot, the field of view for the front view was set at 90° horizontal and 66° vertical. These viewing angles corresponded to a station point (center of projection) of 5.4 cm from the screen center. Because participants viewed the screen from about 55.0 cm (instead of 5.4 cm), this resulted in some distortion for cars near the

driver, which appeared elongated. The projection to the three mirrors (see Figure 1) was also based on a station point of 5.4 cm; thus they showed a wider field of view than normal as well. The presentation rate for the animation was 9.1 frames per second, which resulted in relatively smooth vehicle motion. All the traffic vehicles had the same shape (a sedan) and were colored red.

The duration of the animated scenarios varied randomly between 18 s and 35 s. There were from three to eight cars near the driver at the end of the scenarios. The scenarios were designed so that all traffic cars were within 14 car lengths from the driver at the end of the scenario, although during the scenario, cars could be at any distance.

For the recall probes, participants could not respond to avoid the hazards. They simply waited for the animated scenario to end and then recalled the ending car locations. The recall grid showed the road 17 car lengths ahead of the driver and 9 lengths behind. The driver's car was shown in blue (white in Figure 2) in the correct lane. Participants could use "move" and "delete" buttons to move or delete cars they had placed on the road. After participants finished recalling the car locations for a scene, they clicked on the "done" button to receive feedback and then clicked on "done" again to start the next trial.

For the driving task, which was part of the performance-recall probes, the hazards were clearly defined. In all cases, the hazard (or signal) trials involved cars that would collide with the driver unless a driving response was made. Participants could avoid the hazards by using the *up*, *down*, *left*, or *right* arrow keys to accelerate, decelerate, move to the lane on their left, or move to the lane on their right, respectively. The response interval during which subjects could correctly avoid the hazard lasted from the time the hazard car or cars entered the driver's lane until it was too late to avoid them. These intervals were between 1.1 s and 4.1 s. In the catch trials and in the part of hazard trials before the hazard, no cars came near enough to the driver to require a response. The response interval at the end of each catch trial was 3.0 s. The animated scene ended after the first arrow keypress during the response interval, or, if no response was made, at the end of the response interval.

For both recall and performance-recall probes, when the animated scene ended, participants

were presented with the recall grid (see the left panel of Figure 2). Figure 2 filled the entire 38.0 cm computer screen.¹ The use of the recall grid was described above. Participants could place cars anywhere on the recall grid, not just within the grid cells. After placing cars and clicking on the "done" button, participants saw either the grid on the right showing the correct car locations for the recall probes or both the right-hand grid and textual feedback concerning whether their driving response was correct for the performance-recall probes.

Because pilot data showed that participants could not recognize repeated scenarios, the 84 unique scenarios from Day 1 were repeated on Day 2. The same set of 84 scenarios was presented to each participant but in a different random order for each participant and each day. Of the 84 scenarios on each day, half were hazard scenarios. Of the hazard scenarios, 16 involved a car approaching from the front, 16 from the rear, and 10 from both front and rear. The number of side blocking cars in hazard scenarios ranged from 0 to 2. Participants always had at least one response to safely avoid each hazard.

Procedure

Participants were tested in groups of about 20 in a large computer laboratory. All participants receiving a particular probe type were tested together because instructions for how to use the simulator were given verbally to the entire group.

¹ The recall grid was oriented vertically to enhance stimulus-response compatibility; that is, for both the main (front) view in the animated scene and the recall grid, cars farther from the driver were at the top of the screen. However, because objects higher in the picture plane appear farther away than those lower in the plane, people tend to overestimate vertical extents at the top of the picture plane and underestimate those at the bottom (Masin & Vidotto, 1983). Thus, the use of a vertical recall grid could cause participants to underestimate distances ahead of the driver and overestimate those behind. To test for this possibility, I conducted an experiment in which participants adjusted the location of objects at the top and bottom of the screen to be equidistant from a central object. This experiment found that distances at the top of the screen were overestimated by only 1.3% compared with those at the bottom.

After the instructions and a short practice period, each participant completed one block of 42 trials in the morning and one block of 42 trials after lunch. There was about 1 hr of unrelated cognitive tests between, and each block took about 1 hr to complete. On the following day, participants received training for the new type of probes and completed two blocks of 42 trials on a schedule similar to Day 1.

Results and Discussion

Reliability and Validity of Recall and Performance Measures

Recall measures of situation awareness. The recall measures described above included percentage of cars recalled and recall error. For each trial, percentage of cars recalled was the number of cars placed by the participant that were matched to actual cars divided by the number of actual cars (see Appendix A). Recall error was the average error distance for recalled (matched) cars. This variable ignored missed (nonrecalled) cars and false alarms. Thus participants could trade off between percentage of cars recalled and recall error, recalling a few cars quite accurately or many cars inaccurately. Because of this, it seemed useful to have a composite variable measuring participants' recall accuracy that included information about percentage of cars recalled, error distance for recalled cars, and also false alarms. Composite recall error was defined according to the formula below:

$$E = \frac{\sum_{i=1}^{N_H + N_M + N_F} e_{w_i}}{N_H + N_M + N_F}, \quad (1)$$

where N_H , N_M , and N_F are the number of hits, misses, and false alarms, respectively, and e_{w_i} is a weighted error distance.

The error distances were weighted because it seemed that errors in recalling distant cars are less important to good driving and should be weighted less. To determine how to weight error distances, I conducted a regression analysis with the predictor variable being the distance between each actual car and the driver and the dependent variable being participants' error distance in

recalling these cars. Missed (nonrecalled) cars were also included in this regression analysis; they were assigned an error distance of 8.4 car lengths. This was done because in the algorithm to match recalled car positions to actual ones, cars less than 8.4 lengths from an actual car were counted as potentially recalled.

The data for this regression analysis consisted of participants' error distances for every car for every trial in Experiment 1 and for a very similar pilot experiment, a total of about 47,500 data points. The distance of a car from the driver was a highly significant predictor of the error in recalling it, $t(47476) = 102.6$, $p < .0005$, accounting for 18% of the variance. Adding a quadratic term to the regression equation did not account for significantly more variance. The regression equation was

$$e_{p_i} = 0.33 d_i + 1.14, \quad (2)$$

where d_i is the distance between an actual car and the driver and e_{p_i} is the predicted error distance in recalling that car (units = car lengths). The coefficient and constant from this equation were used to weight recall error distances for the composite variable. For each car in a trial, the weighted error distance assigned was

$$e_{w_i} = e_{A_i} - e_{p_i}, \quad (3)$$

where e_{p_i} is the predicted error distance from equation 2. If the car was recalled (a hit), then the actual error distance, e_{A_i} , was the distance between the participant-placed car and the actual car. If the car was not recalled (a miss) or a false alarm, then e_{A_i} was set to the maximum error distance of 8.4 car lengths. Thus, a car's weighted error distance reflected how much its actual error distance exceeded the expected error distance for cars at that distance from the driver.

The composite recall error variable was validated by using it to rank order 12 recall trials in terms of goodness of recall and comparing this ordering to those of three human raters (none of whom knew the matching algorithm or how composite recall error was calculated). The correlations (Pearson's r) between the rank ordering from the composite recall error variable and the human raters (.94, .91, and .88) were as high as the intercorrelations of the raters (.97, .84, and

.79). Thus, the computer algorithm to match recalled with actual cars and the composite recall error measure based on it did a good job of reflecting people's judgments concerning the goodness of participants' recall.

All of the recall variables were calculated for each trial (scenario) and then averaged across trials. Before I analyzed the recall data, trials where it seemed participants were making no effort to perform the recall task were excluded. These included trials where a participant recalled far too few (zero) or far too many cars (eight or more greater than the actual number of cars), where participants placed recalled cars on top of each other or where participants waited too long to recall the first car. The time limit on placing the first car was 15.7 s, which was 3 SDs greater than the mean time for placing the first car. Only 4.5% of the trials were excluded by these criteria.

Participants' recall was first compared with chance levels because some researchers have found that individuals recalled object locations in real-time tasks at chance levels (Fracker & Davis, 1991). For each trial, chance performance was simulated by randomly choosing a number of cars to recall from a distribution that matched the actual distribution of numbers of cars in the experiment and then randomly placing the cars anywhere from 9.0 car lengths behind the driver to 13.5 lengths ahead. Although the response road extended from 9.0 lengths behind to 17.0 lengths ahead in the experimental scenes, all the cars were within 13.5 lengths of the driver at the end of the scene. Thus, the chance performer was assumed to keep within the range of traffic levels in the scenes and the range of traffic car to driver distances. For each participant, the 168 trials of actual participant data were compared with 1,680 chance performance trials.

Table 2 compares actual average levels for the percentage of cars recalled, recall error, and composite recall error variables with chance levels. In Experiment 1, participants recalled more cars and recalled them more accurately than would be expected given chance performance. The group average performance levels were significantly above chance for composite recall error, $t(33) = 8.97$, $p < .001$, and percentage of cars recalled, $t(33) = 5.65$, $p < .001$, and not significantly above chance for recall error, $t(33) = 1.89$, $p < .07$.

Table 2
Actual Versus Chance Performance Levels
for Recall and Performance Measures

Experiment	Actual	95% CI ^a	Chance	95% CI ^a
Percentage of cars recalled				
1	89	2.1	83	0
2	83	2.5	83	0
3	91	1.1	85	0
Recall error (in car lengths)				
1	2.41	.12	2.52	.04
2	2.42	.05	2.60	.01
3	2.33	.07	2.55	.01
Composite recall error (in car lengths)				
1	0.14	.19	0.97	.01
2	0.07	.16	0.61	.01
3	0.06	.11	1.12	.01
Crash avoidance (%)				
1	56	4	30	0
2	55	5	25	0
3	60	4	25	0
Hazard detection (A') ^b				
1	.97	.01	.50	0
2	.97	.01	.50	0
3	.98	.01	.50	0
Blocking car detection (%)				
1	71	6	47	0
2	73	3	48	0
3	76	4	48	0

^aIndicates half widths of 95% confidence intervals (CIs).

^bSignal detection measure of sensitivity for which 0.5 is chance and 1.0 is perfect performance.

To get another estimate of the accuracy of participants' recall, I regressed the actual distance of cars from the driver on their recalled distances using only distances in the depth dimension (i.e., down the road). This regression showed that participants compressed (underestimated) far distances and slightly overestimated close distances. The relation of actual to recalled distances was highly linear with a slope of 0.55 and an intercept of 1.23 ($r^2 = .97$). Depth compression effects of the same magnitude have been found when people make binocular judgments in real scenes of distances of the magnitude simulated in the driving simulator (Loomis, DaSilva, Fujita, & Fukusima, 1992; Toye, 1986; Wagner, 1985).

Thus, even though the participants in Experiment 1 saw a small viewport that presented a wide angle view, their distance judgments followed the same pattern as participants who viewed real scenes.

Next I looked at the reliability and validity of the recall measures. Even-odd reliabilities, after correction with the Spearman-Brown prophecy formula, were .93 for percentage of cars recalled; .92 for recall error; and .96 for composite recall error. To estimate validity, I correlated performance on the recall measures with global performance on the driving task in the performance-recall probes and with general intelligence (g). Percentage of cars recalled, recall error, and composite recall error correlated .48 ($p < .005$), $-.62$ ($p < .001$), and $-.74$ ($p < .001$), respectively, with crash avoidance, the main global measure of driving performance, and .47 ($p < .01$), $-.64$ ($p < .001$), and $-.73$ ($p < .001$), respectively, with g . The value g was measured with the Cognitive Abilities Measurement Battery that was administered to all subjects during the week-long testing period (Kyllonen & Christal, 1990).

Global measures of driving performance. Global performance was measured by crash avoidance, the percentage of hazard trials on which participants safely avoided all hazard and blocking cars. Chance performance on this variable was estimated by randomly selecting one of the five response options (accelerate, decelerate, go left, go right, and do nothing) on each trial. As Table 2 shows, the average crash avoidance was 56% in Experiment 1. All of the 34 individuals avoided crashes at better than the chance level of 30%. The corrected even-odd reliability of the crash avoidance measure was .68, and its correlation with g was .64 ($p < .001$).

Performance measures of situation awareness. The first imbedded task measure of situation awareness was hazard detection (H). This variable was calculated from the hit rate and false alarm rate with the nonparametric signal detection measure of sensitivity (A') in Grier (1971): $H = 0.5 + [(h - f) \times (1.0 + h - f)] \div [4.0 \times h \times (1.0 - f)]$. The hit rate (h) and false alarm (f) rates were the proportion of signal (hazard) and catch trials, respectively, with participant responses during the response interval. For these rates to be comparable, the average duration of the response intervals should have been equal for

hazard and catch trials. In Experiment 1, the average duration of the response intervals was 1.9 s for hazard trials and 3.0 s for catch trials. Thus, participants had more time to emit false alarms, and the false alarm rate was overestimated relative to the hit rate. To correct for this, I multiplied the false alarm rate by $1.9 \div 3.0$ before calculating H . Because the false alarm rate was very low (.038 before correction), this correction had little effect on H .²

Chance performance on the hazard detection measure (with hit rate = false alarm rate) was 0.5, and perfect performance was 1.0. Participants were very good at detecting hazards with a mean level of .97 and a minimum level of .89 (see Table 2). The even-odd reliability of this measure was .80. It had a correlation of .63 ($p < .001$) with global driving performance (crash avoidance) and .60 ($p < .001$) with g .

The second imbedded task measure was blocking car detection. Participants detected the blocking cars 71% of the time on average. Of the 33 participants for whom blocking car detection could be calculated, 28 detected blocking cars above the chance level of 47%. Because blocking car detection was based only on participants' left and right responses and some participants made as few as six of these responses, reliability was not estimated for this variable. However, blocking car detection had a correlation of .79 ($p < .001$) with crash avoidance and .74 ($p < .001$) with g .

In summary, both the recall and performance measures of situation awareness were reliable measures that were positively associated with global driving performance and with g . Most individuals performed above chance on both the recall and performance measures.

Comparing Recall and Performance Measures of Situation Awareness

The first goal of this experiment was to compare the recall and performance measures of situation awareness and look for any evidence that participants were using implicit knowledge of car locations during maneuvering. One piece of evidence for this would have been a dissociation of the scores on the recall and performance measures. Table 3 shows the correlations (Pearson's r) between the recall measures and both the

imbedded task and global performance measures for the combined data from Experiments 1, 2, and 3 ($N = 190$). Experiments 2 and 3 were replications of Experiment 1. The pattern of correlations in Experiment 1 was very close to that shown in Table 3, and the specific correlations for each experiment are presented in Appendix B. The recall and imbedded task measures of situation awareness were significantly correlated. That is, better explicit recall of car locations was associated with (rather than dissociated from) better performance (hazard detection and blocking car detection). Negative correlations occur only when a recall error score was correlated with a performance percentage correct score.

The most likely time for people to use nonconscious implicit knowledge during driving is at high traffic levels where working memory (i.e., explicit knowledge) is overloaded. Therefore, at high traffic levels dissociations might occur between performance and recall measures of situation awareness, as reflected in participants who showed good implicit knowledge but poor explicit knowledge under these conditions. To check for such a dissociation, I correlated the recall and performance measures after blocking for traffic level by using data from Experiment 1. Blocking did not have much effect on the pattern of correlations shown in Table 3. Even at the highest traffic level (seven to eight cars), hazard detection and blocking car detection were correlated with composite recall error ($r = -.35$, $p < .05$, for both); both recall and imbedded task measures correlated positively ($p < .05$) with global performance. Thus, recall and performance measures of car location knowledge were positively associated at all levels of task difficulty.

This is very different from the pattern found by Berry and Broadbent (1984, 1988) and Buchner et al. (1995) for process control tasks where performance was dissociated from recall (with

² In Experiment 3, which used the same stimuli as Experiment 1, the response interval for catch trials was set at the average duration for hazard trial response intervals (1.9 s), so no correction of the false alarm rate was necessary. The uncorrected false alarm rate in Experiment 3 (0.025) was very close to the corrected false alarm rate for Experiment 1 (0.024), which suggests that the correction yielded an adequate estimate of the true rate.

correlations ranging from $-.15$ to $-.31$). The performance-recall associations found here argue against the conclusion that participants were using primarily implicit knowledge to perform the driving task and a different kind of knowledge in the explicit recall task. Rather, they suggest that to a large extent the performance measures were tapping into the same explicit knowledge of car locations as the recall measures.

On the other hand, the moderate size of the performance-recall correlations leaves open the possibility that implicit knowledge has some independence from explicit knowledge in this task. If the performance measures were only reflecting explicit knowledge, one would expect that they would correlate very highly with the recall measures; yet all the recall-performance correlations are less than $.45$ in Table 3 and less than $.62$ for Experiment 1. One explanation for the lack of a stronger association between performance and recall measures was that the performance variables measured participants' knowledge of cars that were very near (hazard and blocking cars), whereas the recall variables measured participants' knowledge of all the cars on the road. To deal with this inconsistency, I also correlated the performance variables with accuracy in recalling nearby cars only. In Experiment 1, hazard detection correlated $-.53$ ($p < .001$) with composite recall error for hazard cars, and blocking car detection correlated $-.44$ ($p < .01$) with composite recall error for blocking cars. Thus, the relationship of the performance and recall measures of situation awareness was one of moderate association, even when both performance and recall variables focused on participants' knowledge of highly salient nearby cars. Given these moderate associations, only more sensitive measures of implicit knowledge can

determine if there is any influence of implicit knowledge on this driving task.

Effects of Working Memory Load on Recall and Driving Performance

The second goal of the first experiment was to see how participants' knowledge of car locations was affected by changes in working memory load, that is, by changes in the number of traffic cars to be recalled. The bird's-eye view of the road in the recall probe extended 17 car lengths ahead of and 9 lengths behind the participant's car. Participants were instructed to recall as many cars as possible that were within these distances at the end of the scene. The number of cars within these distances ranged from three to eight. However, because only about 7% of the trials had either three or eight cars, data for trials with these traffic levels were combined with the adjacent traffic levels, which were four and seven, respectively.

Only percentage of cars recalled was used to investigate memory load effects. The other two recall measures, recall error and composite recall error, were not used because of an unavoidable confound between working memory load and spatial density of traffic cars. As traffic level increased, the density of traffic cars within the fixed-length response area necessarily increased. This meant that recall error actually decreased at higher traffic levels because any car placed by a participant was more likely to be near one of the more densely placed traffic cars. The greater likelihood of close matches at high traffic levels meant that the two recall measures based on error distances gave a biased picture of the effect of memory load. On the other hand, the percentage of cars recalled measure was less affected by this

Table 3
Correlations Between Situation Awareness Measures for Experiments 1, 2, and 3

Measure	1	2	3	4	5	6
1. Composite recall error	—	$-.81^{***}$	$.46^{***}$	$-.34^{***}$	$-.35^{***}$	$-.44^{***}$
2. % cars recalled		—	$-.06$	$.38^{***}$	$.21^{**}$	$.35^{***}$
3. Recall error			—	$-.11$	$-.37^{***}$	$-.36^{***}$
4. Hazard detection				—	$.30^{***}$	$.51^{***}$
5. Blocking car detection					—	$.67^{***}$
6. Crash avoidance						—

Note. $N = 190$.

$^{**}p < .01$. $^{***}p < .001$.

confound because the wide window for counting a car as recalled within eight car lengths meant that for all traffic levels (even low levels), it was fairly easy for a car to be counted as recalled.

Figure 3 shows that participants in Experiment 1 recalled most of the cars at the lowest memory load but that percentage of cars recalled decreased with increases in memory load, $F(3, 96) = 204, p < .001$. Participants recalled an average of 3.7, 4.5, 5.0, and 5.5 cars when the traffic level was four, five, six, and seven cars, respectively. The effect of traffic level differed for the recall-only and the performance-recall probes. Increases in traffic level led to a larger drop-off in percentage of cars recalled for the performance-recall probes than for the recall-only probes, $F(3, 96) = 9.12, p < .001$. These data show that at high traffic levels participants could not track all the cars; therefore they focused attention on a subset of cars. This focusing of attention was accentuated when participants performed the driving task along with the recall task.

In contrast to the recall measures, working memory load had a much weaker effect on both global driving performance and the imbedded task measures of situation awareness. Although crash avoidance declined with traffic level (59%, 54%, 55%, and 49% for Levels 4–7, respectively), this effect was not significant, $F(3, 96) = 1.86, p = .14$. Hazard detection remained at about .97 for each of the traffic levels, although the effect of working memory load on hazard detection may have been hidden by a ceiling effect for this variable. The effect of traffic level could not be estimated for blocking car detection because there were only about 11 hazards per

traffic level and most participants did not make enough left or right responses to calculate this variable based on this small number of trials.

Cues for Allocating Attention When Working Memory Is Overloaded

The data presented above suggest that when working memory is overloaded on a dynamic spatial task, participants focus attention on a subset of objects. This subsection focuses on the factors that affect where participants allocate attention.

Gibson and Crooks (1938) pointed out that drivers use information (*cues*) about other cars on the road—such as speed, position, and direction—to perceptually define a field of safe travel. If this field is unduly encroached on, drivers must respond by changing their own speed or direction of travel. Under the assumption that drivers will attend most to vehicles that are likely to encroach on their field of safe travel, it was hypothesized that drivers will attend most to cars that are nearby, in front of them, and moving toward them. On the other hand, other cues, such as whether a traffic car is on the driver's right or left, were not expected to be used to focus attention. The recall data were used to test these hypotheses. In particular, I tested which cues concerning traffic cars predicted how accurately participants recalled the cars.

This analysis was based on the assumption that the more attention participants allocate to a particular car, the more accurate they will be in recalling it. This follows Cheal, Lyon, and Gottlob (1994), who found that participants' accuracy in a perceptual task reflected the amount of attention allocated to it. Thus, a stepwise regression analysis was conducted with individual traffic cars as the units of analysis. The dependent variable was participants' error distance in recalling each car. The predictor variables included the distance between the driver and the car being recalled, the relative speed of the car being recalled with respect to the driver, whether the car being recalled was in front the driver, and whether it was to the driver's right.

Additional predictor variables were included in this analysis. One was whether the traffic car was the car that would hit the driver on hazard trials; these cars would be expected to be recalled

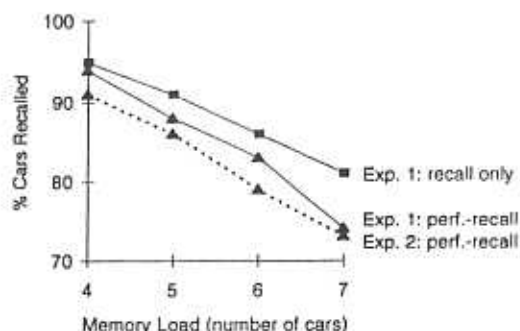


Figure 3. Effect of working memory load on recall for two conditions in Experiments 1 and 2: recall-only and performance (perf.) recall.

with high accuracy. Another predictor was whether the traffic car was in the driver's blind spot. Because cars in the blind spot are near the driver, the above hypothesis suggests they will be recalled accurately. On the other hand, the fact that these cars were out of the driver's sight for up to 10 s before the end of the scene might make them harder to recall accurately. Finally, to account for individual differences in overall recall accuracy, each participant's mean error distance was used as a predictor (see Pedhazur, 1982).

Before presenting the results of this regression analysis, one point should be made concerning the dependent variable. For cars that were judged by the matching algorithm to be recalled, the error distance between the recalled and the actual car was used. For cars participants failed to recall, there was no actual error distance available; yet it seemed important to investigate what cues can predict when participants will not recall a car. Therefore, nonrecalled cars were included in the regression analysis and assigned an error distance of 8.4 car lengths. This was the maximum error distance a recalled car could have according to the matching algorithm.

The results of the regression analysis are shown in Table 4. Because of the large number of degrees of freedom in this analysis (28,184 based on 34 participants $\times$ 168 trials $\times$ about 5 cars per trial), significance testing was too sensitive and thus was not an appropriate measure of the effects of the predictor variables. All the predictor variables were highly significant ($p < .0005$). Therefore, the increment in multiple R for each variable was used to evaluate its effect on the dependent variable. For each predictor variable, Table 4 shows the regression coefficient (B), the order in

which that variable was added in the stepwise regression, and the increment in R . The correlations among the predictor variables were generally low; 17 of 21 were below .20, and all were below .55.

This analysis suggests that participants used cues concerning individual traffic cars to focus attention on the cars most likely to present a driving hazard. In particular, participants focused attention on nearby cars, cars that were not in the blind spot, and, possibly, cars in front of them. Two of the cues the participants used (attending to nearby cars in front) suggest they had been focusing attention on the cars most likely to present a hazard. On the other hand, the participants focused less attention on cars in the blind spot, which were potentially quite hazardous. This was likely due to the much greater working memory demands of tracking blind spot cars.

Effect of Driving Task on Recall

The third goal of the first experiment was to determine how participants' recall of vehicle locations was affected by how much control they had over the driving task. Thus, recall for the more active performance-recall probes and for the more passive recall probes was compared. One hypothesis was that active control would lead to heightened awareness for driving events and thus to better overall recall. However, participants recalled car locations better with the passive, recall probes than with the performance-recall probes, showing a significantly lower composite recall error (0.02 vs. 0.26 car lengths), $F(1, 32) = 16.23$, $p < .001$. Thus, based on this analysis of the data, the hypothesis of better recall

Table 4
Cues Predicting Accuracy in Recalling Individual Cars

Predictor and unit	Experiment 1			Experiment 2		
	B	Order	ΔR	B	Order	ΔR
Distance to driver (car lengths)	0.39	1	.449	0.44	1	.489
Participant mean (car lengths)	1.04	2	.040	1.02	2	.060
In blind spot (1) or not (0)	1.51	3	.036	1.39	3	.028
In front (1) or back (0)	-0.62	4	.012	-0.82	4	.016
Hazard car (1) or not (0)	-0.46	5	.004	-0.43	6	.001
Speed toward driver ^a	-0.02	6	.001	-0.03	5	.004
On right (1) or left (0)	0.10	7	.000	0.18	7	.001

^aIn miles per hour; the approximate metric equivalent of 1 mile is 1.61 km.

with active control, as in the performance-recall probes, was not supported.

Analysis of Larish and Anderson (1991) suggested a further analysis of the data, as Larish and Anderson had found that active control led to better recall of information directly relevant to the control task but not of other information. The information most relevant to the driving task in the performance-recall probes was the locations of cars near the driver. For nearby cars (those less than two car lengths away), the participants recalled better with the more active performance-recall probes, showing a lower composite recall error (0.72 vs. 0.92 car lengths). However, this difference was not significant, $t(33) = 1.80, p < .09$. The data for cars farther than two lengths from the driver showed the same pattern as the overall data, with the recall probes leading to better recall. Thus, in the driving task, active control leads to worse recall of task-irrelevant information (distant cars) but not of task-relevant information (nearby cars). These findings fit with those of Larish and Anderson.

One possible explanation for this result is that for performance-recall probes, participants could have adopted a general strategy of focusing on nearby cars. In this case, then the advantage for performance-recall probes for nearby cars would hold for both hazard and catch trials. Alternatively, participants in the performance-recall probes could have focused attention on nearby cars only after they saw a car that was threatening to hit them. In this case, the recall advantage of performance-recall probes for nearby cars would hold only for hazard trials, not for catch trials, where no car threatened the driver. The data supported the latter explanation. On hazard trials, participants showed significantly lower composite recall error for nearby cars with performance-recall probes (0.81 vs. 1.24 car lengths); $t(33) = 2.96, p < .01$, with distant cars being recalled better with recall-only probes. On catch trials, participants showed better recall with recall probes at all distances.

Thus, having active control of the driving task did lead to better recall of car locations, but only for the cars very near to the driver that were most relevant to driving decisions. The conclusion that drivers focused attention on nearby cars when performing the driving task was also supported by the finding that the decline in percentage of

cars recalled as memory load increased was greater for the performance-recall probes than for the recall-only probes (see Figure 3). However, participants in the active control condition seemed to allocate attention to nearby cars only after a hazard developed. In general, when the data was averaged over all recall distances and trial types (hazard vs. catch), recall was better with the more passive recall-only probes.

The results of Experiment 1 suggest that drivers' knowledge of car locations while maneuvering is largely explicit. The association of the performance and recall measures of location knowledge gave little evidence for the independent use of implicit knowledge. Recall of car locations decreased dramatically with working memory load (i.e., traffic level), as drivers focused attention on a subset of cars. The cues drivers used to focused attention (distance and position) suggest that they were attending more to potentially hazardous cars. Also, drivers focused attention more when they were actively controlling the driving task than when they were not. Experiment 2 was an attempted replication of Experiment 1.

Experiment 2

Method

Participants

The 80 participants were Air Force recruits at Lackland Air Force Base, Texas. They were 17 to 34 years old and took part in the study as part of their basic training. All knew how to drive. Years of driving experience ranged from less than 1 to 16 years, with a mean of 4.4. One participant was dropped for lack of effort; he recalled no cars on 50% of the trials.

Design

All participants (42 men and 36 women) completed 84 performance-recall probes in a single session.

Materials

The same driving simulator and traffic scenarios were used as in Experiment 1 with the

following changes. To increase discriminability, the computer randomly assigned one of five colors to each traffic car before each scenario was displayed, and there were two vehicle shapes: sedan and truck. The recall grid for the recall probes was lengthened to 20 car lengths ahead of the driver and 10 lengths behind. Of the 84 trials on each day, half were hazard scenarios. The response interval for each catch scenario was 3.0 s. Each participant saw the same set of 84 scenarios but in a different random order.

Procedure

Participants were tested in groups of 40 in a large computer laboratory. The instructions and testing were conducted in the same manner as Experiment 1, except that all the testing was completed in a single 3.5-hr session.

Results and Discussion

Reliability and Validity of Recall and Performance Measures

Before analyzing the recall data, I excluded trials where it seemed participants were making no effort to perform the recall task; using the same criteria as in Experiment 1, I excluded 2.9% of the trials. Composite recall error was calculated as in Experiment 1. For Experiment 2, equation 2 was determined based on regression analysis to be: $e_{p_i} = [-0.41 d_i] + 1.04$, where e_{p_i} is predicted error distance and d_i is distance from the driver. Before calculating the A' measure of hazard detection, the false alarm rate was corrected as in Experiment 1 because the average duration of the hazard trials was less than catch trials (1.9 s vs. 3.0 s). The corrected even-odd reliabilities of the measures in Experiment 2 were composite recall error, .97; percentage of cars recalled, .97; recall error, .80; hazard detection, .83; and crash avoidance, .72. As Appendix B shows, all of the recall and performance measures of situation awareness were significantly associated with global driving performance (crash avoidance).

Participants in Experiment 2 showed above-chance performance on all of the performance and recall measures of situation awareness except percentage of cars recalled (see Table 2). Participants as a group were at chance in terms of

percentage of cars recalled but were above chance in terms of their accuracy in locating these cars; for recall error, $t(78) = 6.43$, $p < .001$; for composite recall error, $t(78) = 6.83$, $p < .001$. It seems that subjects in Experiment 2 focused their attention more than those in Experiment 1 and thus recalled fewer cars.

Although percentage of cars recalled was not above chance in Experiment 2, the algorithm for matching participants' cars with actual cars is fairly liberal in allowing any participant-placed car within 8.4 car lengths (38.0 m) of an actual car to be defined as recalled. Thus, the chance level of recalling cars was fairly high (83%). The lower percentage of cars recalled in Experiment 2 was likely due to the fact that this experiment involved all performance-recall trials, which were found in Experiment 1 to have a lower percentage of cars recalled, whereas Experiment 1 had only half performance-recall trials.

Comparing Recall and Performance Measures of Situation Awareness

As shown in Appendix B, Experiment 2 replicated the finding of Experiment 1 that performance and recall measures of situation awareness were significantly but moderately associated. Thus, Experiment 2 supports the same conclusion as Experiment 1. The positive associations suggest that the performance measures are primarily tapping into the same explicit knowledge of car locations as the recall measures; the moderate size of the performance-recall correlations leaves open the possibility that some implicit knowledge is reflected in participants' performance in the driving task.

Working Memory Load and Attention Allocation

As Figure 3 shows, Experiment 2 replicated the effects of working memory load on recall of car locations found in Experiment 1. As traffic level increased, percentage of cars recalled decreased markedly, $F(3, 234) = 317$, $p < .001$, as if participants were focusing attention when overloaded.

Experiment 2 replicated the findings of Experiment 1 concerning how participants use cues from individual cars to focus attention on potentially dangerous cars (i.e., nearby cars in front of

them) but not on cars in the blind spot (see Table 4). The regression coefficients, R increments, and order of cue entry in the stepwise regression were very similar in the two experiments.

One difference between the first two experiments was that in Experiment 2 global driving performance (crash avoidance) declined significantly with memory load, whereas Experiment 1 showed a nonsignificant trend in this direction. Crash avoidance in Experiment 2 was 61%, 55%, 48%, and 55% for Traffic Levels 4–7, respectively, $F(3, 234) = 11.64, p < .001$. Tukey a tests showed the only significant difference ($p < .05$) to be between Traffic Levels 4 and 6. Thus, both the recall and performance measures have been found to be sensitive to changes in memory load. As in Experiment 1, hazard detection remained at about .97 for each of the traffic levels, and the effect of traffic level could not be estimated for blocking car detection because of an insufficient number of trials.

A major conclusion from the first two experiments is that the performance and recall measures of situation awareness were tapping into the same explicit knowledge of car locations, as evidenced by the significant associations between these two types of measures. However, an alternative explanation for these associations follows from the fact that for all of Experiment 2 and half the trials in Experiment 1 (i.e., the performance-recall trials), both performance and recall measures were collected on the same trials. Because of this, both types of measures would be expected to show good performance on trials where participants were attending closely and poor performance when participants were not attending. Experiment 3 was conducted to determine whether the associations between performance and recall measures of situation awareness occur when these measures are collected on separate sets of trials. This should give a more stringent test of whether the performance and recall measures are reflecting the same knowledge base.

Experiment 3

Method

Participants

The 77 participants were Air Force recruits at Lackland Air Force Base, Texas. They were 18 to 25 years old and took part in the study as part of

their basic training. All knew how to drive. Years of driving experience ranged from less than 1 to 9 years, with a mean of 3.6.

Design

All participants completed 84 performance probes and 42 recall probes in one 3.5 hr session. In one session, 25 men and 14 women completed a block of performance probes followed by a block of recall probes. In another session, 24 men and 14 women completed the blocks in the reverse order.

Materials

The same driving simulator and traffic scenarios were used as in Experiment 2. Half of the trials in each block were hazard scenarios. The response interval for each catch scenario was 1.9 s, which is the average duration for the response intervals for the hazard trials. Each participant saw the same scenarios but in a different random order.

Procedure

Participants were tested in groups in a large computer laboratory. The instructions and testing were conducted in the same manner as Experiment 2. All of the instructions and testing for one block of trials was completed before any of the instructions or testing for the second block.

Results and Discussion

Composite recall error was calculated as in Experiment 1. For Experiment 3, equation 2 was determined based on regression analysis to be $e_{p_i} = -0.37 d_i + 0.83$, where e_{p_i} is predicted error distance and d_i is distance from the driver.

As Table 2 shows, participants in Experiment 3 performed above chance on all of the recall and performance measures of situation awareness. Performance on all measures in Experiment 3 was also slightly better than in Experiments 1 and 2, as expected from the fact that in Experiment 3, participants performed both the recall and the driving tasks in isolation, whereas in Experiments 1 and 2, participants performed these tasks together, at least part of the time. The corrected

even-odd reliabilities of the measures in Experiment 3 were composite recall error, .93; percentage of cars recalled, .90; recall error, .84; hazard detection, .46; and crash avoidance, .66. The low reliability for the hazard detection measure occurred because of a ceiling effect for this variable, which was even stronger than in the previous two experiments, probably because the driving task was performed alone.

Appendix B shows that the unreliable hazard detection measure had low correlations with the recall measures. However, the more reliable blocking car detection measure showed a moderate positive association with the recall measures, as in Experiments 1 and 2. The performance measures of situation awareness were also correlated with the measures of recall accuracy for nearby cars. Despite its low reliability, hazard detection correlated $-.25$ ($p < .03$) with composite recall error for hazard cars. Blocking car detection correlated $-.36$ ($p < .002$) with composite recall error for blocking cars. Thus, in Experiment 3 where the recall and performance tasks were done in isolation, I found the same pattern of moderate association between recall and performance measures of situation awareness as in Experiments 1 and 2 where these tasks were performed concurrently. This suggests that the association between recall and performance measures in the first two experiments was not due to the concurrent performance of the two tasks.

General Discussion

Measuring Situation Awareness

The three experiments reported here focused on how people maintain spatial location knowledge (an aspect of situation awareness) while performing a dynamic real-time task. Two methods were used to measure spatial knowledge, recall measures that tapped primarily explicit, conscious knowledge, and performance measures that could assess implicit as well as explicit knowledge.

Two of the performance measures, hazard detection and blocking car detection, are plausible candidates for imbedded task measures of situation awareness. That is, these are performance-based measures that seem to reflect primarily participants' spatial knowledge of nearby

vehicles more than their decision-making and response-execution processes. This emphasis on perceptual and spatial knowledge was attained by assessing participants' ability to respond to and avoid cars in particular locations (e.g., ahead and behind or right and left) without regard for whether participants' overall responses were correct.

Researchers have had a number of problems with recall and performance situation awareness measures. Fracker and Davis (1991) found that their participants were explicitly recalling object locations at only chance levels. Vidulich et al. (1994) found that neither their recall nor their performance measures were sensitive to their experimental manipulation. However, for both the recall and performance measures used in this research, participants exhibited reliable above-chance performance. Moreover, both recall and performance measures were found to be sensitive to experimental manipulations such as varying working memory load that would be expected to affect dynamic spatial knowledge.

There was a slight problem with the hazard detection measure. A ceiling effect for this measure may have dampened its correlation with other situation awareness measures and hidden the effect of increases in working memory load. The ceiling effect can be easily remedied in future research by increasing the difficulty of noticing the hazards in hazard trials and increasing the difficulty of catch trials. Despite the ceiling effect, the hazard detection measure was reliable in two of the three experiments and was significantly correlated with g ($r = .60$; Experiment 1).

Explicit and Implicit Knowledge in Real-Time Tasks

Little evidence was found for the use of implicit knowledge in the simulated driving task. The performance and recall measures of situation awareness showed significant associations, both when performance and recall tasks were done concurrently (Experiments 1 and 2) and separately (Experiment 3). This association is the opposite of the effect found in earlier studies of process control tasks (Berry & Broadbent, 1984, 1988; Buchner et al., 1995). Berry and Broadbent (1984, 1988) suggested that participants were

using separate implicit and explicit knowledge bases to perform the task and answer questions about the task.

However, the current findings fit with more recent research with process control and other tasks (e.g., serial reaction time) where explicit knowledge has been found to track implicit knowledge fairly closely (Buchner et al., 1995; Perruchet & Amorim, 1992). Following the reasoning of Berry and Broadbent (1984), I conclude from the significant performance-recall associations in Experiments 1, 2, and 3 that the performance measures of spatial knowledge are tapping the same explicit knowledge base as the recall measures.

A likely reason for this association was the fact that the knowledge of nearby hazard and blocking cars tapped by the performance situation awareness measures was very important to the driving task. The participants probably knew that this knowledge was important, both while viewing the animated scenes and making their responses. With use of other terminology, it can be said that both the performance and recall probes involved intentional learning of task-relevant knowledge. Prior research has shown that these are the conditions most likely to lead participants to use explicit knowledge and thus show associations of performance and recall memory probes (Berry & Dienes, 1993).

Although the association of performance and recall measures suggests considerable overlap in the knowledge bases used for the driving and the recall tasks, it can be argued that the moderate size of these associations leaves open the possibility that participants are using some implicit knowledge in the driving task. Because the performance measures assess only knowledge of nearby cars, one would most expect a strong performance-recall association when the recall measures also focus on nearby cars. The associations between hazard detection and recall of (nearby) hazard cars and between blocking car detection and recall of blocking cars ranged from .25 to .53 for the three experiments with an average of .35.

Two points should be made about these moderate correlations. First, the correlations underestimate the true ones because they are based on scores that are less than perfectly reliable. Although reliabilities were not available for the

blocking car detection measure, the correlations between hazard car recall and hazard detection can be corrected for unreliability (Ferguson, 1971). This yielded an average correlation of .62 between the performance and recall measures of hazard location for the three experiments. The second point concerning the moderate size of the correlations is, as Perruchet and Amorim (1992) pointed out, that one should not draw major theoretical conclusions from an imperfect correlation because of the many sources of error in psychological experimentation.

One possible reason for the failure to find extensive use of implicit spatial knowledge in the driving task is that the performance measures focused exclusively on knowledge of nearby task-relevant cars. Drivers would have been more likely to use implicit knowledge for less important cars such as those farther away. However, it is difficult to determine how performance-based measures, as used here, can be used to assess knowledge of distant cars. Perhaps, other measures of drivers' knowledge, such as recognition measures, can be used for this purpose.

Effects of Working Memory Load on Dynamic Spatial Memory

Two key aspects of real-time tasks are that people performing these tasks must track a dynamic situation and cope with working memory overloads. Considerable research has been conducted on spatial working memory and on working memory capacity (Baddeley, 1990, 1992). However, almost all of this research has focused on memory for static information, and very little has focused on memory for dynamic spatial information. Pylyshyn and Storm (1988) found that people can track up to five moving objects in a perceptual task, but they did not assess the ability to perceive object locations and distances and did not use a task that tapped working memory. In the driving task, participants can perceptually track some objects but must use working memory to keep track of other objects such as cars in the blind spot.

Both Experiments 1 and 2 showed that percentage of cars recalled decreased in a linear fashion as working memory load increased. Participants seemed to be coping with memory overloads by focusing attention on a subset of the cars. Experi-

ment 1 showed that this focusing was accentuated when participants performed the driving task along with the recall task (i.e., in the performance-recall probes).

The regression analysis for Experiments 1 and 2 showed that when participants' dynamic working memory is overloaded, they focused attention on cars that were nearby, not in the blind spot or in front of them. This analysis is an example of a cue utilization study, which evaluates the extent to which participants' behavior is constrained by environmental cues (cf. Brunswick, 1956; Hursch, Hammond, & Hursch, 1964). Casner (1994) conducted a cue utilization study of the real-time task of piloting a commercial airliner and found that the type of automated flight aids pilots used to respond to air traffic controllers' instructions was constrained by a number of characteristics of those instructions. The cues that participants used in the current experiments suggest they were focusing attention on potentially hazardous cars. Fracker (1990) also found that participants focused attention on hazardous vehicles using a different task—flying—and the same metric of attention used in these experiments (accuracy of location recall).

Despite participants' attempt to focus attention on hazardous cars, their global driving performance also deteriorated at high memory loads, as shown by a significant decrease in crash avoidance in Experiment 2 and a nonsignificant trend in this direction in Experiment 1.

Effects of the Driving Task on Dynamic Spatial Knowledge

One goal of the first experiment was to see how participants' recall of vehicle locations was affected by how much control they had over the driving task. Thus, I compared recall for the more active performance-recall probes where the recall task was performed with the driving task and the more passive recall probes where the recall task was performed alone. The performance-recall probes led to poorer overall recall and to a greater decline in cars recalled with higher memory loads, as compared with the recall-only probes. However, recall with the performance-recall probes was better when participants were recalling nearby cars on hazard trials. These findings

were replicated in a pilot experiment in our laboratory ($N = 46$).

These results suggest that when people have more active control over the driving task, they focus attention on the information necessary for successful driving (i.e., the locations of nearby cars in hazardous situations) more than when they have less control. This replicates Larish and Anderson (1991), who found that active control of a task leads to better awareness and recall of information under active control, but not of other information.

So, which is the better method of assessing peoples' recall of vehicle locations in a dynamic scene: the recall or the performance-recall probes? Although the recall-only probes led to significantly better overall recall accuracy, their advantage was slight: Only 3% more cars recalled and 0.06 car lengths less error per recalled car. On the other hand, there were a number of advantages to the performance-recall probes: participants are in situations more like actual driving, they are more motivated, and, in about the same amount of time as with recall probes, one can collect performance as well as recall data on participants' situation awareness.

Applications of the Situation Awareness Measures

This research has demonstrated that the recall and performance-based measures of situation awareness are reliable measures. The fact that these measures correlated significantly with general cognitive ability (g) and global driving performance in the simulator gives some evidence for their validity. With my colleagues, I am currently conducting a validation study to assess how well the situation awareness measures predict on-road driving performance (e.g., accident and ticket rates). If this study yields positive results, then the simulator-based situation awareness measures should be considered good candidates for a driver licensing battery or for evaluating problem drivers who may need retesting.

The results described above suggest that in developing a more extensive battery of situation awareness tests, the performance-recall probes are more useful than the recall-only probes. If a less extensive situation awareness battery is desired, the association between performance and

recall measures suggests that the shorter performance probes alone would provide useful information about people's situation awareness.

Final Conclusions

In this research, measures of dynamic spatial knowledge were developed. This knowledge is essential for maintaining situation awareness in many real-time tasks such as driving. With these measures, the spatial knowledge used in the simulated driving task was found to be largely explicit with little contribution of implicit knowledge. Working memory limitations seemed to place a strong constraint on drivers' real-time performance, as recall of vehicle locations declined markedly with memory load. Drivers handled memory overloads by focusing attention on the locations of potentially hazardous cars, which was the information most relevant for successful performance. Focusing on hazardous cars was most evident when participants performed a driving task and less evident when they simply watched and recalled driving scenes in a passenger mode. The situation awareness measures have potential applications in selecting people for real-time tasks.

References

- Adams, M. J., Tenney, Y., & Pew, R. (1995). Situation awareness and the cognitive management of complex systems. *Human Factors*, 37, 85-104.
- Baddeley, A. (1990). *Human memory: Theory and practice*. London: Erlbaum.
- Baddeley, A. (1992). Is working memory working: The fifteenth Bartlett lecture. *Quarterly Journal of Experimental Psychology*, 44, 1-31.
- Berry, D., & Broadbent, D. (1984). On the relationship between task performance and associated verbalizable knowledge. *Quarterly Journal of Experimental Psychology*, 36A, 209-231.
- Berry, D., & Broadbent, D. (1988). Interactive tasks and the implicit-explicit distinction. *British Journal of Psychology*, 79, 251-272.
- Berry, D., & Dienes, Z. (1993). Towards a working characterization of implicit learning. In D. Berry & Z. Dienes (Eds.), *Implicit learning: Theoretical and empirical issues* (pp. 1-18). Hillsdale, NJ: Erlbaum.
- Brunswick, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley: University of California Press.
- Buchner, A., Funke, J., & Berry, D. (1995). Negative correlations between control performance and verbalizable knowledge: Indicators for implicit learning in process control tasks. *Quarterly Journal of Experimental Psychology*, 48A, 166-187.
- Casner, S. (1994). Understanding the determinants of problem-solving behavior in a complex environment. *Human Factors*, 36, 580-596.
- Cheal, M., Lyon, D., & Gottlob, L. (1994). A framework for understanding the allocation of attention in location-precued discrimination. *Quarterly Journal of Experimental Psychology*, 47A, 699-739.
- Debnier, J., & Jacoby, L. (1994). Unconscious perception: Attention, awareness and control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 304-317.
- Dorfman, J., & Mandler, G. (1994). Implicit and explicit forgetting: When is gist remembered? *Quarterly Journal of Experimental Psychology*, 47A, 651-672.
- Elander, J., West, R., & French, D. (1993). Behavioral correlates of individual differences in road-traffic crash risk: An examination of methods and findings. *Psychological Bulletin*, 113, 279-294.
- Endsley, M. R. (1995a). Measurement of situation awareness in dynamic systems. *Human Factors*, 37, 65-84.
- Endsley, M. R. (1995b). Towards a theory of situation awareness in dynamic systems. *Human Factors*, 37, 32-64.
- Endsley, M. R., & Bolstad, C. (1994). Individual differences in pilot situation awareness. *International Journal of Aviation Psychology*, 4, 241-264.
- Ferguson, G. (1971). *Statistical analysis in psychology and education*. New York: McGraw-Hill.
- Flinn, D. (1965). Functional states of altered awareness during flight. *Aerospace Medicine*, 36, 537-544.
- Fracker, M. L. (1990). Attention gradients in situation awareness. In *Conference proceedings of the Advisory Group for Aerospace Research and Development: Situational awareness in aerospace operations* (Vol. 478, pp. 6-2-6-10). Copenhagen: NATO Advisory Group for Aerospace Operations.
- Fracker, M. L., & Davis, S. A. (1991). *Explicit, implicit, and subjective rating measures of situation awareness in a monitoring task* (Technical Report AL-TR-1991-0091). Wright-Patterson Air Force Base, OH: Air Force Systems Command.
- Gibson, J., & Crooks, L. (1938). A theoretical field

- analysis of automobile driving. *American Journal of Psychology*, 51, 435-471.
- Grier, J. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin*, 75, 424-429.
- Hirsch, C., Hammond, K., & Hirsch, J. (1964). Some methodological considerations in multiple-cue probability studies. *Psychological Review*, 71, 42-60.
- Johnston, W., & Dark, V. (1986). Selective attention. *American Review of Psychology*, 37, 43-75.
- Kyllonen, P., & Christal, R. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, 14, 389-433.
- Larish, J., & Anderson, G. (1991). Active control versus passive observation in a simulated flight task. In *Proceedings of the Human Factors Society 35th Annual Meeting* (pp. 1570-1573). Santa Monica, CA: Human Factors Society.
- Loomis, J., DaSilva, J., Fujita, N., & Fukusima, S. (1992). Visual space perception and visually directed action. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 906-921.
- Masin, S., & Vidotto, G. (1983). A magnitude estimation of the inverted-T illusion. *Perception and Psychophysics*, 33, 582-584.
- Pedhazur, E. (1982). *Multiple regression in behavioral research: Explanation and prediction*. New York: Holt, Rinehart & Winston.
- Perruchet, P., & Amorim, M. A. (1992). Conscious knowledge and changes in performance in sequence learning: Evidence against dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 785-800.
- Pylyshyn, Z., & Storm, R. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, 3, 179-197.
- Roediger, H., & McDermott, K. (1993). Implicit memory in normal human subjects. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (Vol. 8, pp. 63-131). Amsterdam: Elsevier.
- Sanford, A., & Garrod, S. (1981). *Understanding written language*. New York: Wiley.
- Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 501-518.
- Toye, R. (1986). The effect of viewing position on the perceived layout of space. *Perception and Psychophysics*, 40, 85-92.
- Treat, J., Tumbas, N., McDonald, S., Shinar, D., Hume, R., Mayer, R., Stansifer, R., & Castellan, N. (1979). *Tri-level study of the causes of traffic accidents: Executive summary* (National Technical Information Services Technical Report No. DOT HS-805 099). Bloomington: University of Indiana.
- Triggs, T., & Drummond, A. (1993). A young driver research program based on simulation. In *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting* (pp. 617-621). Santa Monica, CA: Human Factors and Ergonomics Society.
- Vidulich, M., Stratton, M., Crabtree, M., & Wilson, G. (1994). Performance-based and physiological measures of situational awareness. *Aviation, Space, and Environmental Medicine*, 65(Supplement), A7-A12.
- Wagner, M. (1985). The metric of visual space. *Perception and Psychophysics*, 38, 483-495.
- Watson, C., & Nichols, T. (1976). Detectability of auditory signals presented without defined observation intervals. *Journal of the Acoustical Society of America*, 59, 655-667.
- Willingham, D., Nissen, M., & Bullemer, P. (1989). On the development of procedural knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 1047-1060.

Appendix A

Algorithm for Matching Recalled and Actual Cars

The goal is to match the cars placed by the participant (P) on the recall grid with the actual locations of the cars at the end of the scene when given only the location coordinates of the recalled and actual cars. In Figure A1, A2 through A5 represent the locations of the four actual (A) cars at the end of a scene, and P1 through P4 represent the locations of all the cars placed by the participant for that scene. The letters DR stand for the driver's (i.e., the participant's) car. The large rectangle represents the recall grid and is not to scale. Thus, Cars A3 and A5 were beyond the recall grid. The numbers in each car show how the matching algorithm would match the cars given this configuration (e.g., A2 is matched to P2). Recalled Car P1 was not matched to an actual car; that is, it was a false alarm. The matching algorithm follows these steps:

1. Each actual car is paired with all the recalled cars within 8.4 car lengths. All distance units here are in car lengths. This pairing is done even for actual cars outside of the recall grid, which allows participants to get credit for recalling cars that were just outside the recall grid. Thus, A3, A4, and A5 would each be paired to both P3 and P4; A2 would be paired to both P1 and P2.

2. From these pairings, all possible ways of assigning the recalled cars to the actual cars are generated. For example, A2, A3, A4, and A5 could be matched with P2, P3, P4, and P0, respectively, or with P2, P3, P0, P4, respectively, where P0 represents matching an actual car to no recalled car (i.e., a miss).

3. For each possible assignment of recalled cars to actual cars, error score S is calculated:

$$S = \sum_{i=1}^{N_H + N_M + N_F} e_{w_i} \quad (A1)$$

where N_H , N_M , and N_F are the number of hits, misses, and false alarms, respectively. The weighted error distance, e_{w_i} , is given by

$$e_{w_i} = e_{A_i} w_{D_i} w_{L_i} w_{B_i} \quad (A2)$$

If a participant-placed car is matched with an actual car, then e_{A_i} is the distance between the two cars. If no participant car is matched to an actual car (a miss) or there is an unmatched participant car (a false alarm), then e_{A_i} is set to the error distance of 8.4 car lengths, which is the maximum error distance allowed for a participant car matched to an actual car. The variables w_{D_i} , w_{L_i} , and w_{B_i} weight the error distance associated with a car based on its distance from the driver and its lane, and whether it is in the blind spot. More weight is given to error distances associated with near cars, cars in nonadjacent lanes, and cars in the blind spot. For hits, the distance weight, w_{D_i} , is set to $1.0 - 0.0335 d_i$, where d_i is the distance of the actual car from the driver. For misses and false alarms, w_{D_i} is set to 1.0. The lane weight, w_{L_i} , is 1.0, 1.5, and 3.0 if the matched cars are in the same, adjacent, or nonadjacent lanes, respectively, and 1.0 in the case of a miss or false alarm. The blind spot weight, w_{B_i} , is 1.5 if the actual car is in the blind spot, and 1.0 otherwise. This weighting scheme reflects participants' tendency to recall near cars more accurately, to recall cars in their correct lane, and fail to recall cars in the blind spot.

4. The assignment of recalled and actual cars with the lowest error score (S) is chosen as the best match. In the above example, this would be A2-P2, A3-P3, A4-P4, A5-not matched, and P1-not matched.

5. Any actual cars outside the recall grid that are not matched to recalled cars in the chosen

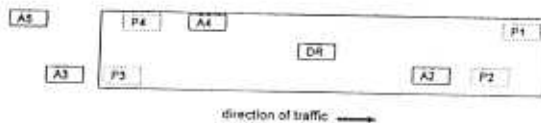


Figure A1. Example locations of actual and recalled cars on the recall grid: A2, A3, A4, and A5 indicate actual car positions; DR = driver's car; P1, P2, P3, and P4 = locations of cars placed by participant.

matching (e.g., A5) are dropped from consideration. The final matching is A2-P2, A3-P3, A4-P4, and P1-not matched. Thus, the participant is given credit for recalling Car A3, which is outside the recall grid but is not penalized for failing to

recall car A5 because participants are not supposed to recall cars outside the grid. The variables discussed in the text—composite recall error, percentage of cars recalled, and recall error—were calculated from this final matching.

Appendix B

Correlations Between Situation Awareness Measures for Experiments 1, 2, and 3

Measure	1	2	3	4	5	6
1. Composite recall error	—	-.77*** -.95*** -.79***	.74*** .14 .75***	-.55*** -.40*** -.07	-.61*** -.24* -.37***	-.69*** -.42*** -.39***
2. Percentage of cars recalled		—	-.17 .16 -.22	.44** .37*** .10	.34* .14 .29*	.46** .31** .36***
3. Recall error			—	-.39* .01 .05	-.57*** -.28* -.28*	-.58*** -.28* -.28*
4. Hazard detection				—	.53** .15 .37***	.63*** .50*** .44***
5. Blocking car detection					—	.79*** .59*** .69***
6. Crash avoidance						—

Note. Data for Experiments 1 ($n = 34$), 2 ($n = 79$) and 3 ($n = 77$) are shown on Lines 1, 2, and 3, respectively.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Received March 11, 1996

Revision received July 18, 1996

Accepted July 25, 1996 ■