CALIFORNIA STATE UNIVERSITY, NORTHRIDGE

THE EFFECTS OF DRIVER DISTRACTION ON SIMULATED DRIVING: ARE MULTIPLE RESOURCE THEORY PREDICTIONS MEDIATED BY CENTRAL EXECUTIVE INVOLVEMENT?

A thesis submitted in partial fulfillment of the requirements for
the degree of Master of Arts in Psychology,
Human Factors and Applied Experimental

by

Maya L. Uhl

December, 2003

The thesis of Maya L. Uhl is approved:

10-8-03

Donald C. Butler, Ph.D. Date

10-8-03

Mr. Barry L. Berson, M.A. Date

Jill L Quilici 10-8-03

Jill L. Quilici, Ph.D., Chair Date

California State University, Northridge

ACKNOWLEDGEMENTS

I would like to thank my committee members for their support throughout my thesis process. They each gave of their time and expertise freely, were responsive to my needs, and still permitted me autonomy. Their approach allowed me to apply what I had learned in the program and to assess my strengths and weaknesses – an invaluable process for which I will always be grateful.

I am especially indebted to Dr. Jill Quilici for her guidance and support. Her encouragement, patience, and kindness have been unparalleled. Thank you for being there for me every step of the way with unconditional support and contribution. Your efforts did not simply facilitate my thesis process; they allowed me a truly worthwhile culminating experience. You are a person of immense integrity, dedication, thoughtfulness, and kindness. I feel blessed to have worked with you.

I wish to thank Dr. Donald Butler for freely offering his time, expertise, and encouragement. Thank you for your input on this project, and more importantly, for teaching me the skills to design, execute, and analyze complex experimental questions. This knowledge will serve me throughout life, regardless of career decisions or life choices. Thank you for your passion and dedication to teaching.

I wish to express my sincere gratitude to Barry Berson for consistently offering his time, expertise, and unwavering encouragement. You have inspired me, challenged me, and been a primary source of support and innovation during my tenure in the program. Thank you for your enthusiasm about the field and your commitment to my success.

I would also like to acknowledge two people who were pivotal to my entry into the field of Human Factors Psychology. Dr. Mark Sanders initially introduced me to the field as an undergraduate and was my first source of inspiration and encouragement. Thank you for opening this new world to me. I credit Dr. John W. Senders with my final decision to enter the field and my resolve during the program. Thank you for your timely intervention, friendship, and support. I could not have asked for a better volunteer guidance counselor.

I would also like to thank Dr. Tyler Blake for his continuous encouragement and support, and Eva Wahlroos, my answer woman, who watched over my progress and always pointed me in the right direction. It is impossible to thank everyone who encouraged me, aided my endeavors, or inspired me. The people mentioned above have been the most influential during my years at California State University, Northridge. Thank you for your support, inspiration, and for dear memories that will last a lifetime.

I'd like to thank my husband, Tad, for his steadfast belief in me, and his unselfish commitment to my education. You have always understood that completing this degree was an important goal in my life, and you have protected it from interference. Thank you for your incredible support and love. I could not have done it without you.

I'd like to thank my parents for their love and support throughout my education, and for instilling in me traits that have enabled me to achieve this goal. Mom, thank you for giving me a thirst for knowledge and strong social morals. Dad, thank you for your work ethic and integrity. Finally, I'd like to thank my family and friends for their support and understanding. I promise I won't miss so many birthdays and get-togethers in the future!

DEDICATION

To my partner in life, Tad.

Without your support, this would still be a dream.

TABLE OF CONTENTS

ABSTRACT

THE EFFECTS OF DRIVER DISTRACTION ON SIMULATED DRIVING: ARE MULTIPLE RESOURCE THEORY PRECIDTIONS MEDIATED BY CENTRAL EXECUTIVE INVOLVEMENT?

by

Maya L. Uhl

Master of Arts in Psychology,
Human Factors and Applied Experimental

The present study was conducted to investigate the impact of central executive involvement on the predictions of multiple resource theory (MRT; Wickens, 1984) with regard to performing concurrent secondary tasks with simulated driving. MRT looks for overlaps between input-processing-output modalities, but does not include consideration of central executive demands on attention. The present study tests a new model, central executive mediating (CEM), in which the central executive uses attention whenever activated. Central executive demand was varied between tasks considered to be automatic (low demand) and controlled (moderate or high demand). Since driving is a visual-input manual-output task, visual demand and manual manipulation were also varied between tasks. Forty-eight undergraduate participants each completed two trial simulations, one established baseline driving, and the other was a dual-task trial with a concurrent secondary task (either eating, changing CDs, or performing a memory search task). Each participant experienced both trials with either low or high traffic complexity. The driving simulator recorded performance measures, including number of collisions, number of road departures, number of speed violations, and mean speed. A 3 (task) by 2 (complexity) by 2 (gender) MANOVA and several ANCOVAS were conducted using the change scores between dual-task and baseline driving. The number of years licensed to drive and computer/video game experience were used as covariates. Four hypotheses were tested. (1) All dual-task trials were expected to result in more driving errors than baseline driving. This was not completely supported. Errors increased with eating, but not with the memory task, and CD-changing only resulted in an increase in road departures. (2) A main effect of complexity was expected, with more errors occurring in high complexity; however, this was not supported. Errors were not statistically different between complexity levels. (3) Mean speed was expected to be lower during dual-task trials to compensate for the additional mental effort, and mental effort was expected to correlate with mean speed. This hypothesis was also not supported. Neither mean speed nor mental effort were significantly different between baseline and dual-task trials, nor were they correlated. (4) Results were expected to support the CEM model. While this was not entirely supported, the results posed a problem for MRT. Several problems were encountered which highlight areas for future study. Mental effort and complexity may not have been sufficiently manipulated, and though restricted randomization was used in assigning participants, groups were not equivalent, most significantly with regard to the number of miles they reportedly drove per week. Taking these issues into consideration, a pattern emerged in the results which modified the model. The new CEM-overlap model incorporates central executive demand into MRT as a processing modality and requires that both tasks vie for central executive involvement in order for interference to occur.

Introduction

Driver distraction has been the subject of much recent research. This has been due primarily to the use of cellular phones and the evolution of computer technology that makes a variety of in-vehicle information and automation systems possible. Interface designs for these systems will require an evaluation of how much they interfere with or facilitate the driving task.

Driving distraction can be operationalized as a measure of decrement in performance of the primary driving task that is attributable to the addition of a concurrent secondary task. Capacity theories of attention assume that attentional capacity is limited, but a limited capacity does not imply an inability to do two or more tasks simultaneously. Rather, it predicts that several demanding tasks cannot be carried out at the same time without negative interference. Wickens, Sandry, and Vidulich (1983) assert that the degree of similarity between the tasks is a key indicator of negative interference. This assertion, known as *multiple resource theory* (MRT) describes separate resource pools for different modalities, and predicts that less interference will occur when performing dissimilar sensory and motor tasks concurrently than when performing tasks that require the same modalities.

Wickens et al. (1983) and Wickens (1984) proposed perceptual codes as either visual or auditory, processing codes as either spatial or verbal, and associated responses as manual or vocal. Consequently, if we consider driving as primarily a visual-spatial-manual task, secondary tasks which involve these modalities should interfere more with driving than secondary tasks which involve other modalities (e.g., auditory-verbal-vocal). Wickens' model suggests the importance of defining the input, processing, and response modalities required by each task. Further, it uses the overlap between these modalities to predict where performance decrements will occur.

All these tasks take place in working memory, and as such, MRT shares many features with the working memory model expounded by Baddeley (1986). The working memory model consists of three components: (1) a *phonological loop* that maintains and manipulates speech-based information, (2) a *visuospatial sketchpad* that maintains and manipulates visual or spatial information, and (3) a *central executive* that integrates information and selects processing strategies. Note that MRT has incorporated the first two components, while ignoring the third. More recently, Baddeley (1996, as cited by Wickens and Hollands, 1999) has concentrated on the functioning of the central executive component, describing four core functions: (1) to coordinate performance on multiple tasks, (2) to temporarily hold and manipulate information stored in long-term memory, (3) to change retrieval strategies from long-term memory, and (4) to attend selectively to stimuli.

Baddeley (1996, as cited by Wickens and Hollands, 1999) proposed that executive processing is controlled processing, whereas tasks that are automatic are allocated to the subsystems. This implies that tasks drawing on the central executive will only receive interference from tasks that cannot be performed automatically. Automatic processing has been described as a generally fast, parallel, and fairly effortless process that is not limited by short-term memory capacity (Shiffrin and Schneider, 1984; Schneider, Dumais, and Shiffrin, 1984; Ruthruff, Johnston, and Van Selst, 2001). Automatic processing typically develops when participants process stimuli in a consistent fashion over many trials. It is not under our direct control, and once learned, is difficult to suppress, modify, or ignore (Shiffrin and Schneider). By contrast, controlled processing is often slow, generally serial, effortful, capacity limited, under our control, and is used to deal with novel or inconsistent information (Shiffrin and Schneider). It is needed in situations where the required responses vary from trial to trial or between situations. Controlled processing is easily modified, suppressed, or ignored (Shiffrin and Schneider; Schneider et al.). Shiffrin and Schneider further assert that all tasks are carried out by complex mixtures of controlled and automatic processes used in combinations.

This is particularly important to the present discussion since learning to drive a car can be considered a mixture of controlled and automatic processes, with the mixture of the two varying with experience (Schneider et al., 1984). The inexperienced driver learns to control the vehicle by employing a conscious and voluntary mode of control, which is demanding on many resources, including visual-spatial perception and processing, manual manipulation, and decision making (Briem and Hedman, 1995), leaving little capacity for other tasks, including unexpected road events. In contrast, the experienced driver is capable of performing the basic control tasks automatically without appreciable voluntary effort, and has additional capacity to attend to other tasks, such as unexpected road events, listening to the radio, and talking. Thus, the control tasks of driving become automatic with practice.

Schneider et al. (1984) describe driving as a complex task that encompasses a mixture of automatic and controlled processes that they suggest may be organized in a systematic network with many of the automatic processes operating in parallel. According to Michon (1985, as cited in Briem and Hedman, 1995), the driving task can be conceptualized in three levels: (1) the lowest level includes automatic, sensory-motor sequences, such as steering, accelerating and braking, (2) the tactical level includes attending to traffic signs, pedestrians, and other vehicles, and (3) the strategic level includes route selection and temporal planning. This model is similar to the one proposed by Rasmussen (1986) addressing the control of human actions in human-system interaction. Rasmussen distinguishes three levels of human performance: skill-based, rule-based, and knowledge-based behavior. Performance at the skill level is based on the processing of automated sensory-motor patterns, at the rule level, on handling familiar task problems, and at the knowledge level, on dealing with novel problems that occur during the task.

This discussion serves to illuminate the complexity and components of the primary driving task. Much of the literature dealing with driver distraction has supported predictions based on MRT (Martens and Van Winsum, 2000; Mollenhauer, Hulse, Dingus, Jahns, and Carney, 1997; see Wickens and Seppelt, 2002, for a literature review). Consequently, the trend in research and development has shifted toward auditory and speech-based interfaces. However, there is mounting evidence that speech-based tasks interfere with driving as well. A statistical analysis of accident data showed that drivers with cellular phones are four times more likely to be involved in a crash and that hands-free models provide no significant safety benefit (Redelmeier and Tibshirani, 1997). Recent studies show increases in driver response times during telephone conversations (Alm and Nilsson, 1995; Strayer and Johnston, 2001), and studies investigating various visual and speech tasks while driving have found similar increased reaction times with speech tasks to those requiring moderate-to-high visual demands (Lee, Caven, Haake, and Brown, 2000; Olsson and Burns, 2000). These results have led several researchers to suggest that central executive processing limits may play a more prominent role than previously thought (Briem and Hedman, 1995; Hegarty, Shah, and Miyake, 2000; Lee et al., 2000; Strayer and Johnston, 2001).

Previous Research

Strayer and Johnston (2001) conducted two studies to investigate the nature of distraction in cell phone use while driving. To represent driving, they used a pursuit-tracking task in which participants used a joystick to control the position of a cursor presented on a computer screen. The object was to keep their cursor as closely aligned with a moving target as possible. Occasionally, the target would flash green or red. Participants were told to press the "brake button" located in the thumb position on the joystick each time the target flashed red. In Experiment 1, participants each saw three phases of the experiment: a practice phase, a single-task phase (the pursuit tracking task), and a dual-task phase (the pursuit tracking task with a concurrent secondary task). Three secondary tasks were varied between participants and consisted of listening to the radio, conversing with a confederate over a hand-held cell phone, and conversing with a confederate over a hands-free cell phone. Participants in the conversation conditions were asked to discuss the impeachment of President Clinton or the Olympic Committee bribery scandal, both hot topics at the time. "Braking" reaction time and missed signals were measured.

Comparing the single-task scores with the dual-task scores, twice as many red signals were missed in both conversing conditions (which were not significantly different from each other) than in the listening condition, which yielded the same results in both the single and dual-task phases. Similarly, reaction time was significantly slower with the conversing conditions than the listening condition.

In Experiment 2, Strayer and Johnston (2001) wanted to explore the source of the interference conversing exerted on the pursuit-tracking task. They did this by using two tasks, a word-shadowing task and a word-generation task. Both were presented on hand-held cell phones while the participant performed the same pursuit tracking task and three experimental phases used in Experiment 1. In the word-shadowing task, the experimenter presented a word verbally to the participant through a hand-held cell phone, and the participant was asked to repeat the word. In the word-generation task, the experimenter presented a word verbally to the participant through a hand-held cell phone, and the participant was asked to generate a new word that started with the last letter of the given word. In this way, the only thing varied between the two conditions was the cognitive effort of word-generation over that of word-shadowing.

Results showed no significant differences between the single-task and dual-task with word-shadowing, but a significant difference was obtained between the single-task and dual-task in the word-generation condition. Strayer and Johnston (2001) proposed their results were consistent with an “attention-based” interpretation of cell phone use in which errors are due primarily to the diversion of attention from the driving task to the phone conversation itself. They argued that these results were inconsistent with multiple-resource models which would have predicted similar patterns of dual-task interference for the two conditions in Experiment 2. Instead, Strayer and Johnston suggest that cell phone use may disrupt performance by diverting attention to an engaging cognitive context other than the one immediately associated with driving.

Strayer and Johnston's (2001) results pose two questions with which the present study is concerned. First, would their results hold with a simulated driving task? Since Strayer and Johnston used a pursuit-tracking task, a task that lacks the complex environment of a driving scene, the generalizability of their results to a more realistic driving task is unknown. Secondly, how does the attention-based interpretation fit with or refute MRT predictions? Does the central executive play a role in complex dual-task performance not predicted by MRT? In looking through the literature, it becomes apparent that many cognitive tasks used in studies of driving distraction have exerted little demand, such as counting backwards (Olsson and Burns, 2000), or conversing at one's own pace (Breim and Hedman, 1995; Jenness, Lattanzio, O'Tolle, and Taylor, 2002). Since Strayer and Johnston used two auditory-verbal-vocal secondary tasks in Experiment 2, and varied the cognitive load, it would be interesting to compare a high-load auditory-verbal-vocal secondary task with two visual-spatial-manual tasks – one automatic and one controlled, to vary central executive involvement. Would both modality combinations cause similar interference when cognitive complexity is high, indicating that central executive involvement plays a role in interference with the driving task? Or will MRT predominate regardless of mental load, indicating that the use of similar modalities and resource pools predicts the most interference?

Olsson and Burns (2000) compared three secondary tasks while driving on a country road in Sweden. One task was visual-auditory-manual (tuning a radio), one was visual-spatial-manual (changing CDs), and one was auditory-verbal-vocal (backward counting). As a tertiary task, participants were asked to perform a peripheral detection task (PDT). The PDT requires participants to press a button each time they detect a target in their peripheral field of view (a visual-spatial-manual task). Olsson and Burns expected that the visual tasks would increase reaction time (RT) and missed signals to the tertiary task. However, backward counting was found to increase mean RT most significantly and, along with CD-changing, produced the most missed signals.

Though these results show promise for a central executive mediating model (CEM), they cannot be considered reliable since the PDT, which facilitated all the dependent measures, was found to be problematic. The peripheral targets were displayed on a head-up display (HUD) at random intervals, but were often difficult to detect and thus display problems may have accounted for many missed signals. Additionally, the PDT method itself may be unreliable as a measure of distraction. As in Olsson and Burns (2000), the PDT is commonly presented as a tertiary task and provides measurement of RT and missed signals. These measures are interpreted as an indication of the secondary task's interference with the primary driving task. However, this is not the only logical explanation. The same measure could be interpreted in a number of ways. Missed signals on the tertiary task may indicate a proper allocation of attention on the part of the driver, as it could be asserted that a missed target indicates that the PDT task did not interfere with the primary driving task. Further, a tertiary measure can only indicate whether attentional capacity is overloaded by the combination of tasks. It cannot decipher where performance decrements may exist – in the primary task, the secondary task, or only in the tertiary task itself. Only performance measures of the tasks can yield this information.

Jenness et al. (2002) measured driving performance with four secondary tasks during simulated driving. One task was visual-verbal-vocal (reading driving directions aloud), one was auditory-verbal-vocal (voice dialing a cellular phone), one was controlled visual-spatial-manual (changing CDs), and one was automatic visual-spatial-manual (eating a cheeseburger). Driving time, minimum speed violations (driving below 10 mph), and lane-keeping errors were measured in a counterbalanced within-subjects design. Reading driving directions aloud and changing CDs produced the most errors and the slowest driving times and were not statistically different from each other. Voice dialing and eating a cheeseburger produced fewer errors and faster driving times and were not statically different from each other. Driving without performing a secondary task yielded the fewest errors and fastest driving time.

This pattern of results yields interesting implications for MRT and CEM models. While all secondary tasks interfered with the primary driving task, the two controlled secondary tasks with visual-input demands (reading driving directions aloud and changing CDs) yielded the most interference, despite different processing and output modalities. However, the automatic visual-input task (eating a cheeseburger) yielded better performance in line with the auditory-input task (voice dialing). This lends evidence that automatic tasks are not limited by short-term memory capacity (Shiffrin and Schneider, 1984; Schneider et al., 1984). Thus, secondary tasks drawing on the central executive may mediate effects of interference, in proportion to their central executive demands, to the effects predicted by MRT (Wickens et al., 1983; Wickens, 1984), while the interference of automatic secondary tasks may depend only on their input-processing-output modality overlaps with the primary task.

While Jenness et al. (2002) findings are compelling, it is important to note that they did not interpret their results through MRT or CEM models, nor any theoretical model. Their discussion simply addressed their results and discussed implementation problems. One problem they encountered affected the voice dialing condition. Participants were told by the experimenter to voice dial certain numbers and relay information back to the experimenter. For example, the participant was told the number for the weather service and asked to voice dial the number and tell the experimenter what the weather forecast was. However, they experienced numerous problems obtaining and maintaining cellular phone signals long enough to complete the tasks.

The simulator used in Jenness et al. (2002) was a Sony PlayStation™ with steering wheel and foot pedal controls. The simulation consisted of one of the tracks in the Grand Turismo™ racing game with all other cars removed from the track. Thus, the simulation consisted of a single road and did not include normal driving cues such as intersections, traffic controls, other vehicles, pedestrians, or unexpected events. Whether their results would hold during an interactive driving event will require further research.

Olsson and Burns (2000), Strayer and Johnston (2001), and Jenness et al. (2002) did not control for differential auditory distraction. When a CD player or radio was used, the auditory stimuli were presented as well, which would have created an additional demand in those conditions. This demand was neither controlled nor assessed. The content of the radio and its volume varied by participant since the participant controlled tuning. Similarly, participants were able to control CD content and volume as well, with one study offering a choice of music or recorded literature (Strayer and Johnston).

Current Study

The studies discussed above suggest an emerging pattern. Strayer and Johnston (2001) suggest that cognitive demand of the secondary task is important, though they did not test these results in conjunction with a visual-input task which would offer information about relative ranking. Olsson and Burns (2000) compared a cognitive task with a visual task and found support for equal interference with the driving task – though this was determined using the questionable tertiary measure of the peripheral detection task (PDT). Jenness et al. (2002) found evidence that MRT predictions could be mediated by the level of central executive involvement required by the task – controlled vs. automatic. However, their driving task lacked important components for generalization. The primary question remains: with regard to the driving task, do automatic secondary tasks depend only on their input-processing-output modality overlaps with the primary task as predicted by MRT (Wickens et al., 1983; Wickens, 1984), while secondary tasks drawing on the central executive exert mediating effects of interference, in proportion to their central executive demands?

To test this, the current study compares three secondary tasks during simulated driving: (1) a controlled visual-spatial-manual task (changing CDs), (2) an automatic visual-spatial-manual task (eating), and (3) a controlled auditory-verbal-vocal task (memory search task, to be described later). A control condition in which participants simply perform the primary driving task will establish baseline performance. This study will differ from previous studies by using an interactive driving scenario constructed to represent realistic driving conditions and events that participants could expect to encounter during an average driving experience. The present study intends to manipulate input, processing, and output modalities, and mental effort in search of contrary or favorable evidence of central executive involvement and its mediating effects upon the predictions of multiple resource theory as it applies to simulated driving.

Olsson and Burns (2000) found both a main effect and interactive effects by varying driving complexity (traffic density). Reaction times increased with higher density. The current study includes a between-subject complexity manipulation in which each task is presented with both high and low traffic conditions.

Figure 1a illustrates the expected outcome of this study if the central executive mediating model (CEM) is supported. The automated task (eating) is expected to yield the best driving performance, with minimal interference to the driving task caused by the shared input-output modalities. The controlled auditory-verbal-vocal task (to be discussed later) will involve central executive processing and therefore is expected to impose significant interference with the primary driving task. It is predicted that the interference will be similar to that of the controlled visual-spatial-manual task (changing CDs). These predictions are drawn from the findings of Olsson and Burns (2000), Strayer and Johnston (2001), and Jenness et al. (2002).

If CEM is not supported, and the results instead support MRT predictions, then the outcome should resemble Figure 1b. Since driving has been classified as primarily a visual-spatial-manual task, the auditory-verbal-vocal task should exert the least interference with the primary driving task. The two visual-spatial-manual tasks (eating and changing CDs) would be expected to yield the most interference since input-processing-output modalities overlap. Because the visual and manual demands of changing CDs will exceed those of eating, the CD-changing condition is expected to yield the worst driving performance.

Hypotheses. The following hypotheses will be tested:

(1) All secondary tasks will interfere with the primary driving task, resulting in more driving errors during dual-task trials.

(2) A main effect of complexity will be obtained, with more driving errors occurring in the high complexity conditions.

(3) Slower speeds have been associated with increased cognitive load (Jenness et al., 2002, Wickens and Seppelt, 2002), therefore, it is predicted that participants will drive more slowly while performing dual-task trials, and that a correlation will be found between speed and mental load.

(4) Results will follow the CEM predictive model (see Figure 1a). The effects of secondary task overlap with input-processing-output modalities will be mediated by central executive demands. Higher central executive processing demands of the secondary task will result in more driving errors. Thus, a controlled auditory-vocal task (verbal memory search) during driving should lead to greater interference with the primary driving task than an automated visual-manual task (eating), and similar interference as would be found with a controlled visual-manual task (changing CDs).

Though measures of the primary driving task are of ultimate interest, these measures alone may not fully elucidate the underlying cognitive load. Cognitive load is assessed by measuring mental load, mental effort, and performance; where mental load and mental effort are empirically derived using subjective rating scales, and performance data is obtained using dual-task techniques (Paas, Tuovinen, Tabbers, and Van Gerven, 2003). These methods have been found to be very sensitive with instructional research (Gopher and Braune, 1984; Paas, 1992; Paas, Van Merrienboer, and Adam, 1994; Paas et al., 2003). For example, Paas et al. (1994) used a 9-point symmetric category rating scale to measure perceived task difficulty (1-very, very low mental effort to 9-very, very high mental effort; F. Paas, personal communication, April 12, 2003). Participants were asked to use the scale to indicate the amount of mental effort they invested on a number of tasks. A physiological measure of heart rate was also recorded. Paas et al. (1994) found that participants were able to rate mental load easily (as did Gopher and Braune). Results suggest that the subjective workload measure was more sensitive and reliable than the heart rate measure of workload, which appeared to be sensitive only to large differences in invested mental effort. Additionally, subjective workload measures are sensitive to variations undetected by performance measures. Dual-task studies have shown that when time-sharing between tasks is well handled, equivalent levels of performance may arise which do not provide information on the cognitive costs involved in obtaining the performance (Gopher and Braune; Paas, 1992; Pass et al., 1994; 2003).

The current study adopted the subjective mental workload rating scale used by Paas et al. (1994). Participants were asked to indicate their mental effort using the scale after completing each driving trial, and then asked to re-evaluate their scores at the conclusion of the study. This served as a manipulation check to insure that tasks rated as more demanding in this study were actually experienced as more demanding by participants.

Method

Design

The current study utilized a 4-way factorial: 3 (secondary tasks) by 2 (complexity levels) by 2 (genders) by 2 (trials), with the first three variables varying between participants and the last variable varying within participants. Secondary task included three levels: an automatic visual-spatial-manual task (eating), a controlled visual-spatial-manual task (changing CDs), and a controlled auditory-verbal-vocal task (a memory search task). Complexity included two levels: high traffic and low traffic. The third variable, gender, was used as a blocking variable. Each participant performed a practice trial, a control trial to establish baseline driving (driving only), and one of the dual-task trials (driving while performing a secondary task), with both the baseline driving and dual-task trials sharing the same complexity level. The order in which the baseline driving and dual-task trials were presented was counterbalanced between participants.

To control for expectancy effects and guard the dependent measures against practice effects, two versions of the same driving scenario route were shown to each participant. The two versions differed only in the timing of signal lights and the placement of events to which the driver must respond. The order of version presentation and its coupling with control and dual-task trials was also counterbalanced across participants (see Figure 2).

The difference between baseline driving trial scores and dual-task trial scores for each participant yielded the dependent measures of interest. For each trial several dependent variables were measured. A composite dependent variable, driving errors, consisted of three commensurate measures: the number of collisions, road departures and speeding violations. These measures were recorded for each trial by the driving simulator. In addition, mean speed was recorded for each trial.

Subjective workload ratings (Paas et al., 1994; obtained from F. Paas, personal communication, April 12, 2003) for each task were devised as a manipulation check on mental demand ratings of the tasks. An exit questionnaire (see Appendix A) supplied a re-evaluation of subjective workload measures and identified covariates for analysis.

Three underlying manipulations were targeted in this study. The first involved the central executive processing demands of the secondary task. Central executive involvement cannot be directly measured, but can be inferred from prior cognitive research. Hegarty et al. (2000) equates central executive involvement with cognitive load, which Paas et al. (1994, 2003) asserts can be successfully measured by subjective ratings of mental demand. Schneider et al. (1984) has demonstrated that less demand is required for automatic processes than controlled processes. Based on the literature, the tasks used in the present study have been ranked as to their demand for central executive processing (see Table 1).

Two tasks (eating and changing CDs) considered primarily visual-spatial-manual were employed to test MRT (Wickens, 1984) against the proposed CEM model. The first proposes that both these tasks will exert similar amounts of interference with the primary driving task, in proportion with their reliance upon the visual and manual modalities. The second model predicts that interference with the eating task will be mediated (reduced in demand) because it is an automatic process and requires no central executive involvement (as predicted by Schneider et al., 1984; Baddeley, 1996, as cited by Wickens and Hollands, 1999). The memory search task (controlled auditory-verbal-vocal), according to MRT, should cause little interference with the primary driving task, whereas CEM would predict that this task will be mediated (increased in demand) due to its demands on the central executive since strategy selection for memory search and retrieval will be present (Baddeley, 1986; 1996, as cited by Wickens and Hollands).

The second manipulation targeted in this study involves the visual-input demand of the secondary tasks. This was measured by the number of glances away from the monitor during trial simulations. Trial simulations were filmed during the pilot study and glances away from the monitor were tallied. The results indicated that a successful visual manipulation was present in the current design, with CD-changing requiring the most visual demand, followed by eating. Visual demand for the memory search task was very low (see Table 1).

The third manipulation targeted in this study is the manual-output modality that is part of the visual-spatial-manual tasks. The CD-changing and eating tasks will both involve manual-output. Since participants will be instructed to drive normally, a number of variations in manual movements are possible. During normal driving, participants periodically remove their hands from the steering wheel to scratch or stretch; some drive with both hands on the steering wheel while others drive with only one. Therefore, manual manipulations for the secondary task were operationalized as the time spent performing manual manipulations directly related to the secondary task. Manual manipulations were timed using the film record taken during the pilot study. The results show that both eating and CD-changing tasks require a similar amount of time to complete manual manipulations (see Table 1). There are certainly other important aspects of manual manipulation, but these are beyond the scope of the current study.

Participants

Fifty-six undergraduate psychology students participated in partial fulfillment of course requirements. All participants had a valid driver’s license, and were able to eat 2 ounces of potato or corn chips and drink 17 ounces of water during the study.

Participants ranged in age from 18 to 27, with a mean age of 19.6 years (*SD* = 1.74). Additional demographic information was collected using an exit questionnaire (refer to Appendix A). Participants self-selected into the experiment through a volunteering system that allowed each participant to select a convenient day and time from several offered.

Each participant performed a practice trial, a control trial, and a dual-task trial. Half of the participants experienced both the control and dual-task trials with high traffic complexity and the other half experienced both trials with low traffic complexity. To control for gender differences, gender was held constant by assigning an equal number of males and females to each condition.

Conditions were assigned to time slots using restricted randomization. A die was rolled to determine condition and a coin was tossed twice, once to determine order and once to determine version, with the restriction that each configuration occur once before it was allowed to repeat. Male and female participants were assigned separately to enable each configuration to be presented to each gender once before it was repeated for that gender. Each participant was tested individually.

Apparatus

The STI driving simulator (Systems Technology, 2002) was used to present driving scenarios. The simulator was composed of a 27-inch computer monitor which sat atop a table, a Microsoft forced-feedback steering wheel clamped to the table in front of the computer screen, and two foot pedals situated on the floor under the steering wheel to simulate brake and accelerator pedals (see Figure 3). Participants sat in a chair in front of the simulator while they were presented with an interactive moving driving scene on the computer screen. Participants were asked to use the foot pedals and steering wheel to move through a programmed driving simulation as though they were driving. Driving errors were measured and recorded by the simulator.

During the eating condition, a universal adjustable cup holder purchased at an auto parts store was affixed to the simulator table with Velcro about a foot to the right of the steering wheel (see Figure 4). During the CD-changing condition, a portable CD player (Fisher Studio-Standard, model PHDS200) was affixed to the simulator table with Velcro directly to the right of the steering wheel. A cloth CD holder (Axius visor model) was affixed with Velcro just beneath the CD player, hanging securely over the edge of the table (see Figure 5). All CDs used in this study were generic re-writable CDs, each with the same white label.

The simulation scene presented on the monitor allowed a complete windshield view of the forward driving scene. Below the windshield, at the bottom of the screen, two analog dials represented the speedometer and tachometer. Toward the top of the screen and off center to the right, a rectangular box acted as a rear view mirror displaying the scene directly behind the vehicle (see Figure 6). Two levers located within easy reach on the back of the steering wheel, to the left and right, when depressed allowed the driver to see the scene 90-degrees to each side of the vehicle.

Stimulus

The base scenario. The base driving scenario was constructed to mimic driving environments close to campus, and presented events that participants could expect to encounter while driving to campus. The scenario began on a two-lane road in a residential neighborhood and traveled briefly through an industrial area and into a four-lane business district lined with office buildings. The scene included oncoming traffic, cross traffic and pedestrians at intersections, and pedestrians walking along the side of the road. In addition, 17 events that impinged upon the driver's path were distributed throughout the scenario, each with varying degrees of immediacy. Eleven of these events involved merging with other cars. Three required the driver to change lanes in order to pass stopped traffic, and the remaining eight involved other cars pulling into the participant’s lane with varying closure times between them (2-10 seconds). Six more events involved pedestrians crossing the road in unexpected areas (not at intersections or marked crossings) with closure time varying between 3 and 8 seconds.

Time to impact with regard to the 17 events was varied to avoid producing an expectancy effect. Eleven of the 17 events allowed over 5 seconds notice and were easily avoided if the participant was paying attention. Of the remaining six events, one allowed 2 seconds to impact, one allowed 3 seconds to impact, two allowed 4 seconds, and the remaining two allowed 5 seconds.

Since each participant experienced two trial scenarios (control and dual-task), two base simulations were created using the same route, in which only signal light timing and event placement were varied between them to avoid expectancy effects. The complete simulations each covered a total distance of 14,500 feet and took about 8 minutes to complete if the driver adhered to the speed limit.

Secondary task placement. Since drivers in the real world cannot know what will happen during a driving event, tasks in this study were prompted independently of events presented in the scenario. Eight blocks (one long task followed by two short tasks) were presented at equal distances throughout the driving scenario. The number of events was determined by first timing the entire run length, then timing the tasks. The long tasks took up to 20 seconds. Run time was divided by task time, yielding 24 separate 20-second time periods. Task instructions (each lasting 3.5 – 4 seconds) were then inserted into the simulation not by time, but by a measure of distance in feet. Task instructions began at 50 feet and a new instruction occurred every 600 feet thereafter. This insured that all tasks in all conditions occurred at the same location in the scenario. It also created a slight variation in the time lag between tasks within each participant's run and between participants. This had the desirable effect of prohibiting the formation of expectancy effects for tasks based upon time.

Automatic visual-spatial-manual: eating. To avoid potential food handling and storage problems, the food chosen for this study was a 1-2 ounce bag of potato or corn chips (each brand varied in actual weight). This was chosen since it does not require any handling, cooking or preparation, is a common food to be eaten while driving, and could be presented to participants sealed in its original packaging. Because thirst may accompany eating chips, a 16.9-ounce bottle of water was added. The water bottle was held in a constant position in the cup holder affixed to the simulator table, and the bag of chips was laid on the table between the steering wheel and the cup holder (see Figure 4).

The act of eating a chip – retrieving a chip from an open bag, placing it in one's mouth, chewing, and swallowing – is an automatic process that should require little mental effort and central executive involvement. It required only one or two glances away from the road and took about 15-20 seconds to complete (as observed in the pilot study). The act of taking a sip of water – retrieving the bottle from the cup holder, unscrewing the cap, taking a sip, screwing the cap back on, and placing the bottle back in the cup holder – is also considered an automatic process, requiring little central executive involvement. This required more glances away from the road (two or three glances as observed in the pilot study), and a little more time (about 20 seconds).

Task blocks were created which included taking one sip of water followed by eating two chips, one at a time (see Table 1). Consequently, each participant in the eating condition was asked to eat 16 chips and drink 8 sips of water during the study. Sound files were constructed and inserted into the scenario with a female voice instructing participants to "take a sip of water" or "eat a chip."

Controlled visual-spatial-manual: changing CDs. This task also encompassed two separate tasks arranged in eight blocks of one longer task followed by two shorter tasks to equate it with the eating tasks. The first task was inserting a new CD, which encompassed opening the CD player, taking out the old CD, placing it in the holder, removing a new CD from the holder, placing it in the player, and closing the player (see Figure 5). This task cannot be considered automatic and was regarded as requiring mental effort and central executive involvement. It took about 20 seconds to perform, and four or five glances away from the road (as observed in the pilot study).

The second CD task was song selection, which entailed pressing a button the correct number of times to select the song (e.g., two presses for song two, four presses for song four, etc.). The first song to be selected was always a smaller number than the second song, minimizing the cognitive load and enabling the participant to use only one button. This task was assessed to require moderate central executive involvement, took about 7-9 seconds to perform, and required two glances away from the road (as observed in the pilot study).

Sound files using the same female voice were constructed and inserted into the scenario instructing participants to “insert a new CD” or “select song [three].” Sound files were inserted into the scenario at the same places as the eating instructions (see Table 1). Thus, voice instructions for the tasks occurred at the same points in the scenario for both eating and CD-changing conditions, with eight blocks of one instance of the longer task (drinking or inserting a new CD) followed by two instances of the shorter task (eating a chip or selecting a song) occurring every 600 feet. All CD labels had the same appearance (white), and the CD volume remained off during the entire experiment to minimize cognitive load.

Controlled auditory-verbal-vocal: memory search. The idea was to select a task that would mimic important aspects of real-world cell phone conversations and allow experimental control, particularly with regard to central executive demand. Two tasks were selected to equate time expenditures and task switching with those of eating and CD-changing. Two verbal memory search tasks using words in a semantic context were employed. Both were prompted aurally. The first, the long task, was a category search task that required the participant to say aloud all examples of a given category. For example, sound files using the same female voice were created which stated, “Examples of [birds] are,” to which the participant responded by listing all the birds he/she could think of until the next voice prompt sounded. Eight categories (taken from Uyeda and Mandler, 1980) were included (see Table 2).

The second verbal search task involved antonyms and required the participant to say aloud an antonym of a given word. Sixteen antonyms were selected from a dictionary of synonyms and antonyms (see Table 2). Sound files using the same female voice were constructed in the form of “The opposite of [noise] is,” to which the participant responded by giving one answer. Switching from examples to antonyms and back again along with the instruction to continue listing examples with one and not the other, was devised to increase central executive processing (Schneider et al., 1984).

The placement of the voice instructions followed the same pattern as both the eating and CD-changing conditions (see Table 1). Thus, all three dual-task conditions offered voice instructions for the secondary tasks at the same locations in the scenario. During real-world phone conversations, the other party lacks awareness of the driving scene. Therefore, the placement of these memory search tasks independent of events (as was the case with eating and CD-changing) enhances both construct and external validity.

Varying complexity. Each of the three secondary task conditions – eating, CD-changing, and memory search – were split into two groups. Both groups experienced the same scenarios with the same events, but the number of opposing vehicles, cross traffic vehicles, parked vehicles, and pedestrians in the high complexity conditions were twice that of the low complexity conditions (see Table 3).

Practice scenario. A practice scenario was created to introduce participants to the simulator and allow learning effects to stabilize prior to performing the trial scenarios. The STI simulator’s brakes were highly responsive. Therefore, braking response time and distance were difficult to gauge without practice. The practice scenario presented 15 intersections, with 11 requiring a full stop. It also included a few road obstacles and events that required merging, lane changing, and interacting with other cars in a similar manner as in the trial simulations. Voice instructions were embedded in the practice scenario to insure all participants learned to maneuver at a necessary baseline level. The practice simulation included ticketing feedback for exceeding the posted speed limit by five miles per hour, and for failing to stop at intersections with red lights or stop signs. The ticketing feedback was included to help participants acclimate to the simulator. The practice scenario covered a total distance of 12,500 feet and took about 8 minutes to complete.

Feedback from participants indicated that the practice scenario was effective. All participants reported feeling comfortable with the brakes, steering, and the simulator in general at the conclusion of the practice trial. All participants reported having begun to drive normally toward the end of the practice scenario.

A note about scenario variability. The STI driving simulator bases events upon a measure of closure time (time to collision) between the participant's vehicle and the other vehicles, pedestrians, and timing of traffic signals, which is computed using the driver's current speed and trajectory at the time the event is programmed to begin, but does not take into account changes in the driver's behavior that may occur after the event has begun. Some events also use the participant's average speed to that point in the scenario in the calculation as well. Because of these factors, each participant experienced a slightly different scenario; pedestrians may have started walking a little early or late at signalized intersections, as the signals change independently of the activation of pedestrian movement. Care was taken to minimize these inconsistencies within a reasonable range of driving speeds and behaviors.

This unpredictability of events within a reasonable range has several benefits. Most real-world driving events lack precise predictability and strict adherence to laws. Other drivers commonly act inconsistently with our expectations. Pedestrians sometimes cross against the light or too late. Thus, the slight inconsistencies in the scenario between participants can aid both construct validity and external validity in the current study.

Materials

Exit questionnaire. The exit questionnaire served three basic purposes. It was used to identify covariates for analysis, it was used to insure equivalency among groups on measures likely to exert an effect upon outcome, and it allowed participants to re-evaluate their subjective workload ratings of baseline driving and dual-task driving trials. The first fourteen questions addressed the first two goals, and the last addressed the third. A 5-point Likert scale was used where appropriate (questions 6-13) and percentage anchors were used in addition to word anchors with several items (questions 8-12). The questionnaire was used and refined during the pilot study to eliminate ambiguous wording, and is shown in its final form in Appendix A.

Age (question 1). Age is an important measure since research has shown significant effects of age on driving performance. Bolstad (2001) found that older drivers had more difficulty attending to important information while driving when compared with younger drivers. Kennedy, Jentsch and Smither (2001) found that younger drivers were better at detecting distances and closure rates, and that this ability was independent of visual acuity. See Staplin et al. (1997) for a comprehensive meta-analysis of the effects of age upon driving.

Gender (question 2). The experimental design of the current study treated gender as a blocking variable. Gender was included on the questionnaire to anchor answers so that any non-equivalence on the various questionnaire items between genders could be readily identified.

Driving experience (questions 3 and 4). Question 3 provided a measure of driving experience associated with the length of time participants had been licensed to drive. Question 4 provided a measure of driving experience associated with the number of miles participants drive on a normal basis. As discussed earlier, experience can automate a task, reducing central executive involvement (Shiffrin and Schneider, 1984). Thus, experienced drivers may find the primary task of simulated driving less demanding than novice drivers. This could exert an effect upon the dependent measures (driving errors). In addition, other affects have been associated with driving experience, and these affects could have an indirect impact upon the current results. For example, studies have shown different visual-scan patterns associated with novice drivers and experienced drivers (Sanders and McCormick, 1993).

Experience with the apparatus (questions 5-7). Since the STI driving simulator presents very much like a computer or video game, ascertaining the level of experience with computer and video games could be an important measure of experience – perhaps influencing the level of central executive involvement of the primary driving task in the current study. Question 5 asked whether participants had any prior experience with driving simulators. Question 6 asked participants to rate their level of experience with computer or video games. Question 7 asked participants to rate their experience with driving games in particular.

Experience with secondary tasks (questions 8-11). These questions provided data regarding prior experience with the types of tasks that served as secondary tasks in the current study. This was important since automation with the secondary task could influence performance measures. Question 8 asked participants to rate how often they engaged in eating while driving, providing a measure of experience related with the eating condition presented in the current study. Question 9 asked participants to rate how often they engaged in cell phone use while driving, and question 11 asked how often participants drove with a passenger in the car. Both these questions were devised to provide a measure of experience commensurate with the memory search task in the current study. The memory search task involved listening (auditory input), comprehension and memory search (verbal and cognitive processing), and vocal response (vocal output). Talking on a cell phone or engaging in passenger conversation while driving would also employ these modalities; however, they include other modalities as well, such as the manual manipulation of the cell phone, and the visual components associated with both cell phone use and conversing with a passenger. Question 10 asked participants to rate how often they engage in changing CDs, tapes, and tuning the radio while driving. This provided a measure of experience commensurate with the CD-changing task employed in the current study. Though no auditory stimulus accompanied the task during the study (the volume remained off), the visual input, spatial processing, and manual manipulation of the tasks are very similar.

Comparisons of the STI simulator with actual driving (questions 12 and 13). These questions provided information regarding how participants perceived the simulated driving. This subjective measure can serve to elucidate how closely the present study replicated real-world driving, and ascertain the difficulty associated with simulated driving when compared with actual driving.

Physiological effects of eating (question 14). Because participants in the eating condition were asked to eat while performing the dual-task trial, the physiological effects of eating may produce confounding by influencing performance differentially between task groups. If physiological effects, such as blood sugar level, had an effect upon the dependent measures in the current study, then a correlation should be found between the number of hours prior to the study that participants last ate and their performance.

Re-evaluations of verbal subjective mental ratings (questions15a and 15b). After each experimental trial (baseline driving and dual-task) participants were shown the 9-point rating scale shown in Appendix A, and asked to rate the mental effort they used during the last trial. The questionnaire allowed another assessment at the end of the experimental session after both trials had been completed. This should increase validity of the subjective mental load ratings (Paas et al., 1994).

Procedure

Because some participants were asked to eat while they completed the simulated driving scenario, to avoid any food allergies or adverse reactions, sign-up sheets for this study clearly stated that participants "must be able to eat 2 ounces of corn or potato chips and drink a 17-ounce bottle of water during the study." To guard against differential motivation between groups asked to eat and those not asked to eat, all participants not in the eating condition were informed that they would receive a bag of chips at the conclusion of the study.

Before the experiment began, each participant was asked to sign an informed consent form. Following this, each participant was shown how to use the simulator and received a practice trial. Participants were then asked to complete two trials, one baseline driving trial and one dual-task trial. Before presenting dual-task trials, participants were introduced to the secondary task required in that trial.

Before the eating trial, participants were allowed to choose from a variety of packaged corn and potato chips supplied by the researcher. The researcher attached the cup holder to the table (see Figure 4), described and demonstrated the steps involved with the two tasks to be performed – drinking a sip of water and eating a chip. The participant was asked to open their bag of chips and lay it on the table between the steering wheel and the cup holder, and to break the seal on the bottle of water and position it in the cup holder. The participant was asked to demonstrate understanding of the steps involved in the secondary tasks before being asked to complete a dual-task trial that prompted them to eat chips and drink water while driving.

Prior to the CD-changing trial, the researcher affixed the CD holder and CD player to the table (see Figure 5) and described and demonstrated the steps involved in the CD insertion task and the song selection task. The participant was asked to demonstrate the secondary tasks before completing a dual-task trial that prompted them to change CDs and select songs while driving. Volume level was set to "0" and each participant was advised that the volume would remain off during the study.

Before the memory search trial, the researcher explained each memory search task and provided an example. The participant was asked to demonstrate their understanding of the tasks through a verbal example with the researcher before performing a dual-task trial that prompted the memory tasks while they drove.

The apparatus from the prior trial was removed prior to beginning a new trial, and only apparatus used in the current driving trial was visible to the participant. Before each trial, participants were instructed to obey posted speed limits and normal traffic laws, not to turn onto any street, and to drive as they normally would. They were informed that no ticketing feedback would be given during the experimental trials. At the end of each trial, participants were asked to rate the mental effort they experienced during the preceding trial. The 9-point scale (taken from Paas et al., 1994) was presented visually to the participants and their ratings were recorded by the researcher. After all trials had been completed, each participant was asked to complete an exit questionnaire (see Appendix A). Participants were then told the purpose of the study, and thanked for their participation. Each session lasted approximately 45 minutes.

Results

Scoring and Data Preparation

The STI simulator collected data on each driving trial. In addition, film of the trial scenarios was reviewed during the course of the experiment to verify several areas. Compliance with instructions regarding both the primary driving task and the secondary task were verified. Since the scenarios were expected to present differently with each participant, the experimenter visually verified collisions and road departures for each trial. A few anomalies occurred which required manual adjustments of simulator scores.

Collisions were separated into three categories: program errors, stopped collisions, and true collisions. Collisions classified as program errors included any collision that, due to the interaction of the simulated scenario with the participant, offered an unrealistic situation with no visible warning of collision. For example, an event vehicle may be programmed to slow considerably and then exit the road when passed by the participant's vehicle. However, if the participant does not pass the slow vehicle, but instead remains behind it, the vehicle will not exit the road. Instead, it will continue on the road and enter another programmed event, causing overlap. Since vehicles other than the participant's vehicle do not recognize each other, they may pass through each other. This can result in the lead vehicle passing through a stopped car in the lane, virtually leading the participant into a collision with the stopped car. This is clearly not realistic, as it would not afford the participant any cue or warning. Collisions rated as program errors were deducted.

Stopped collisions entailed the participant stopping behind a stopped car too close to pass it. The function to reverse the direction of the vehicle was not available during the experiment, so the only manner in which to continue through the scenario entailed colliding with the stopped car, often at the instruction of the experimenter. Collisions rated as stopped collisions were deducted from the total collision score. All other collisions were rated as true collisions and were retained in the total collision score.

Road departures, as recorded by the STI simulator, included both entering the right shoulder and crossing the center divider into opposing traffic lanes. The simulator did not differentiate between legal and illegal crossings of the center divider. However, crossing the center divider to pass a car is legal when the centerline is not solid. Therefore, any deliberate crossing of the center divider for the purpose of passing another vehicle or avoiding a pedestrian when the center divider was a broken line was deducted from the road departure score.

Data from eight participants were discarded. Data from two males were discarded due to instrumentation errors – the foot pedals were not securely attached to the floor with one, and the sound equipment malfunctioned during task prompts with the other.

Data from one female and one male were discarded for failing to comply with instructions – the first failed to eat when prompted, and the second, a clear outlier, did not treat the driving task seriously and crashed into every object he could manage during the trial scenarios. Data from another outlier, a female, was discarded because she was very ill during the experiment. Her scores showed a steep deterioration over time that could well have been a function of her physiological state rather than experimental tasks. Five replacement participants were obtained to substitute this data. In addition, the last three female participants' data were discarded because their data was not needed and their inclusion would have thrown off the gender blocking and complete balance of the study.

Change scores were computed for each of the 48 retained participants as the difference between the dual-task trial and the control trial (baseline driving) for each dependent variable (number of collisions, number of speed violations, number of road departures, and mean speed). Mean speed violated normality with positive skew, and was subjected to a log transformation before continuing with analysis. A composite dependent variable, "number of driving errors," was computed as the total number of collisions, speed violations, and road departures. A change score on this composite measure was determined for each participant.

Raw Score Comparisons of Dual-task and Baseline Driving Performance

Paired t-tests were conducted to explore whether each of the secondary tasks increased driving errors significantly above baseline driving (hypothesis 1). Tests were performed on the raw scores obtained from baseline driving and during the dual-task condition on both the composite measure "driving errors" (a sum of the number of collisions, speed violations and road departures) and the individual measures: mean speed, number of collisions, number of speed violations, and number of road departures. Scores were collapsed across both gender and complexity. Alpha was set at .05 for all comparisons.

Table 4 shows the means and t-values for all paired t-tests. With CD-changing, the number of road departures increased significantly over baseline, *t*(15) = 2.42, *p* = .029, change *M* = .69, *SD* = 1.14. While changing CDs, participants departed the road more (*M* = 1.13, *SD* = 1.31) than they did during baseline driving (*M* = .44, *SD* = .89). The number of driving errors, collisions, speed violations, and mean speed during the CD-changing trial did not significantly differ from those found with baseline driving.

With eating, the number of driving errors increased over baseline driving, *t*(15) = 2.19, *p* = .045, change *M* = 1.81, *SD* = 3.31. Participants made more driving errors while eating (*M* = 2.94, *SD* = 2.74) than during baseline driving (*M* = 1.13, *SD* = .96). The number of road departures also increased during the eating trial, *t*(15) = 2.64, *p* = .019, change *M* = 1.06, *SD* = 1.61. Participants departed the road more often while eating (*M* = 1.50, *SD* = 1.46) than while driving without a secondary task (*M* = .44, *SD* = .63). The number of collisions, speed violations, and the mean speed while eating and driving did not differ significantly from those obtained with baseline driving.

With the memory task, no significant differences were found between driving while performing the memory task and baseline driving on any of the measures.

Group Equivalency and Questionnaire Results

To test equivalency between groups, a 3 (task) by 2 (complexity) by 2 (gender) between-subjects ANOVA was conducted for each item on the exit questionnaire (shown in Appendix A). Alpha was set at .05 for all tests. Means and standard deviations for all groups are shown in Table 5.

Question 1: Age. No significant age differences were found between groups. The mean age of participants was 19.6 years, *SD* = 1.74 years. Participants ranged in age from 18 to 27 years.

Question 3: How long have you been licensed to drive a motor vehicle (years)? No significant differences were found between groups on the number of years they reported having been licensed to drive. Participants reported having been licensed an average of 2.9 years, *SD* = 1.82 years, with a range of 1 month to 11 years.

Question 4: How much do you drive (miles per week)? Significant differences were found between groups. A main effect of gender was obtained, $F(11, 36) = 1.64$, $p = .129$. Males reported driving longer distances per week ($M = 292.7$ miles, $SD = 254.96$ miles) than females ($M = 159.8$ miles, $SD = 120.38$). Additionally, a significant interaction of complexity by gender was obtained, $F(1,36) = 7.15$, $p = .011$. As shown in Figure 7, both genders in the low complexity conditions reported driving similar distances per week (males $M = 200.0$, $SD = 218.87$, females $M = 217.1$, $SD = 116.63$); however, males in the high complexity conditions reported driving more ($M = 385.4$, $SD = 263.17$), while females in the high complexity conditions reported driving less ($M = 102.5$, $SD = 97.64$). The miles driven per week for males and females in the high complexity conditions were significantly different from each other using the Tukey method, $HSD = 283$, $p < .01$; however, no significance was found between genders in low complexity conditions or between complexity levels within genders. Main effects of task and complexity were not found, nor were any other interactions.

Question 5: Have you ever used a driving simulator before? No significant differences were found between groups on this question. All participants reported having no prior experience with a driving simulator.

Question 6: How often do you play computer or video games? The five options were transformed into a Likert scale for scoring: 1 – Never, 2 – Have played a few times, 3 – Sometimes, 4 – More than 10 hours per month, 5 – More than 5 hours per week. Groups were found to differ significantly on this item. A main effect of gender was obtained, $F(1,36) = 37.12$, $p < .001$. Males reported playing computer or video games ($M = 3.63$, $SD = .970$) more often than females ($M = 2.17$, $SD = .817$). Thus, males were more likely to report that they played more than 10 hours per month, while females were more likely to report that they had only played a few times. No significant main effects of task or complexity were obtained and no significant interactions were found.

Question 7: How often do you play computer or video games that involve driving? The five options were transformed into a Likert scale in the same manner as question 6. Groups were found to differ significantly on this item as well. A main effect of gender was found, $F(1,36) = 14.22$, $p = .001$. Males reported playing computer or video games that involved driving more often than females ($M = 2.54$, $SD = .658$; $M = 1.88$, $SD = .680$, respectively). On average, males reported having played driving games more than a few times while females reported having played less.

A main effect of task was also obtained, $F(2,36) = 5.56$, $p = .008$. Participants in the eating group reported having slightly more experience with driving games ($M = 2.6$, $SE = .50$) than did participants in the CD-changing and memory task groups ($M = 2.0$, SD

= .82; *M* = 2.0, *SD* = .73, respectively). A main effect of complexity was not obtained, nor were any interactive effects.

Question 8: How often do you eat while driving? The five options were transformed into a Likert scale for scoring: 1 – 0% Never, 2 – 25% Not often, 3 – 50% Sometimes, 4 – 75% Often, 5 – 100% Always. Groups were not equivalent. A main effect of gender was obtained, *F*(1, 36) = 8.27, *p* = .007. Females reported eating more often while driving than males (females *M* = 2.88, *SD* = 1.076; males *M* = 2.08, *SD* = .717). Thus, on average females reported eating while driving "50% sometimes," while males reported that they ate while driving "25% not often."

Question 9: How often do you talk on a cell phone while driving? The five options were transformed into a Likert scale in the same manner as question 8. Groups were found to be equivalent with regard to this item. Participant reports of talking on a cell phone while driving ranged from "0% never" to "100% always," and were found to be fairly normally distributed with an average answer of "50% sometimes" (*M* = 2.8, *SD* = 1.08).

Question 10: How often do you change CDs, tapes, or the radio station while driving? As in the two preceding questions, the same Likert scale was applied to the five options. Groups were found to be equivalent with regard to this question. Participants reported engaging in these activities, on average, "75% often" (*M* = 3.9, *SD* = .94), with a range in answers from "0% never" to "100% always."

Question 11: How often do you drive with a passenger in the car? Using the same Likert score, groups were found to be equivalent on this item as well. On average, participants reported driving with a passenger in the car a little more than "50%

sometimes" (*M* = 3.4, *SD* = .86). Again, responses ranged from "0% never" to "100% always."

Question 12: How would you compare the simulation you just completed with actual driving? The five options were transformed into a Likert scale as follows: 1 – 0% Not like driving at all, 2 – 25% A little like driving, 3 – 50% Similar to driving, 4 – 75% Very similar to driving, 5 – 100% Exactly like driving. A main effect of task was obtained, *F*(2,36) = 3.25, *p* = .050. Participants in the CD-changing group rated the simulations as more similar to actual driving (*M* = 3.2, *SD* = .83) than did those in the eating group (*M* = 3.0, *SD* = .97), and those in the memory task group rated the simulations as the least similar to actual driving (*M* = 2.4, *SD* = .81). Post-hoc Tukey tests revealed a significant difference between the CD-changing group and the memory task group, HSD = .75, *p* = .049. No other main effects or interactions were obtained.

Question 13: Compared with actual driving, the simulated driving was . . . ? The five options were transformed into a Likert scale as follows: 1 – Much harder than actual driving, 2 – A little harder than actual driving, 3 – About the same as actual driving, 4 – A little easier than actual driving, 5 – Much easier than actual driving. A main effect of gender was obtained, *F*(1,36) = 4.12, *p* = .050. Females rated the simulations as more difficult when compared with actual driving (*M* = 2.2, *SD* = .87) than males (*M* = 2.7, *SD* = .96). No main effects of task or complexity were found to be significant, nor were any interactions.

Question 14: When was the last time you ate? How may hours ago? Non-equivalence was found between groups on this self-report measure. A main effect of gender was found, *F*(1, 36) = 5.160, *p* = .029. Males reported having gone without eating twice as long before the experiment (*M* = 6.5 hours, *SD* = 5.63) than females (*M* = 3.4, *SD* = 3.05). Neither a main effect of task nor a main effect of complexity was found, nor were any interactions.

Spearman correlations were conducted on the composite measure "driving errors" for baseline driving to ascertain whether the number of hours participants ate prior to engaging in the study had any effect upon their baseline performance. The number of hours was set to "0" for those participants in the eating condition who performed the baseline driving trial second (after eating during the experiment). A significant correlation was not obtained, $r_s = -.092$, $p = .533$.

To further explore whether eating during the study had an immediate effect upon performance, an independent t-test was conducted comparing the driving errors of participants who ate before performing the baseline driving trial (performed dual-task trial first) and those who didn't (performed dual-task trial second). The comparison involved baseline driving scores only. Alpha was set at .05. No significant difference was found between the baseline performance of those who ate (*M* = .75, *SE* = .313) and those who didn't (*M* = 1.50, *SE* = .327); $t(14) = -1.66$, $p = .12$.

To rule out the possibility that this result was attributable to order effects, independent t-tests were conducted in a similar manner for each of the other tasks. Alpha was set at .05. No significant differences were obtained as a function of the order in which baseline driving was performed with either the CD-changing group, $t(14) = 1.08$, $p = .295$; or the memory search group, $t(14) = .33$, $p = .746$.

Question 15a: Rate how much mental effort you used [during baseline driving trials]. Participants rated how much mental effort they used during baseline driving twice during the experiment: upon completion of the baseline driving trial, and on the exit questionnaire (shown in Appendix A). Of the 48 participants, only nine altered their ratings of baseline driving. Seven participants reduced their scores, while two increased their scores. For each of the nine participants who changed their ratings, the average of the two scores was used for analysis.

Groups were not equivalent with regard to their ratings of the mental effort used during baseline driving. A main effect of gender was obtained, *F*(1, 36) = 5.09, *p* = .030. Overall, females rated baseline driving as more effortful (*M* = 5.3, *SD* = 1.16) than males (*M* = 4.3, *SD* = 1.87). Thus, on average, females rated the task of baseline driving on the simulator as requiring neither low nor high mental effort, while males rated it as requiring rather low mental effort. No significant main effects of task or complexity were obtained, nor were any interactions.

Question 15b: Rate how much mental effort you used [during dual-task trials]. Participants rated how much mental effort they used during the dual-task trial twice during the experiment: upon completion of the dual-task trial, and on the exit questionnaire (see Appendix A). Of the 48 participants, only six altered their ratings. Two participants in the CD-changing condition reduced their scores, while three in the eating condition and one in the memory condition increased their scores. For each of the six participants who changed their ratings, the average of the two scores was used for analysis.

Again, groups were found to differ. A main effect of gender was obtained, *F*(1, 36) = 7.14, *p* = .011. Females rated driving while performing a secondary task as more effortful (*M* = 6.8, *SD* = 1.38) than males (*M* = 5.6, *SD* = 1.79). On average, females rated the dual-task trial as requiring high mental effort, while males rated these trials as requiring some mental effort. No significant main effects of task or complexity were obtained.

Non-equivalence between groups was obtained most often in the form of gender differences (as in questions 4, 6, 7, 8, 13, 14, 15a, 15b). This was handled by treating gender as a factor in all further analyses. Significant differences between task groups were found with questions 7, and 12. Question 12 revealed that participants in the CD-changing group rated the simulator as more similar to actual driving than did those in the eating group, with those in the memory task group rating the simulator as the least similar to actual driving. This may indicate that the groups were not equivalent to begin with, or it may indicate that the tasks exerted an effect upon the perception of similarity between the simulator and driving.

Question 7 dealt with driving game experience. Participants in the eating group reported having slightly more experience with driving games than in the CD-changing and memory task groups. This could pose a problem if driving game experience is highly correlated with the dependent measures used in this study.

Correlations

Spearman correlations were conducted on all dependent measures, gender, and questionnaire items. Three factors emerged as possible covariates: game experience (question 6), driving game experience (question 7), and number of years licensed to drive (question 3). These correlations are shown in Table 6.

Driving game experience (question 7) was correlated with only one dependent measure: road departure change scores, $r_s = -.325$ $p = .024$. However, significant correlations between driving game experience and two questionnaire items emerged: with gender, $r_s = -.466$, $p = .001$, and with computer/video game experience, $r_s = .572$, $p < .001$. As Table 6 shows, game experience was significantly correlated with the number of driving errors, road departures, speed violations, gender, and driving game experience. Because game experience was highly correlated with driving game experience, and was correlated with more dependent measures than driving game experience, it was selected as a possible covariate.

The number of years licensed to drive (question 3) was correlated with three dependent measures: number of driving errors, number of road departures, and mean speed. Since these correlations did not completely overlap with those found with computer/video game experience, and the two covariates were not correlated with each other, both were selected as the strongest possible covariates to use in further analyses.

The number of miles driven per week (question 4) did not emerge as a covariate. Though groups were not equivalent due to a significant gender effect and gender by complexity interaction on this questionnaire item, it could not be used as a covariate since none of the dependent measures used in this study varied with it. However, it was found to correlate significantly with the independent variable gender ($r_s = -.342$, $p = .017$). Thus, interpretations of the effects of gender may have been confounded with the number of miles driven per week, and complexity could be involved due to its interaction with gender on this measure.

Analyses with Change Scores on the Composite Measure "Driving Errors"

The discriminating feature between MRT and CEM predictions lays in the relative position of the eating and memory conditions with regard to overall driving errors (see Figures 1a and 1b). To test this, the composite dependent variable, consisting of change scores on the commensurate measures (number of collisions, number of speed violations, and number of road departures) was analyzed in a 3 (task) by 2 (complexity) by 2 (gender) between-subjects ANOVA, with alpha set at .05.

The ANOVA revealed a significant main effect of gender, *F*(1,36) = 5.86, *p* = .021 (see Table 7). The mean number of driving errors increased between baseline driving and driving while performing a secondary task with females (*M* = 2.00, *SD* = 3.23), while no mean difference between baseline and dual-task driving with males, (*M* = .00, *SD* = 2.41). No other main effects or interactions were obtained.

However, both the game experience covariate and the years licensed to drive covariate were significantly correlated so an ANCOVA with both covariates was also performed with alpha set at .05 (shown in Table 8). With the inclusion of the game experience covariate, *F*(1,34) = 8.29, *p* = .007, and the years licensed to drive covariate, *F*(1,34) = 6.42, *p* = .016, the gender effect disappeared, *F*(1,34) = .034, *p* = .854 (see Table 8). No significant main effects were obtained.

Instead, the ANCOVA revealed a significant complexity by task by gender interaction, *F*(2,34) = 6.77, *p* = .003. Complexity affected male and female performance on the three tasks quite differently. As shown in Figure 8a, with low complexity traffic, the number of driving errors decreased while changing CDs relative to baseline driving for females (*M* = -1.956, *SE* = 1.511), and increased for males (*M* = 3.480, *SE* = 1.535). The eating task resulted in the highest errors for both genders (males *M* = 2.651, *SE* = 1.389; females *M* = 3.154, *SE* = 1.282). The memory task yielded a similar an increase in errors for females (*M* = 1.633, *SE* = 1.456), while it resulted in fewer driving errors than baseline driving for males (*M* = -.779, *SE* = 1.308). With high complexity (see Figure 8b), a different pattern emerged. Male errors were reduced with the addition of the CD (*M* = -.589, *SE* = 1.305) and eating tasks (*M* = -.858, *SE* = 1.272), and increased with the memory task (*M* = 2.684, *SE* = 1.395), while female performance displayed the opposite pattern (*M* = 1.492, *SE* = 1.284; *M* = 2.242, *SE* = 1.284; *M* = -1.154, *SE* = 1.357, respectively). A list of the means in rank order appears in Table 9. Figures 9a and 9b allow a visual inspection of the different trends found between genders. The Tukey method was used to locate significant differences; however, none were found with this conservative method. The greatest difference occurred between males and females who changed CDs under low complexity traffic conditions, though this difference was not significant.

Analyses with Change Scores on the Separate Measures

To investigate the dependent measures separately, change scores by participant for each measure were analyzed using a 3 (task) by 2 (complexity) by 2 (gender) between-subjects MANOVA with alpha set at .05 (shown in Table 10). No significant covariates emerged in the multiple measure analysis.

Number of collisions. A main effect of gender on the number of collisions was obtained, *F*(1, 36) = 5.00, *p* = .032. Females had more collisions while performing a secondary task than they did during baseline driving (*M* = .417, *SE* = .198), while males had fewer collisions while performing a secondary task than they did during baseline driving (*M* = -.208, *SE* = .198).

A main effect of task was not obtained. The change in number of collisions from baseline driving was slightly greater with the eating task (*M* = .312, *SE* = .242), followed by the memory task (*M* = .125, *SE* = .242), and CD-changing (*M* = -.125, *SE* = .242), though these were not statistically different from each other.

A main effect of complexity was not obtained. The number of collisions increased slightly above baseline driving in the low complexity dual-task condition (*M* = .292, *SE* = .198), but did not increase in the high complexity dual-task condition (*M* = .083, *SE* = .198). Again, these differences were not significant.

A task by gender interaction on the number of collisions was obtained, *F*(2, 36) = 3.80, *p* = .032 (see Figure 10). With males, the number of collisions decreased relative to baseline driving with the eating task (*M* = -.500, *SE* = .342), and the memory task (*M* = -.125, *SE* = .342), while the CD-changing task had no mean effect on the number of collisions (*M* = .000, *SE* = .342). With females, the number of collisions increased with the eating task (*M* = 1.125, *SE* = .342), and the memory task (*M* = .375, *SE* .342), and decreased with the CD-changing task relative to baseline driving (*M* = -.250, *SE* = .342). Tukey comparisons were used but did not reveal any significant differences. Power may have been insufficient to fully elucidate the interaction.

Number of road departures. A marginal main effect of gender was found on the change in the number of road departures between dual-task and baseline driving, *F*(1,36) = 4.05, *p* = .052. Females showed a larger increase (*M* = 1.000, *SE* = .264) than males (*M* = .250, *SE* = .264).

A main effect of task was not obtained. Participants who performed the eating task departed the road slightly more often (*M* = 1.063, *SE* = .323) than did those who changed CDs (*M* = .687, *SE* = .323) or those who performed the memory task (*M* = .125, *SE* = .323), though these differences were not significant at alpha .05.

A main effect of complexity was also not obtained. The mean difference in number of road departures while performing a secondary task slightly increased when presented with low complexity traffic (*M* = .958, *SE* = .264), and increased less when presented with high complexity traffic (*M* = .292, *SE* = .264). Again, these differences were insignificant. No interactions were found with road departure change scores.

Number of speed violations. No main effects or interactions were obtained to demonstrate a change in the number of speed violations between dual-task and baseline driving. With gender, females showed a slight increase (*M* = .583, *SE* = .366), while males barely showed an increase (*M* = .042, *SE* = .366). With task, both eating and the memory task showed a slight increase (both *M* = .437, *SE* = .448), while CD-changing barely increased at all (*M* = .063, *SE* = .448). With complexity, high complexity traffic increased the number of speed violations over baseline slightly more than low complexity traffic (*M* = .375, *SE* = .366, *M* = .167, *SE* = .366, respectively). None of these differences were statistically significant at the .05 level.

Mean speed. No main effects or interactions were found suggesting a change in mean speed in miles per hour (mph) between dual-task trials and baseline driving. Though not statistically significant, participants actually tended to increase speed while performing the memory task (change *M* = 1.19, *SE* = .88), increase speed while eating (*M* = .10, *SE* = .50), and decrease speed while changing CDs (*M* = -.29, *SE* = .57). No differences were obtained between complexity levels (low *M* = .470 mph, *SE* = .547; high *M* = .213 mph, *SE* = .547), or between genders (males *M* = -.422 mph, *SE* = .547; females *M* = 1.105 mph, *SE* = .547).

Mental Effort

To ascertain whether mental effort ratings increased with the addition of a secondary task, paired t-tests were conducted on the mental ratings of the dual-task trials and the ratings of the baseline driving trials for each task condition. The change in rating of mental load increased significantly with the addition of all three tasks: memory task $t(15) = 4.516$, $p < .001$; eating task $t(15) = 4.111$, $p = .001$; and CD-changing task $t(15) = 2.809$, $p = .013$. Table 11 lists the means and standard deviations for each task and trial, as well as t and p values. On average, all three conditions rated baseline driving as requiring neither high nor low mental effort, and the dual-task trial as requiring rather high mental effort.

To explore mental effort ratings further, change scores were computed as the difference between each participant's rating of the dual-task trial and baseline driving trial (question 15b – 15a). These change scores were analyzed using a 3 (task) by 2 (complexity) by 2 (gender) ANOVA (no covariates were significant) with alpha set at .05. No significant main effects were revealed. However, a significant task by complexity interaction was obtained, $F(2, 36) = 3.52$, $p = .040$. As Figure 11 shows, participants in low complexity conditions rated eating while driving as the most different from baseline driving ($M = 2.3$, $SD = .92$), followed closely by CD-changing ($M = 2.1$, $SD = .88$), and rated driving while performing the memory task as the most similar to baseline driving ($M = .6$, $SD = 1.30$). In high complexity conditions, the opposite pattern emerged. Participants rated the driving while performing the memory task as the most different from baseline driving ($M = 1.8$, $SD = 1.20$), followed by changing CDs while driving ($M = 1.2$, $SD = 2.15$), and rated eating while driving as the most similar to baseline driving (*M* = .8, *SD* = 2.36). The Tukey method was employed to locate significant pairwise differences; however, none were significant using this conservative method.

Discussion

Hypotheses

Hypothesis 1: All dual-task trials will result in more driving errors than baseline driving. This hypothesis was only partially supported. Paired t-tests using the composite measure of driving errors revealed an increase over baseline driving only while eating. However, paired t-tests conducted with the separate dependent measures revealed significant differences between two of the tasks and baseline driving on the number of road departures. Road departures increased with CD-changing and eating. However, with both the composite measure and the separate measures, paired t-tests did not reveal any significant differences between the memory task trials and baseline driving.

It should be noted that the paired t-tests conducted to test this hypothesis collapsed both gender and complexity effects. Significant gender effects and interactive effects with both gender and complexity were obtained in later analyses (ANCOVA and MANOVA). Collapsing on gender and complexity could have reduced power sufficiently to conceal lesser effects.

Hypothesis 2: A main effect of complexity will result with more errors in high complexity. This hypothesis was not supported. The ANOVA and ANCOVA performed on change scores of the composite measure “driving errors” revealed no significant main effect of complexity. The ANCOVA did reveal an unexpected interaction of complexity with both task and gender. Figures 8a and 8b illustrate this interaction. The highest increase in driving errors over baseline driving was found with males in the low complexity CD condition. The smallest difference between baseline and dual-task was found with males in the high complexity CD condition, and the largest decrease in errors was found with females in the low complexity CD condition (refer to Table 9). The differences between these conditions were not statistically significant using the Tukey method; however, this may have been due to low power and a small sample size. In addition, the MANOVA using the separate dependent measures revealed no significant main effect of complexity, nor any interactions with complexity.

The ability to detect a main effect of complexity may have been impeded by the nonequivalence between groups on the number of miles driven per week (questionnaire item 4). Since this measure was significantly correlated with gender, and gender interacted with complexity, the ability to detect an effect of complexity may have been confounded as well. Females in high complexity conditions reported driving 283 miles less per week on average than males in high complexity conditions (see Figure 7). This difference was significant between males and females overall and between males and females in high complexity conditions. Therefore, high complexity scores may have been biased by an interaction of males with more experience, or by females with less experience, or both. However, it could be argued that since the number of miles driven per week did not significantly covary with any of the dependent measures, it's power to explain the absence of a main effect of complexity is weak at best. Further, the presence of a significant 3-way interaction indicates a reliance on both task and gender with regard to complexity. However, the current analysis cannot claim to distinguish gender from miles driven per week, nor measure the magnitude of its effect upon complexity. It would be prudent to assume that this measure could have had an effect since it is a measure of driving experience. The only other measure of driving experience in this study, the number of years licensed to drive, was correlated with many dependent measures and was used as a covariate in analyses.

Hypothesis 3: Mean speed will be lower during dual-task trials than during baseline driving and slower speed will be correlated with mental effort. This hypothesis was not supported. Paired t-tests revealed no significant differences in mean driving speed between each of the dual-task conditions and their respective baseline driving trials. This prohibited the finding of any relationship between mental load and mean speed for each task. Further, an exploratory correlation matrix found no correlation between the subjective mental effort ratings and any of the questionnaire items or dependent measures used in the current study. This may indicate a failure of this study to properly manipulate mental effort, or it may indicate a failure of the subjective rating scale to capture the mental effort of the tasks used in this study. This will be discussed in detail later.

Hypothesis 4: Results will follow the CEM predictions shown in Figure 1a, and mental effort will increase with errors. This hypothesis was not supported at the level predicted; however, some interesting implications arose from comparison of the current data with the predictive models. The CEM predictive model shown in Figure 1a was based upon the composite measure “driving errors” and did not take gender differences into account, since no significant gender differences were anticipated. For comparison purposes, Figure 12 shows the data from the current study collapsed over gender in the same manner as the predictive model. At first inspection, it appears that the data collected from the current study does not support the CEM model, nor does it support the MRT model shown in Figure 1b. However, eating and memory scores were not statistically different which would support the CEM model over the MRT model. Both of these were not found to significantly differ from baseline driving. The relative placement of the eating condition, the only task found to significantly differ from baseline driving, appears to differ only when presented with low complexity traffic.

While allowing direct comparison with the predictive models, viewing the results collapsed over gender is not appropriate since a significant three-way interaction (gender by complexity by task) was obtained (see ANCOVA, Table 8). Taking this into account, Figures 9a and 9b allow comparison of the predictive models separately with each gender at each level of complexity. With males (Figure 9a), low complexity scores resemble MRT predictions, with the visual-spatial-manual tasks (CD-changing and eating) resulting in more driving errors than baseline driving, and the auditory-verbal-vocal task (memory task) resulting in fewer driving errors than baseline driving. However, with high complexity, the opposite pattern emerged with males. The memory task resulted in an increase in driving errors over baseline driving, while the CD-changing and eating tasks resulted in fewer errors than baseline driving. Though this pattern does not match the CEM predictive model, it poses a problem for MRT, and suggests that with a more complex driving task, central executive involvement may exert a detrimental effect upon the primary driving task with males.

With females, a very different pattern emerged (see Figure 9b). High complexity scores approximate the MRT predictive model, with the visual-spatial-manual tasks (eating and CD-changing) resulting in an increase in driving errors relative to baseline driving, and the auditory-verbal-vocal task (memory task) resulting in a decrease in driving errors. Low complexity results did not match either predictive model, though the relative positions of the memory and CD tasks present a problem for MRT. Since CD-changing overlaps input-processing-output modalities with driving, MRT predicts that it will interfere with driving, resulting in increased driving errors relative to baseline driving. However, the CD task actually reduced the number of driving errors for females in the low complexity condition. MRT also predicts that the memory task will result in fewer errors than CD-changing since it does not share input-processing-output modalities with driving. Yet in the current study, the memory task resulted in increased driving errors over baseline with females in low complexity traffic. This could suggest the CEM model if it weren't for the relative position of the eating condition, which would be expected to yield the fewest errors since the activity of eating is thought to be automated. Instead eating resulted in the most errors. Further, eating exerted the most errors in both high and low complexity for females. This could suggest that some aspect of eating is inherently more difficult for females while operating the simulator, regardless of complexity level. It could also suggest that some aspect related to the experience of the females involved in this study interacted with their performance while eating.

Though gender differences were identified on the questionnaire – most notably, females reported less computer/video game (question 6) and driving game experience (question 7) than males – none of the questionnaire data logically explains why the eating condition in particular resulted in more errors for females than males. Further, females reported eating while driving on a more frequent basis than did males (question 8). The gender by complexity interaction found with question 4 (miles driven per week) showed that females in high complexity conditions reported driving 283 miles less per week than males in the high complexity conditions. Both males and females in the low complexity conditions reported driving the same average number of miles per week (200 and 217, respectively). Since the number of miles driven per week was significantly correlated with gender in this study, it is important to recognize that comparisons between males and females in high complexity are confounded by this measure.

Mental Effort Ratings

Rating scale techniques are based on the assumption that people are able to introspect on their cognitive processes and report the amount of mental effort expended. Though self-ratings may appear questionable, it has been demonstrated that people are quite capable of giving a numerical indication of their perceived mental burden (Gopher and Braune, 1984). Most subjective measures are multidimensional in that they assess groups of associated variables, such a mental effort, fatigue, and frustration, which are highly correlated (Nygren, 1991). However, studies have shown that reliable measures can be obtained with unidimensional scales, like the one used in this study (Paas, 1992; Paas et al., 1994; 2003). Further, it has been demonstrated that such scales are sensitive to relatively small differences in cognitive load and that they are valid and reliable (Paas et al., 1994).

However, prior studies have used the unidimensional scale to evaluate mental load with instructional material. For example, Paas et al. (1994) studied the effects of several computer-based training strategies on training, transfer, and cognitive load. Performance measures were accompanied by measures of cognitive load (using the same 9-point subjective rating scale used in the current study), and by spectral analysis of heart rate. The rating scale was more highly correlated with performance measures than the heart rate analysis. Further, the scale was more reliable and sensitive to differences in the type of problems, the variability in training conditions, and training transfer than heart rate analysis.

In the current study, the subjective mental effort ratings were not found to correlate with any of the performance measures, nor could they be qualified as artifacts of any questionnaire items. Though participants rated mental load as significantly higher for the dual-task trials than for baseline driving, the ratings for each task were not statistically different from each other, although performance measures differed between them. It could be argued that participants assigned comparative ratings of mental effort rather than reporting the actual mental effort of each trial. Each participant was asked to rate the first trial they performed immediately upon completion of that trial. Their answers were given verbally and recorded by the researcher. Half of the participants received the baseline driving trial first and half received the dual-task trial first. Ratings for baseline driving ranged from 1 (very, very low mental effort) to 8 (very high mental effort), while ratings for the dual-task trial ranged from 3 (low mental effort) to 8. After the second trial, participants were again asked to rate the trial they had just completed. Several participants asked the researcher to remind them of the rating they assigned to the first trial. Though the researcher politely declined to reveal the first rating, the questions imply that participants may have been seeking to rate comparatively. This could account for the pattern of results: a significant difference between baseline and dual-task ratings with all tasks, and the lack of difference between tasks.

Further, there is anecdotal evidence that the rating scale may not have been sufficiently meaningful with regard to the complex driving simulation tasks used in the current study. Several participants made comments to the researcher while assessing their ratings of mental load. These comments were unsolicited and did not garner a response, but were noted by the researcher. These shed doubt on the usefulness of the rating scale with regard to driving simulation tasks.

Two participants indicated that their enjoyment of the task affected their rating of mental effort. The first, a male participant in the high complexity eating condition performed the dual-task trial first and rated it 2 (very low mental effort), then performed the baseline trial, rating it 5 (neither low nor high mental effort). While completing the exit questionnaire, he recalled rating the first trial (eating) too low in comparison with how he'd rated baseline driving, so he re-rated the eating trial as 3 (low mental effort). He commented while making his decision that eating made the task of driving more enjoyable and therefore deserved to be rated lower than baseline driving. Another participant, a male in the low complexity memory condition performed the dual-task trial first, rating it 3 (low mental effort), and baseline driving 1 (very, very low mental effort). He said his rating wasn't really based on mental effort because he didn't feel that he used more mental effort while performing the memory task. Rather, his rating was based upon emotional discomfort. He said he "didn't like the questions; they were annoying."

The definition of mental effort did not appear to be universal among participants. For example, a male in the low complexity CD condition rationalized his rating of the dual-task trial by saying that it didn't take mental effort to switch CDs since he was not asked to select and insert a specific CD. Because of this, he felt that it required physical effort, but not much mental effort. He performed the dual-task trial first and rated it 4 (rather low mental effort), and rated baseline driving 3 (neither low nor high mental effort).

In contrast, a female in the high complexity memory condition, who rated both trials equally (6: rather high mental effort), commented that she didn't use more mental effort when she performed the memory task, but rather that the same level of mental effort was split between the two tasks during the dual-task trial. Thus, she believed that she used the same level of mental effort during both trials. The only difference was in how effort was allocated.

Going a step further, a male in the low complexity memory condition who performed the dual-task trial first said at the end of the experimental session that the task prompts coincided with events he needed to attend to, making it more difficult. His statement implied that he found the dual-task trial more mentally demanding. However, his mental ratings contradicted this. He rated the dual-task trial 5 (neither low nor high mental effort), and baseline driving 7 (high mental effort).

Finally, a male in the high complexity memory condition performed the dual task trial first, rating it 3 (low mental effort), and rating baseline driving as 5 (neither low nor high mental effort). He commented after the baseline trial, but before rating it, that he "zoned out . . . almost fell asleep." He confirmed his scores on the exit questionnaire. Perhaps he did not understand the progressive nature of the rating scale? Or perhaps his notion of "neither high nor low" better fit his perception of falling asleep than "very, very low mental effort." Perhaps he was rating the effort it took to stay awake during the task rather than the mental effort required by the task? Perhaps he was fatigued by the time he performed his second trial and the ratings appropriately reflected this? His performance suffered during baseline driving as well, resulting in one collision and one speeding violation, compared with one road departure during his dual-task trial. Though he may not have been aware of the road departure or the speeding violation, the participant was aware of the collision. Perhaps his ratings were simply congruent with his performance, irrespective of his sleepiness?

Subjective ratings could involve many evaluative processes, such as cognitive dissonance, hypothesis guessing, and issues of self portrayal. This is conjecture since there is no data by which to evaluate the possible processes involved with subjectively rating mental effort in the current study. What does seem clear is that participants did not use a shared interpretation of the rating scale, nor did they assign values to tasks based purely on the mental effort involved.

Mean speed and mental effort. Prior studies found a negative correlation between speed and mental effort. Participants tend to decrease speed as the mental effort of the task increases (Jenness et al., 2002, Wickens and Seppelt, 2002). This compensatory effect was not found in the current study. Though not statistically significant, on average participants actually tended to increase speed while performing the memory task and eating task, and decreased speed only while changing CDs.

On average, males rated mental effort with lower scores than females for both baseline and dual-task trials. This may have been an artifact of prior experience with computer/video games, with which males reported having more experience. Mental load was not correlated with any measure in the study. However, gender and computer/video game experience and driving game experience were highly correlated, and their possible effects upon the measures in this study cannot be overlooked.

The mental effort rating scale was used as a manipulation check of the level of central executive demand accorded to each task. The questions raised regarding its usefulness with the tasks involved in the present study also raise questions regarding the manipulation of mental effort, and hence, central executive demand. It is possible that a meaningful manipulation of central executive demand between tasks did not occur.

Does the STI Simulator Sufficiently Mimic Driving?

Generalizations of the results of the present study must rely on the assumption that the STI simulator sufficiently replicated driving and induced true driving behavior. However, there is no data that correlates scores from the STI simulator with actual driving measures (W. Allen, STI, personal communication, October 15, 2002). The argument could be made that the STI simulator more accurately induces behavior associated with video/computer games than that of driving. Though the STI simulator is controlled by a steering wheel and foot pedals, and the scenarios were carefully crafted to represent real-world driving conditions, the scenario is presented on a computer monitor as a 2-dimensional graphical scene with the consequences of errors far less severe than actual driving. In addition, participants consistently reported enjoying participating in the study. Some asked about the consequences of various actions, and whether they could come back to “play” with the simulator. It bears noting that if the study had been conducted in a real vehicle driving on a real road, participants would not have asked about the potential consequences of various actions.

While it is likely that the present study induced gaming behavior, it cannot be claimed that it did not also induce driving behavior. The presence of the number of years licensed to drive as a significant covariate indicates that the current results bear some relation to driving. Since both game experience and years licensed to drive were significant covariates, it is appropriate to assert that the present study probably induced both gaming and driving behaviors.

The Absence of an Effect of Complexity

The present study hypothesized a main effect of complexity, with high complexity expected to result in more driving errors than low complexity. This was not supported. A main effect of complexity was not found. As discussed earlier, this may have been due to the non-equivalence between groups on the number of miles driven per week, though a lack of correlation between this measure and any of the dependent measures makes this an unlikely culprit. It is more likely that the current study failed to manipulate complexity, or perhaps a different process was evoked in the low complexity conditions, such as underload (Young and Stanton, 2002).

Was complexity sufficiently manipulated? The complexity manipulation in the current study involved the amount of activity surrounding interactive events, and not the events themselves. Thus, complexity was defined as the number of pedestrians, parked cars, opposing traffic, and cross traffic. The number and types of events that required interaction from participants were held constant in all trials. This was necessary to provide all participants with equal opportunities to commit driving errors. The object of high complexity was to create more distraction and movement in the driving scene, though it could be argued that the presence of more oncoming traffic, cross traffic, pedestrians and parked cars created more collision opportunities for participants in the high complexity condition.

Levels of traffic complexity are arbitrary. In real world driving, there are many gradations of traffic complexity, with definitions that are not universally fixed. The volume of activity in the traffic environment is only one concept of traffic complexity. Others may include speed of travel, consistency of speed, deviance from traffic laws and expected conduct, or a ratio of pedestrian to vehicular activity, and may be strongly influenced by road characteristics such as road curvature, incline, visibility, number of lanes, number of intersections, type of intersections, and distance between intersections.

Previous studies have manipulated complexity by intersection density, traffic density, and scenery (Lee, Caven, Haake, and Brown, 2000) or simply traffic density (Olsson and Burns, 2000). The National Highway Traffic Safety Administration suggests important factors are the level of congestion and the complexity of the road (Hankey et al., 2000). The current study combined these suggestions and manipulated the volume of traffic congestion. However, since levels of congestion were not quantified in previous studies, there is nothing to compare with the current complexity manipulation. For this reason, the complexity levels in the current study must be considered arbitrary. The failure to find a main effect of complexity may indicate that the complexity manipulation failed. Perhaps the manner in which complexity levels were operationalized did not allow enough difference between them. Doubling the number of pedestrians, parked cars, and oncoming cars may not have produced a robust difference between the levels. Although no main effect of complexity was obtained, a significant difference between complexity levels while eating was obtained (more driving errors were found with low complexity). Thus, some difference between the complexity levels seems to exist. However, it is not clear why an interaction with eating facilitated an effect. Whether this is due to a less than robust complexity manipulation, or is indicative of different processes evoked by the manipulation is not obvious. For example, one complexity level may have required a sufficient level of attention while the other may have required too little and exerted an effect of underload.

Underload. Young and Stanton (2002) studied the effects of underload on performance in a driving simulator experiment. They manipulated several levels of vehicle automation, and measured mental workload through measures of eye movements and performance on a secondary task that used the same resource pools as driving (visual-spatial-manual), as posited by MRT. They were particularly interested in whether attentional capacity changed with attentional demands. Participants were told to perform the secondary task only when they had the capacity to do so without impairing the primary task. Results showed that responses on the secondary task did not vary consistently with the amount of attention directed to the task. Secondary task scores increased as automation increased; however, the allocation of attention to the secondary task became less efficient, suggesting that the size of the resource pool can change. This provided evidence for an association between task demands and attentional resource capacity.

Young and Stanton (2002) proposed malleable attentional resources theory (MART) to account for the effects of underload on performance. Their theory posits that attentional capacity can change size in response to changes in task demands. Thus, performance decrements associated with mental underload can be explained by a lack of appropriate attentional resources.

Considerable effort was taken to create a realistic driving scenario for the present study, and according to feedback obtained from participants this was achieved. However, several aspects of a real-world driving environment are not present, such as diverse plant life, advertisements, street noise, radio or recorded entertainment, and interactive factors such as interest in the surroundings and route selection. The real world is filled with stimulation and subtle variation that a simulator cannot duplicate. It is possible that the realistic driving events, without the presence of real-world variation, produced an effect associated with underload. Hence, the primary driving task in this study may have required less mental effort than studies with less realistic driving tasks, such as pursuit-tracking (Strayer and Johnston, 2001) or driving laps in a Sony PlayStation^{TM} racing game (Jenness et al., 2002). Along this line, poor results obtained with low complexity traffic may indicate an effect of mental underload (Young and Stanton, 2002), and the high traffic complexity may approach an effective level of mental stimulation for performing the driving task.

Level of Driving Errors

Another aspect that challenges the realistic qualities of the driving task used in the current study is the number of driving errors that were present in baseline driving (see Table 4). For example, one third of all participants experienced at least one collision during baseline driving. Clearly, one-third of all drivers engaging in an eight-minute drive would not be expected to have a collision in the real world. The number of potential collision events was higher in the simulation than would normally be encountered in a typical eight-minute drive. This was partially intentional in order to gain sufficient data in a short period of time for interpretable results. However, care was taken not to produce an expectancy effect for potential collision events during the trials. Perhaps the mix of more potential collision events and the care not to induce abnormal vigilance or expectancy effects aided in producing a high collision rate.

Additionally, two potential collision events in each trial involved a vehicle stopped in the lane ahead of the participant's vehicle. According to Sanders and McCormick (1993), people are not very good at determining closure rate. We are better able to discern closure, than to discern the speed with which the closure is occurring. Evans (1991) asserts that an important factor in rear-end accidents, in which a vehicle was stopped in the driver's lane, was a strong expectation that vehicles in traffic lanes are always moving at some rate of speed, rather than standing still. In addition, the perception of distance and closure rate may have been complicated by the simulator's two-dimensional presentation of the driving scene. Though a stopped car in the lane ahead is not an uncommon sight, particularly in the driving environment around campus, these two scenario events in each trial may have induced a higher collision rate in this study.

It may also be argued that the reaction time allotted to some of the events in the current study was insufficient for participants to avoid a collision. According to Oglesby (1975), the U.S. standard for perceptual reaction time used by traffic engineers in highway design is 2.5 sec (as cited by Sanders and McCormick, 1993). However, from a review of the literature to that point, Hooper and McGee (1983) asserted that the standard needed to be increased to 3.2 sec to accommodate the 85th percentile driver (as cited by

Sanders and McCormick). Events presented to participants in the current study ranged in reaction time from 2 to 10 sec. Only three events allowed less than 3-second reaction time. The first involved a partial pullout, in which the event vehicle pulled out from a parked position on the side of the road and began to enter the participant's lane, then stopped after partially entering the lane. The driver had 2 seconds to react to this event. The partial pullout accounted for 16% of all collisions in the study, with the majority (12%) occurring during baseline driving. The other two events involved several cars pulling into the participant's lane from a line of stopped vehicles, the first allowing a 2-second reaction time, and the second allowing 3 seconds. The line event accounted for less that 9% of the total collisions in the study.

Fifty-four percent of the collisions in the study were attributed to an event involving a line of stopped vehicles, some in the traffic lanes, and some parked alongside the road, with a pedestrian crossing the street in front of them. In version 1, the scene resembled a car accident, with an ambulance and police car among the vehicles in the line of stopped cars. Version 2 involved the same movement of vehicles and the pedestrian, but the vehicles were busses. Both versions allotted more than a 5-second reaction time. Version 1 yielded 9 collisions (18% of all collisions), all 9 during dual-task trials. Version 2 yielded 18 collisions (36% of all collisions), with 14 from dual-task trials and 4 from baseline driving trials. The increased collisions with version 2 were probably due to reduced visibility of the pedestrian due to the height of the buses. This may have slowed perceptual recognition of the event and reduced reaction time. The presence of a police car in version 1 may have exerted an effect as well. However, the difference in number of collisions with dual-task and baseline driving shows that the collisions were avoidable in both versions and more apt to occur when the participant was performing a secondary task while driving.

Generalizing to Other Populations

Driving behavior relies heavily on visual perception of the driving environment. A measurable aspect of this process involves visual-scan patterns of the driving scene. Studies have shown that novice drivers sample the roadway environment more narrowly than experienced drivers (Sanders and McCormick, 1993; Recarte and Nunes, 2000). Thus, novice drivers receive less information from their peripheral field of view. Since visual input forms the basis for making judgments about aspects of the driving task, the difference in visual-scan patterns is thought to be indicative of the higher incidence of driving errors found with novice drivers.

The sample used in the present study is representative of young, novice drivers, and cannot be construed to represent the general driving public. On average, participants had been licensed to drive 2.9 years. One participant had been licensed for 11 years, another for 8 years, a third for 6 years, and the remainder had been licensed for less than 5 years. In addition, the average age was 19.6 years with a mode of 19 years. Many of the effects found in this study may be partly attributable to the age and inexperience of the participants.

Driving errors in the present study have been quantified, but not qualified. In essence, driving errors have been attributed to human operators of the STI driving simulator who committed errors of the type identified and recorded in this study. These errors were conceptualized to vary as a function of the addition of a secondary task. This interpretation is simplistic since it does not consider the entire system in which the human participant operated. Whether errors were a function of the attentional demands of secondary tasks, of the stimuli (driving simulator and scenarios), of the controls (steering wheel and foot pedals), of the perception of risk and the consequences of error, or some mixture of these, cannot be ascertained. The present study did not attempt to evaluate errors through any formal theory or procedure of error analysis.

CEM-overlap: A New Model

The CEM model proposed in this paper posits that the central executive is active whenever a complex task is performed, during task selection, or when switching between tasks (Baddeley, 1986; Baddeley, 1996, as cited by Wickens and Hollands, 1999). MRT deals chiefly with overlaps in input-processing-output modalities. The data in the current study suggests that with complex tasks, the central executive may not be separate from resource pools, but rather may act as a modality itself that requires an overlap of demand to produce a decrement in performance. This revised model will be referred to as the CEM-overlap model.

CEM-overlap emerged as an explanation of the pattern found in the current results, and as such it must be treated with scrutiny. The model revises CEM to include the central executive as a modality rather than a separate entity. It can also be seen as a revision to MRT – to include central executive processing as a modality. When the central executive modality is not required by two concurrent tasks, interference will not occur and predictions will match MRT. When the central executive modality is demanded by both tasks, CEM predictions will predominate (see Table 12). To illustrate, this section will interpret the results of the current study through CEM-overlap. Material from previously presented discussions will also be integrated to form a more cohesive picture.

First, it is important to review the effects of three measures from the exit questionnaire. The first, number of years licensed to drive (question 3), was used as a covariate in analyses. Additionally, groups were equivalent with respect to this measure, so its use as a covariate served to reduce error variance and reduce the probability of a Type II error. No residual issues are anticipated nor addressed with regard to this measure in the following discussion, save for its affirming affect that driving behavior has been evoked to some degree in the current study.

The second measure, computer/video game experience (question 6), was also used as a covariate. However, groups were not equivalent with regard to this measure. A main effect of gender was obtained, with females reporting less experience than males. The use of gender as a factor and game experience as a covariate should have compensated for any threat of confounding. However, the following discussion will include the impact of inequalities between groups on this measure.

The third measure, number of miles driven per week (question 4), was not used as a covariate since it was not correlated with any dependent measures. However, groups were not equivalent on this measure. A main effect of gender was obtained as well as an interaction of gender with complexity. This presents problems for interpretations of gender effects, complexity effects, and gender by complexity effects. Proper conclusions must treat gender and miles driven per week as a package, and take care to interpret complexity through its interaction with this package. It is possible that the number of miles driven per week had no actual impact upon the acquired results; however, the following discussion will assume that it did since a pattern emerges in the results with this assumption.

Integrating the previous discussions, the current study can be interpreted through the CEM-overlap model as follows. If experience with computer/video games and driving games exerted an effect between genders in the current study, then males in the study could be processing the task using different cognitive mechanisms than females. Experience can automate a task, reducing or removing central executive involvement (Shiffrin and Schneider, 1984). Revisiting Table 5, males in the low complexity conditions reported playing computer/video games more frequently than any other group in the study. Assuming that the task of driving the STI simulator evoked computer/video game behavior, and that playing games was somewhat automatic for these participants, there may have been little central executive demand involved in the task of driving. If little or no central executive demand were present in the driving task, then tasks involving central executive demand could not produce interference. In this case, CEM-overlap would predict MRT as shown in Figure 1b. As Figure 9a shows, the pattern found with males in the low complexity conditions did match the MRT model. Any possible effects of underload would be expected to strengthen the MRT pattern. Underload could also cause the participant to attend more to the secondary task than would otherwise be expected, if the primary task lacks sufficient stimulation, further accentuating the pattern.

Referring again to Table 5, males in high complexity conditions reported less computer/video game experience than males in low complexity conditions, though the difference was not significant. Males in high complexity conditions reported driving 185 more miles per week, on average, than males in the low complexity conditions. While not statistically significant, it is important to note that the average number of miles driven per week by males in high complexity is nearly twice that of low complexity. Thus, males in high complexity should have more expertise with driving, and slightly less expertise with games. Additionally, an effect of underload was not suspected in the high complexity conditions, allowing a greater degree of central executive involvement and attention to the primary task.

The pattern of results for males in the high complexity conditions (see Figure 9a) does not match MRT. Instead, these results present a problem for MRT, particularly since the STI simulator induced gaming behavior as well as driving behavior. Under this interpretation, males in the high complexity condition would have experienced higher central executive demand while performing the primary driving task, thus creating a situation in which central executive demands of the secondary task could produce overlap, and hence, interference in the form of decrements in performance. The results support this assertion to a great extent. Consistent with CEM-overlap, males in the high complexity memory task, the task thought to place the most demand on the central executive, resulted in the most driving errors. The central executive demand in the high complexity memory condition may have been compounded by the number of miles reportedly driven per week by this group (see Table 5). On average, this group reported driving 144 miles per week, as opposed to the low complexity memory group who reported driving an average of 450 miles per week.

The explanation of the results of the other tasks is more elusive. All secondary tasks sharing modalities with the primary task would be expected to result in decrements in performance. However, both CD-changing and eating slightly facilitated driving performance, though not significantly. Looking further into the questionnaire data (Table 5), males in the high complexity eating and CD-changing conditions had more experience eating while driving (question 8) and changing CDs while driving (question 10), than did their male counterparts in the low complexity conditions. Perhaps some level of expertise with the secondary tasks aided this result, though it is still puzzling.

Females in the low complexity conditions reported very little experience with computer/video games. In fact, participants in this condition reported less game experience than all other groups in the study. Hence, this group can be seen as having low game expertise, and the simulated driving task itself may not have been automatic, requiring more central executive involvement with the primary task. The CEM-overlap model would predict interference from all three tasks; however, underload effects may interact with these predictions by introducing a lack of attention, or tendency to attend more to the secondary task if the primary task lacks sufficient stimulation. The results fit these assumptions to an extent (see Figure 9b). Decrements in performance on the memory task support the notion of an overlap of demand on the central executive between the primary and secondary tasks. Interpretation of the eating and CD-changing conditions is more obscure. Perhaps the level of stimulation provided by the eating task drew sufficient attention away from the primary driving task to produce the resulting decrements in performance. The facilitation effect found with the CD-changing task cannot be explained through prior experience, CEM-overlap, MRT, or by any means available in the present study. Its position relative to the memory task presents a problem for MRT predictions, and in this respect it provides useful information.

Finally, females in the high complexity conditions also reported little computer/video game experience. In addition, they reported driving less than half as much per week as both genders in the low complexity conditions, and nearly one-fourth as much as males in the high complexity conditions (see Table 5 and Figure 7). If the number of miles driven per week has an impact, this lack of expertise should translate into higher central executive involvement with all tasks since the driving task itself should require a higher level of involvement with novice participants. However, rather than supporting CEM-overlap predictions, the results approximate MRT predictions with the two visual-spatial-manual tasks producing the greatest decrements in driving performance (see Figure 9b), and the memory task (auditory-verbal-vocal) facilitating primary task performance. It is possible that novice performance and expert performance, requiring different processes, may require different predictive models as well.

A small sample size was used in the current study. It could be argued that the current results are simply the artifact of the small sample size, and that numerous replications of this study could each yield different results. However, an interpretable pattern of results would not be expected to emerge from spurious data. The emergence of a pattern in the results of the current study indicates that these findings carry some validity.

A single study cannot sufficiently elucidate a trend or fully support a model, and no such assertions are made here. Even the best design coupled with clear, hypothesized results should not garner undue reliance, but rather should inspire further investigation to test the reliability and limitations of its results. The current study does not purport to have answered questions so much as it has posed them, and hopes to stimulate a more comprehensive and cohesive approach to future investigations of driving behavior.

Suggestions for Future Study

There is a need to refine and test the CEM-overlap model. Some level of central executive involvement is present in every task. It would be important to ascertain the level of demand at which decrements begin to appear, whether that level must be met by one of the tasks alone, or by both separately or in combination, and whether thresholds modulate. CEM-overlap should be tested in a more direct fashion with tasks of varying levels of complexity and varying types of overlap before it can be interpretably applied to compound complex behaviors like those attempted in this study. Further, for CEM-overlap to be meaningful, an appropriate measure of mental effort is needed. Once this is refined, tasks must be identified which reliably use different levels of mental effort, as well as other types of modalities, and eventually, combinations of the two.

More generally, studies involving driving behavior should take care to measure covariates dealing with game experience, driving experience, and any relevant variables in the study which could interact with known or suspected gender differences. The current study suggests that gender blocking is a necessity and should be considered early in the design process. It would be prudent to administer entrance questionnaires, match participants on a number of variables, and assign matching pairs within genders to conditions. This would guard against unforeseen problems with non-equivalent groups.

In the present study, analysis on the change scores of the composite measure “driving errors” was conducted without covariates (ANOVA) and with two covariates (ANCOVA with computer-video game experience and number of years licensed as covariates) to highlight the need to use covariates. The ANOVA revealed a significant main effect of gender: on average, adding a secondary task increased driving errors with females but not with males. This was a seemingly robust effect (p = .021). However, with the introduction of the two covariates into the analysis, the gender effect was completely ameliorated (p = .854). Instead, a three-way interaction was obtained (shown in Figures 8a, 8b, 9a, and 9b). Task, complexity, and gender must be viewed in terms of their combined effects upon "driving errors." If analyses concluded prior to identifying pertinent covariates, the main findings of this study would have highlighted gender differences with driving behavior. Instead, this paper can more responsibly offer gender differences attributable to game experience and driving experience.

Even with these precautions, any obtained gender differences should be scrutinized and alternative causes of differences explored. Covariates should be used in analysis, and the reader should be wary of reported results in which covariates are not used. These precautions will guard against faulty findings that unfairly negatively label one group of people with a specific detriment or behavior, as could easily have arisen with gender in the present study.

Secondary task performance should be measured and analyzed in addition to primary task performance. This is because decrements in primary task performance will only occur when the participant over-allocates attentional resources to the secondary task. According to Hegarty et al. (2000), and Bourke, Duncan, and Nimmo-Smith (1996), people tend to allocate more resources to the task they perceive to be most important or most demanding. During a driving experiment, this should be the primary task of driving. Hegarty et al. asserts that it is essential to report decrements with secondary task performance as well as those with primary task performance, particularly when results of a time-restricted study are to be generalized to tasks that are commonly performed for longer durations of time. The cognitive load of the concurrent tasks may provide some insight into possible performance decrements that may be seen if tasks are performed for longer durations, thus illuminating areas for possible future study. However, this will require that both primary and secondary tasks be studied alone and in combination, to ascertain baseline performance for each.

Eye movements, particularly visual-scan patterns would also be a useful measure. Studies have shown visual-scan patterns associated with novice drivers, following closely behind a lead vehicle (Sanders and McCormick, 1993), and traveling familiar routes (Mourant and Rockwell, 1970). Recarte and Nunes (2000) found an association between eye movements and cognitive load, and Young and Stanton (2002) found eye movements to be a useful measure of mental workload in a simulated driving task.

For safety reasons, we rely on facsimiles of driving to study driving behavior. It is widely thought that driving simulators provide good construct validity, approximating the complex and interactive task of driving much more closely than say, a pursuit-tracking task. However, there is no data to support this assertion. Effects of the apparatus must be considered when interpreting results.

Driving is a complex task that requires the full range of human capabilities: perception, decision-making, and motor skills, all in a highly coordinated fashion and often under stressful conditions. To rate it as simply a visual-spatial-manual task with intermittent central executive involvement, as partially controlled and partially automatic, may be the best description we currently have, but it is still an oversimplification. One of the problems with studying driving behavior is that the driving task itself is so complex and its components are difficult to isolate. Efforts to isolate components may obstruct efforts to effectively study authentic driving behavior since driving errors are likely caused by a dynamic interaction of environment, experience, physiological state, emotional state, individual differences, perceptual-, manual- or cognitive- abilities and the distractions of secondary tasks. However, we have no choice but to investigate components in a methodical fashion, ruling out those whose effects are not robust. Once we identify the components of distraction, and the behavioral factors that contribute to accidents, we must then discover how to apply these findings to reduce accidents and create safer vehicular systems.

References

Alm, H., & Nilsson, L. (1995). The effects of a mobile telephone task on driver behavior in a car following situation. *Accident Analysis and Prevention, 27*(5), 707-715.

Baddeley, A. D. (1986). *Working memory*. New York: Oxford University Press.

Bolstad, C. A. (2001). *Age-related factors effecting the perception of essential information during risky driving situations.* Paper presented at the Human Factors and Ergonomics Society 45th Annual Meeting, Santa Monica, CA. Retrieved March 10, 2002, from http://www.satechnologies.com/html/papers/2001.shtml

Bourke, P. A., Duncan, J., & Nimmo-Smith, I. (1996). A general factor involved in dual-task performance decrement. *Quarterly Journal of Experimental Psychology, 49*A, 525-545.

Briem, V., & Hedman, L. R. (1995). Behavioral effects of mobile telephone use during simulated driving. *Ergonomics, 38*(12), 2536-2562.

Evans, L. (1991). *Traffic safety and the driver.* New York: Van Nostrand Teinhold.

Gopher, D., & Braune, R. (1984). On the psychophysics of workload: Why bother with subjective measures? *Human Factors, 26*(5), 519-532.

Hankey, J. M., Dingus, T. A., Hanowski, R. J., Wierwille, W. W., Monk, C. A., & Moyer, M. J. (2000). The development of a design evaluation tool and model of attention demand. *NHTSA Driver Distraction Internet Forum: methods and techniques for measuring distraction.* Retrieved January 15, 2003. http://www-nrd.nhtsa.dot.gov/departments/nrd-13/driver-distraction/PDF/8.PDF

Hegarty, M., Shah, P., & Miyake, A. (2000). Constraints on using the dual-task methodology to specify the degree of central executive involvement in cognitive tasks. *Memory & Cognition, 28*(3), 376-385.

Jenness, J. W., Lattanzio, R. J., O'Tolle, M., & Taylor, N. (2002). Voice-activated dialing or eating a cheeseburger: Which is more distracting during simulated driving? *Proceedings of the Human Factors and Ergonomics Society*, *USA, 46,* 592-596.

Kennedy, R. S., Jentsch, F., Smither, J.A. (2001). Looming detection among drivers of different ages. *Proceedings of the Human Factors and Ergonomics Society, USA, 45*, 240-244.

Lee, J. D., Caven, B., Haake, S., & Brown, T. L. (2000). Speech-based interaction with in-vehicle computers: The effect of speech-based e-mail on drivers' attention to the roadway. Iowa City, Iowa: University of Iowa, Department of Industrial Engineering, National Center for Injury Prevention and Control Grant #R49-CCR703640-10. *NHTSA Equipment Design Features (Impacts on Safety).* Retrieved January 15, 2003. http://www-nrd.nhtsa.dot.gov/departments/nrd-13/driver-distraction/PDF/27.PDF

Martens, M. H., & Van Winsum, W. (2000). Measuring distraction: The peripheral detection task. *NHTSA Driver Distraction Internet Forum: methods and techniques for measuring distraction.* Retrieved January 15, 2003. http://www-nrd.nhtsa.dot.gov/departments/nrd-13/driver-distraction/PDF/34.PDF

Mollenhauer, M. A., Hulse, M. C., Dingus, T. A., Jahns, S. K., & Carney, C. (1997). Design decision aids and human factors guidelines for ATIS displays. In Y. I.

Noy (Ed.), *Ergonomics and Safety of Intelligent Driver Interfaces* (pp. 23-61). Mahwah, NJ: Lawrence Erlbaum Associates.

Mourant, R., & Rockwell, T. (1970). Mapping eye-movement patterns to the visual scene in driving: An exploratory study. *Human Factors, 12*(1), 81-87.

Nygren, T. E. (1991). Psychometric properties of subjective workload measurement techniques: Implications for their use in the assessment of perceived mental workload. *Human Factors, 33*, 17-33.

Olsson, S., & Burns, P. C. (2000). Measuring driver visual distraction with a peripheral detection task. Gothenburg, Sweden: Volvo Technological Development Corporation. *NHTSA Driver Distraction Internet Forum: methods and techniques for measuring distraction.* Retrieved January 15, 2003. http://www-nrd.nhtsa.dot.gov/departments/nrd-13/driver-distraction/PDF/6.PDF

Paas, F. G. W. C. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology, 84*(4), 429-434.

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*(1), 63-71.

Paas, F. G. W. C., Van Merrienboer, J. J. G., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills, 79*, 419-430.

Rasmussen, J. (1986). *Information processing and human-machine interaction: An approach to cognitive engineering.* New York: North-Holland.

Recarte, M. A., & Nunes, L. M. (2000). Effects of verbal and spatial-imagery tasks on eye fixations while driving. *Journal of Experimental Psychology: Applied, 6*(1), 31-43.

Redelmeier, D. A., & Tibshirani, F. J. (1997). Association between cellular-telephone calls and motor vehicle collisions. *The New England Journal of Medicine, 336*, 453-458.

Ruthruff, E., Johnston, J. C., & Van Selst, M. (2001). Why practice reduces dual-task interference. *Journal of Experimental Psychology: Human Perception and Performance, 27*(1), 3-21.

Sanders, M. S., & McCormick, E. J. (1993). *Human factors in engineering and design* (7th ed.). New York: McGraw-Hill.

Schneider, W., Dumais, S. T., & Shiffrin, R. M. (1984). Automatic and control processing and attention, in R. Parasuraman and D. R. Davies (eds.), *Varieties of Attention*. London: Academic.

Shiffrin, R. M., & Schneider, W. (1984). Theoretical note: Automatic and controlled processing revisited. *Psychological Review, 91*(2), 269-276.

Staplin, L., Ball, K., Park, D. Decina, L. E., Lococo, K. H., Gish, K. W., & Kotwal, B. (1997). *Synthesis of human factors research on older drivers and highway safety, Volume 1: Older driver research synthesis.* PA: Scientex Corporation Traffic Safety Division; sponsored by Office of Safety and Traffic Operations R & D, Federal Highway Administration (NTIS No. FHWA-RD-97-094).

Strayer, D. L., & Johnston, W. A. (2001). Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular telephone. *Psychological Science, 12*(6), 462-466.

Systems Technology (2002). STISIM Drive II, (Version 2.00.10) [Computer software]. Hawthorne, CA: Systems Technology, Inc.

Uyeda, K. M., & Mandler, G. (1980). Prototypicality norms for 28 semantic categories. *Behavior Research Methods and Instrumentation, 12*(6), 587-595.

Wickens, C. D. (1984). Processing resources and attention. In R. Parasuraman and R. Davies (Eds.), *Varieties of Attention*. New York: Academic Press.

Wickens, C. D., & Hollands, J. G. (1999). *Engineering Psychology and Human Performance* (3rd ed.). New Jersey: Prentice Hall.

Wickens, C. D., Sandry, D. L., & Vidulich, M. (1983). Compatibility and resource competition between modalities of input, central processing and output. *Human Factors, 25*(2), 227-248.

Wickens, C. D., & Seppelt, B. (2002). Interference with driving or in-vehicle task information: The effects of auditory versus visual delivery. *Aviation Human Factors Division, Institute of Aviation, University of Illinois at Urbana-Champaign* (GM contract TCS16231 WIC). Retrieved March 27, 2003. http://www.aviation.uiuc.edu/new/html/arl/TechPdf/02-18.pdf.

Young, M. S., & Stanton, N. A. (2002). Malleable attentional resources theory: A new explanation for the effects of mental underload on performance. *Human Factors, 44*(3), 365-375.

Tables

Table 1

Task blocks for each experimental condition, with visual demand and central executive involvement ratings, and approximate time, glance and manual manipulation measures from pilot study.

Task	**Visual Demand (Number of Glances)**	**Manual Manipulation (Seconds)**	**Central Executive Involvement**	**Total Task Time**
Changing CDs	**High**		**Moderate**	
Insert new CD	5	11.4		20 sec
Select song	2	2.2		9 sec
Select song	2	2.2		9 sec
Eating	**Moderate**		**Low**	
Drink water	2	8.9		20 sec
Eat chip	1	3.9		10 sec
Eat chip	1	3.9		10 sec
Memory Search	**Low**		**High**	
Category	0	0		20 sec
Antonym	0	0		9 sec
Antonym	0	0		9 sec

Table 2

Category and antonym blocks presented in the memory search task.

Block	**Categories**	**Antonyms**	
1	Tools	Selling	Stiff
2	Trees	Creation	Agreement
3	Musical Instruments	Praise	Shallow
4	Fruit	Deny	Rejection
5	Clothing	Noise	Permanent
6	Animals	Joy	Assisting
7	Furniture	Alert	Chaos
8	Birds	Shy	Ignorance

Table 3

Elements of scenarios with high vs. low complexity conditions.

Element	**High Complexity**	**Low Complexity**
Opposing traffic	152 vehicles	74 vehicles
Cross traffic	22 vehicles	11 vehicles
Parked cars	111 vehicles	55 vehicles
Pedestrians	96	48
Stop signs	2	2
Traffic signals	5	5
Events	17	17

Table 4

Means and t-values for paired t-tests and percent of participants who committed each type of error (%). Means are based on all participants, not just those who committed errors.

Secondary Task	**Dual-task**			**Baseline Driving**			**Change**			
	%	***M***	***SD***	***%***	***M***	***SD***	***M***	***SD***	***t***	***p***
CD										
Driving Errors	68%	2.44	2.55	68%	1.94	2.08	.50	2.99	.67	.514
Road Departures	56%	1.13	1.31	25%	.44	.89	.69	1.14	2.42	**.029**
Collisions	44%	.50	.63	44%	.63	.81	-.13	.89	-.57	.580
Speed Violations	31%	.81	1.33	38%	.88	1.63	-.06	1.98	-.13	.901
Mean Speed		21.97	2.90		22.26	1.81	-.29	2.27	-.51	.620
Eating										
Driving Errors	81%	2.94	2.74	68%	1.13	.96	1.81	3.31	2.19	**.045**
Road Departures	75%	1.5	1.46	38%	.44	.63	1.06	1.61	2.64	**.019**
Collisions	31%	.63	1.02	25%	.31	.60	.31	1.35	.92	.370
Speed Violations	31%	.81	1.52	31%	.38	.62	.44	1.55	1.13	.276
Mean Speed		22.08	1.73		21.98	1.14	.10	2.00	.20	.847
Memory										
Driving Errors	94%	2.50	1.83	75%	1.81	2.20	.69	2.68	1.03	.320
Road Departures	56%	.69	.70	38%	.56	.89	.13	1.09	.46	.652
Collisions	44%	.86	.73	31%	.44	.81	.13	.96	.52	.609
Speed Violations	63%	1.25	1.34	38%	.81	1.42	.44	1.86	.94	.362
Mean Speed		22.04	3.54		20.86	1.78	1.19	3.50	1.36	.195
Overall										
Driving Errors	81%	2.63	2.37	71%	1.63	1.83	1.00	3.00	-	-
Road Departures	63%	2.80	4.06	33%	.76	1.90	.63	1.33	-	-
Collisions	40%	.56	.80	33%	.46	.74	.10	1.08	-	-
Speed Violations	42%	.96	1.38	35%	.69	1.29	.27	1.78	-	-
Mean Speed		22.03	2.76		21.70	1.69	.34	2.69	-	-

Table 5

Descriptive statistics from the exit questionnaire shown by complexity, gender, and task.

Item			**Male**						**Female**					
			CD		**Eating**		**Memory**		**CD**		**Eating**		**Memory**	
			M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
1	Age	H	19.3	1.3	19.0	0.8	19.8	1.0	18.8	1.0	21.0	1.6	20.3	2.5
		L	20.8	4.2	19.3	1.0	18.8	1.5	19.0	0.0	19.8	1.5	19.5	1.7
3	Years licensed	H	2.4	1.1	2.6	0.8	2.9	1.4	3.0	0.9	3.0	1.8	4.0	2.7
		L	5.0	4.0	2.0	0.8	2.1	1.2	2.3	0.8	2.3	1.4	3.3	1.9
4	Mi per wk driven	H	388	25	319	163	450	464	68	36	131	113	109	135
		L	329	348	128	45	144	139	213	130	291	86	148	108
6	Game exp.	H	3.5	.6	2.8	1.0	4.0	0.8	2.5	0.6	2.5	0.6	2.0	0.8
		L	4.0	1.2	4.0	1.2	3.5	1.0	1.5	0.6	3.0	0.8	1.5	0.6
7	Driving game exp.	H	2.3	1.0	2.8	0.5	2.5	0.6	2.3	0.5	2.3	0.5	1.5	0.6
		L	2.3	1.0	3.0	0.0	2.5	0.6	1.3	0.5	2.6	0.5	1.5	0.6
8	Eat while drive	H	2.3	0.5	2.3	1.0	2.0	0.8	3.5	0.6	2.5	0.6	2.5	1.3
		L	1.8	1.0	2.3	1.0	2.0	0.0	2.8	1.0	2.5	1.7	3.5	1.0
9	Cell while drive	H	3.0	1.4	3.0	0.8	2.5	1.0	2.8	1.0	2.5	1.7	3.0	0.8
		L	2.8	1.5	2.3	1.0	2.0	0.8	3.8	1.0	2.8	1.0	3.5	1.0
10	CD while drive	H	4.5	0.6	4.0	0.8	3.8	0.5	3.5	0.6	4.3	1.0	3.8	1.3
		L	4.3	1.0	4.3	.05	3.0	1.8	3.8	1.0	4.3	1.0	3.8	1.0
11	Talk while drive	H	3.8	0.5	3.0	.08	3.3	0.5	3.3	1.0	3.0	0.8	3.3	1.5
		L	3.5	0.6	3.5	0.6	3.5	0.6	3.8	0.9	4.0	0.8	2.5	1.0
12	Compare to driving	H	3.8	1.0	3.0	0.8	2.0	0.8	3.3	0.5	2.3	1.3	2.8	0.5
		L	2.8	1.0	3.3	0.5	2.3	1.0	3.0	0.8	3.5	1.0	2.8	1.0
13	Compare difficulty	H	2.8	1.0	3.5	1.0	2.0	0.0	2.5	0.6	1.8	1.0	2.5	1.3
		L	2.3	1.3	3.0	0.8	2.8	1.0	2.0	0.8	1.8	0.5	2.3	1.3
14	Hours ago last ate	H	6.0	6.2	8.0	6.5	3.0	1.5	5.8	5.5	3.3	1.3	2.5	0.6
		L	6.6	7.7	6.5	7.2	9.0	4.8	2.3	2.0	1.8	1.0	5.0	4.1
15	Mental load (change)	H	1.8	2.6	0.3	1.7	2.3	1.3	0.5	1.7	2.0	2.9	1.5	1.3
		L	2.5	1.0	1.8	0.5	0.5	1.9	2.0	0.8	3.0	0.8	1.0	0.0

Table 6

Spearman correlations of the most highly correlated items from the exit questionnaire with gender and the dependent measures. Bold items are significant at alpha .05.

Change Scores	Driving Errors	Road Depart.	Mean Speed	Collision	Speed Viol.	Gender	Game Exp.	Driving Game Exp.	Years Licensed
Driving Errors	r_s =1.0	**r_s = .704** **p = .000**	**r_s = .414** **p = .003**	**r_s = .646** **p = .000**	**r_s = .709** **p = .000**	**r_s = .331** **p = .022**	**r_s = -.361** **p = .012**	r_s = -.206 p = .160	**r_s = -.336** **p = .020**
Road Depart.		r_s =1.0	r_s = .044 p = .769	**r_s = .426** **p = .003**	r_s = .189 p = .199	r_s = .283 p = .051	**r_s = -.294** **p = .043**	**r_s = -.325** **p = .024**	**r_s = -.304** **p = .035**
Mean Speed log			r_s =1.0	**r_s = .362** **p = .012**	**r_s = .483** **p = .001**	r_s = .269 p = .065	r_s = -.188 p = .198	r_s = .186 p = .207	**r_s = -.295** **p = .042**
Collision				r_s =1.0	r_s = .166 p = .259	**r_s = .292** **p = .044**	r_s = -.089 p = .545	r_s = -.113 p = .443	r_s = -.173 p = .240
Speed Violation					r_s =1.0	r_s = .123 p = .405	**r_s = -.307** **p = .034**	r_s = -.037 p = .800	r_s = -.258 p = .077
Gender						r_s =1.0	**r_s = -.654** **p = .000**	**r_s = -.466** **p = .004**	r_s = -.008 p = .958
Game Exp.							r_s =1.0	**r_s = .572** **p = .000**	r_s = .062 p = .675
Driving Game Exp.								r_s =1.0	r_s = -.032 p = .828
Years Licensed									r_s =1.0

Table 7

3 (task) by 2 (complexity) by 2 (gender) ANOVA source table for change scores on the composite dependent variable "driving errors."

Source	**Sums of Squares**	***df***	**Means Squares**	***F***	***p***
Complexity	8.33	1	8.33	1.02	.320
Task	16.13	2	8.06	.98	.384
Gender	48.00	1	48.00	5.86	**.021**
Complexity by task	2.54	2	1.27	.15	.857
Complexity by gender	1.33	1	1.33	.16	.689
Task by gender	2.38	2	1.19	.15	.866
Complexity by task by gender	48.29	2	24.15	2.95	.065
Error	295.00	36	8.19		
Total	470.00	48			

Table 8

3 (task) by 2 (complexity) by 2 (gender) ANCOVA source table for change scores on the composite dependent variable "driving errors."

Source	**Sums of Squares**	***df***	**Means Squares**	***F***	***p***
Game Covariate	53.31	1	53.31	8.29	**.007**
Yrs Licensed Cov.	41.27	1	41.27	6.42	**.016**
Complexity	6.30	1	6.30	.98	.329
Task	14.48	2	7.24	1.13	.336
Gender	.22	1	.22	.03	.854
Complexity by task	11.96	2	5.98	.93	.404
Complexity by gender	4.33	1	4.33	.67	.418
Task by gender	21.68	2	10.84	1.69	.200
Complexity by task by gender	87.01	2	43.50	6.77	**.003**
Error	218.57	34	6.43		
Total	470.00	48			

Table 9

Rank ordered means indicating change from baseline driving from the ANCOVA three-way interaction of complexity with gender and task on the composite measure "driving errors."

Gender	**Complexity**	**Task**	**Mean Change**	***SE***
Male	Low	CD	+ 3.48	1.54
Female	Low	Eating	+ 3.15	1.28
Male	High	Memory	+ 2.68	1.40
Male	Low	Eating	+ 2.65	1.39
Female	High	Eating	+ 2.24	1.28
Female	Low	Memory	+ 1.63	1.46
Female	High	CD	+ 1.49	1.28
Male	High	CD	- 0.59	1.31
Male	Low	Memory	- 0.78	1.31
Male	High	Eating	- 0.86	1.27
Female	High	Memory	- 1.15	1.36
Female	Low	CD	- 1.96	1.51

Table 10

3(task) by 2(complexity) by 2(gender) between-subjects MANOVA on change scores.

Source	Sums of Squares	*df*	Means Squares	*F*	*p*
Complexity					
Collisions	1.687	1	1.687	1.800	.188
Speed Violations	.521	1	.521	.162	.690
Road Departures	5.333	1	5.333	3.200	.082
Log Mean Speed	.000	1	.000	.064	.801
Task					
Collisions	1.542	2	.771	.822	.448
Speed Violations	2.667	2	1.333	.415	.664
Road Departures	7.125	2	3.563	2.138	.133
Log Mean Speed	.007	2	.003	1.301	.285
Gender					
Collisions	4.688	1	4.688	5.000	**.032**
Speed Violations	4.687	1	4.687	1.458	.235
Road Departures	6.780	1	6.780	4.05	.052
Log Mean Speed	.008	1	.008	2.959	.094
Complexity by task					
Collisions	1.625	2	.812	.867	.429
Speed Violations	1.167	2	.583	.181	.835
Road Departures	.042	2	.021	.012	.988
Log Mean Speed	.003	2	.001	.496	.613
Complexity by gender					
Collisions	1.688	1	1.688	1.800	.188
Speed Violations	6.021	1	6.021	1.873	.180
Road Departures	.000	1	.000	.000	1.000
Log Mean Speed	.003	1	.003	.974	.330
Task by gender					
Collisions	7.125	2	3.563	3.800	**.032**
Speed Violations	3.500	2	1.750	.544	.585
Road Departures	.125	2	.063	.038	.963
Log Mean Speed	.000	2	.000	.060	.941
Complexity by task by gender					
Collisions	2.375	2	1.188	1.267	.294
Speed Violations	15.167	2	7.583	2.359	.109
Road Departures	3.875	2	1.938	1.163	.324
Log Mean Speed	.005	2	.002	.929	.404
Error					
Collisions	33.750	36	.938		
Speed Violations	115.750	36	3.215		
Road Departures	60.000	36	1.667		
Log Mean Speed	.094	36	.003		
Total					
Collisions	55.000	48			
Speed Violations	153.000	48			
Road Departures	102.000	48			
Log Mean Speed	.121	48			

Table 11

Means and t-values for paired t-tests on mental load ratings.

Secondary Task	***M***	***SD***	***t***	***p***
CD				
Dual-task	6.33	1.64		
Baseline Driving	4.78	1.53		
Change	1.55	2.20	2.81	**.013**
Eating				
Dual-task	6.38	1.78		
Baseline Driving	5.02	1.50		
Change	1.36	1.32	4.11	**.001**
Memory				
Dual-task	5.97	1.74		
Baseline Driving	4.84	1.87		
Change	1.48	1.31	4.52	**.000**

Table 12

Like MRT, the CEM-overlap model requires that two or more concurrent tasks compete for the same modality for interference to occur. Central executive processing has been added to the MRT model as a processing modality.

Input Modalities	**Processing Modalities**	**Output Modalities**
Auditory	Verbal	Vocal
Visual	Spatial	Manual
	Central Executive	

Figures

Figure 1a. Driving error result patterns predicted by the Central Executive Mediating Model (CEM) proposed in this paper.

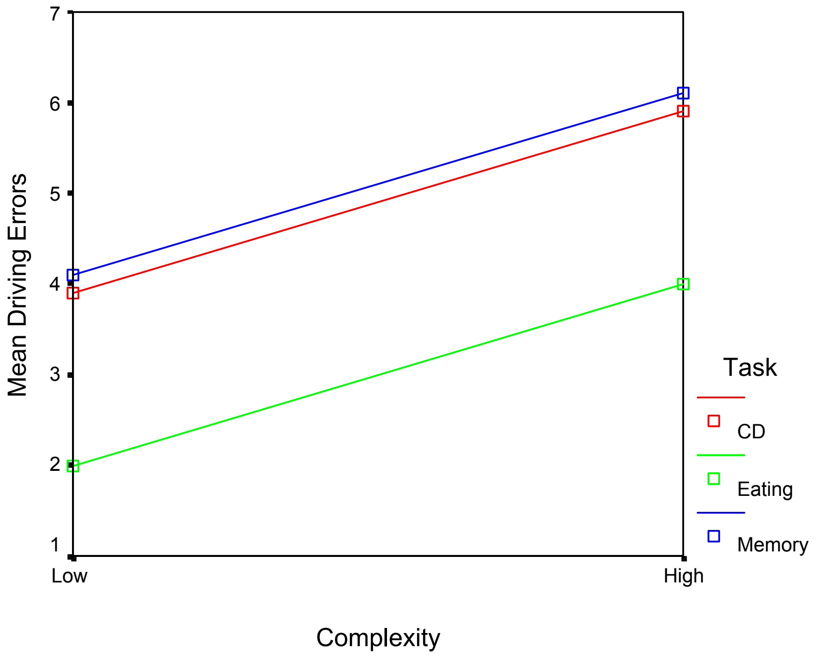


Figure 1b. Driving error result patterns predicted by Multiple Resource Theory (MRT).

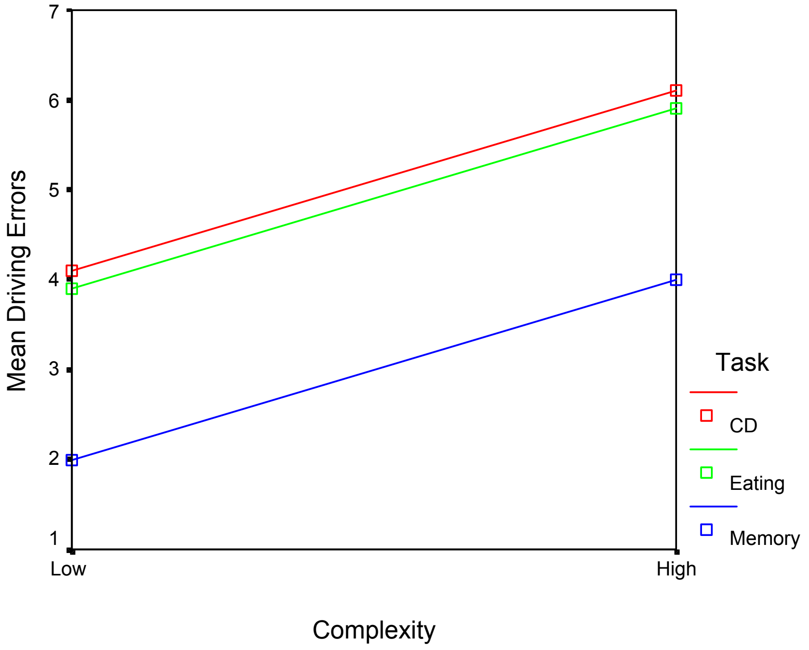


Figure 2. Experimental design. Subscripts refer to scenario versions.

		Traffic Complexity	
		Low	High
Secondary Task	Changing CDs	Practice → Control $_1$ → Dual-task $_2$ Practice → Control $_2$ → Dual-task $_1$ Practice → Dual-task $_1$ → Control $_2$ Practice → Dual-task $_2$ → Control $_1$	Practice → Control $_1$ → Dual-task $_2$ Practice → Control $_2$ → Dual-task $_1$ Practice → Dual-task $_1$ → Control $_2$ Practice → Dual-task $_2$ → Control $_1$
	Eating	(Same pattern as above)	(Same pattern as above)
	Memory Search	(Same pattern as above)	(Same pattern as above)

Figure 3. The STI Driving Simulator.

Figure 4. Apparatus used in eating trials.

Figure 5. Apparatus used in CD-changing trials.

Figure 6. Participant's view of the driving scene. A lever on each side of the steering wheel allowed a 90-degree view to each side of participant's vehicle.

Figure 7. Interaction of gender and complexity on questionnaire item 4, number of miles per week participants reported driving.

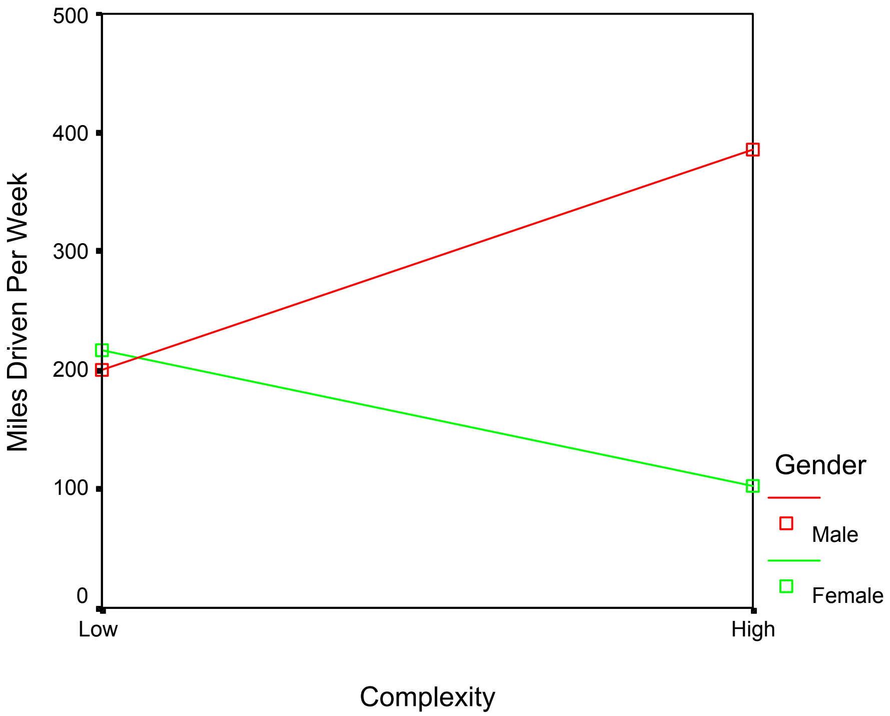


Figure 8a. Change in number of driving errors as a function of task and gender at low complexity.

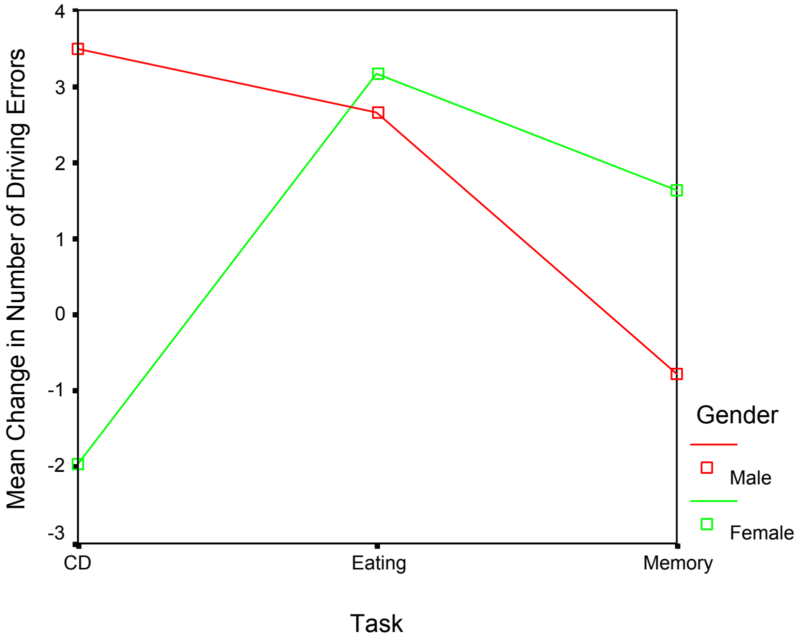


Figure 8b. Change in number of driving errors as a function of task and gender at high complexity.

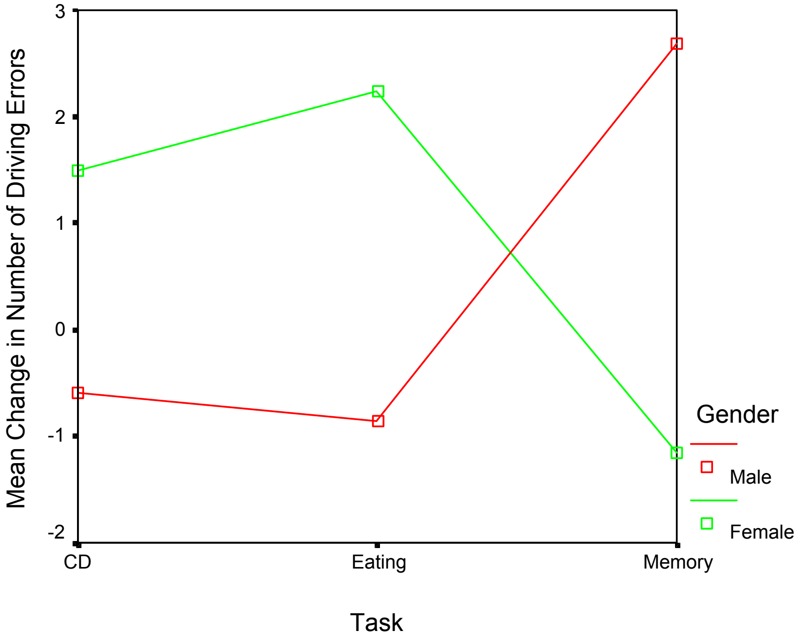


Figure 9a. Change in number of driving errors as a function of task and complexity for males.

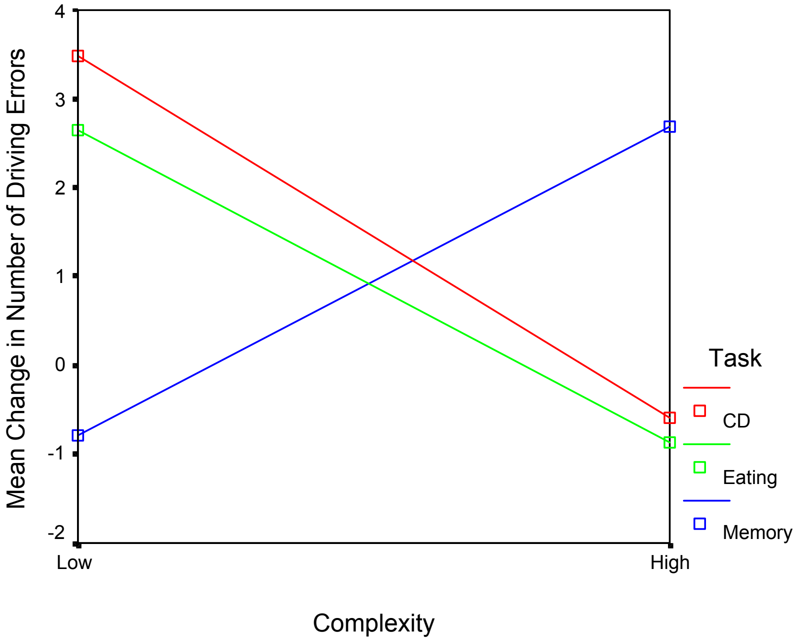


Figure 9b. Change in number of driving errors as a function of task and complexity for females.

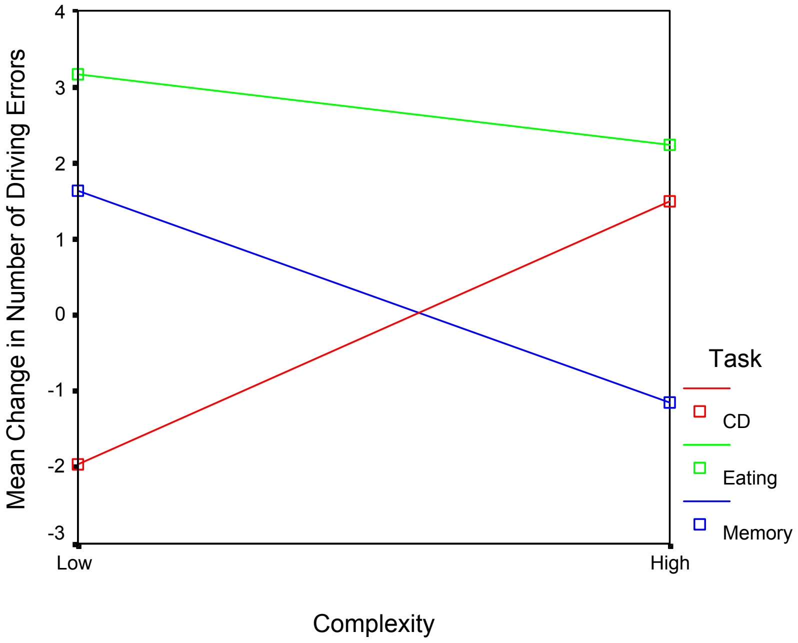


Figure 10. Change in number of collisions as a function of task and gender.

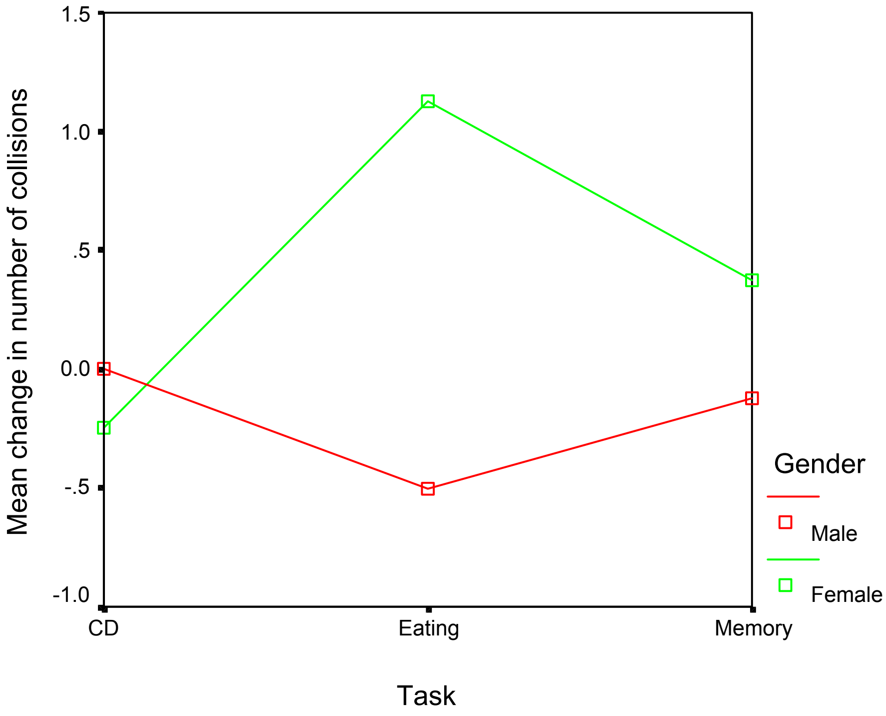


Figure 11. Change in mental ratings as a function of task and complexity.

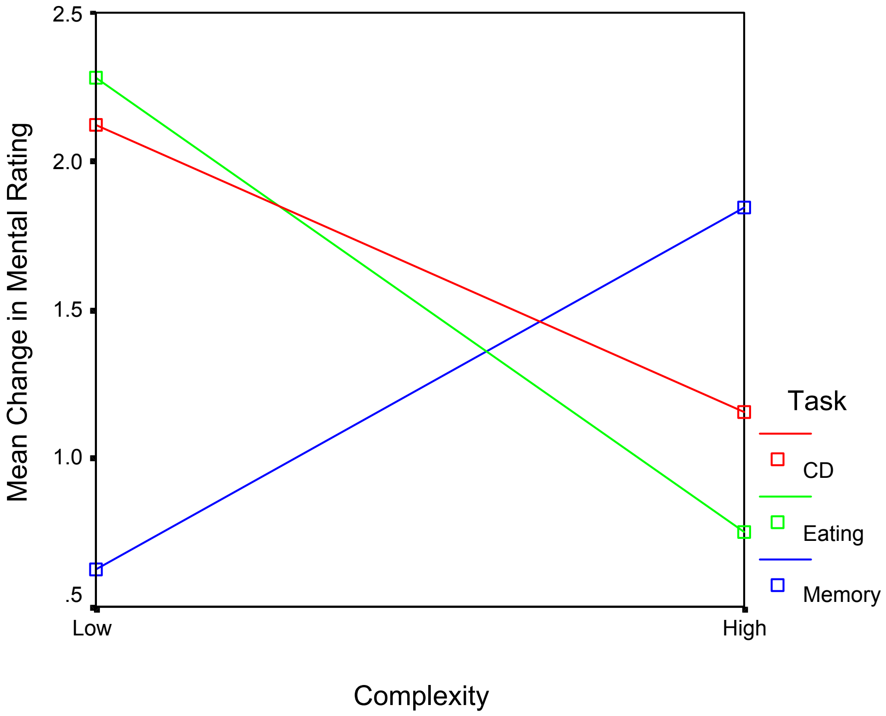
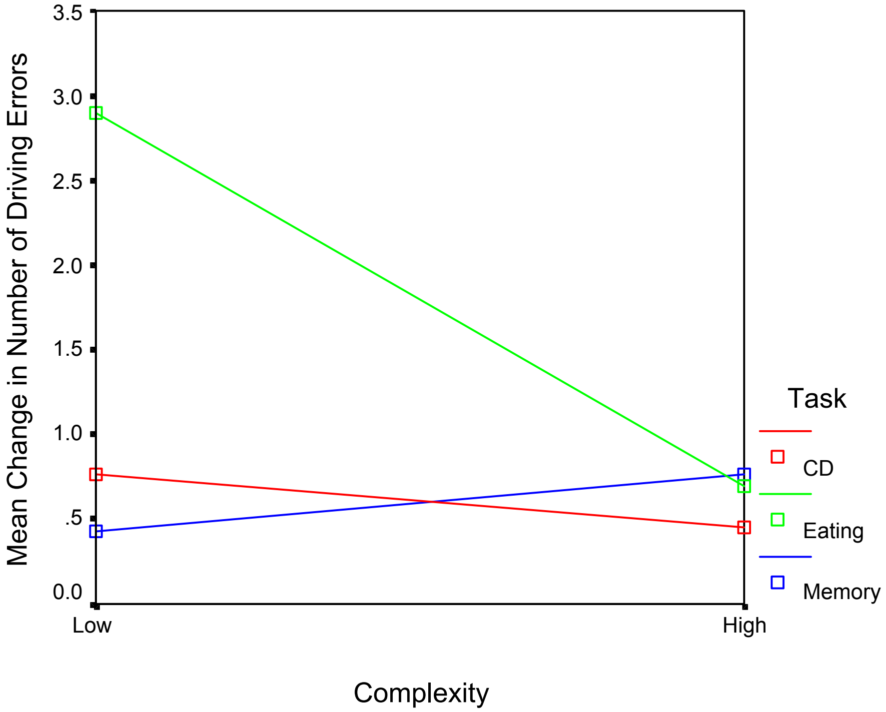


Figure 12. Complexity by task collapsed over gender from ANCOVA on driving errors, to compare with predictive models.



APPENDIX

CALIFORNIA STATE UNIVERSITY, NORTHRIDGE
Driving Simulation Project
EXIT QUESTIONNAIRE

1. Age: ________ (years)

2. Gender: M F (circle one)

3. How long have you been licensed to drive a motor vehicle? ________ (years)

4. How much do you drive? _____ miles per week

5. Have you ever used a driving simulator before? __ Yes __ No

6. How often do you play computer or video games? (circle one)

Never	Have played a few times	Sometimes	More than 10 hours per month	More than 5 hours per week

7. How often do you play computer or video games that involve driving? (circle one)

Never	Have played a few times	Sometimes	More than 10 hours per month	More than 5 hours per week

8. How often do you eat while driving? (circle one)

0%	25%	50%	75%	100%
Never	Not often	Sometimes	Often	Always

9. How often do you talk on a cell phone while driving? (circle one)

0%	25%	50%	75%	100%
Never	Not often	Sometimes	Often	Always

10. How often do you change CDs, tapes or the radio station while driving? (circle one)

0%	25%	50%	75%	100%
Never	Not often	Sometimes	Often	Always

11. How often do you drive with a passenger in the car? (circle one)

0%	25%	50%	75%	100%
Never	Not often	Sometimes	Often	Always

12. How would you compare the simulation you just completed with actual driving? (circle one)

0%	25%	50%	75%	100%
Not like driving at all	A little like driving	Similar to driving	Very similar to driving	Exactly like driving

13. Compared with actual driving, the simulated driving was: (circle one)

Much harder than actual driving	A little harder than actual driving	About the same as actual driving	A little easier than actual driving	Much easier than actual driving

14. When was the last time you ate? How many hours ago?

____

15. How much mental effort did you use during each of the following (check one):

a. Driving only

	1	Very, very low mental effort
	2	Very low mental effort
	3	Low mental effort
	4	Rather low mental effort
	5	Neither low nor high mental effort
	6	Rather high mental effort
	7	High mental effort
	8	Very high mental effort
	9	Very, very high mental effort

b. Driving while (secondary task inserted)

	1	Very, very low mental effort
	2	Very low mental effort
	3	Low mental effort
	4	Rather low mental effort
	5	Neither low nor high mental effort
	6	Rather high mental effort
	7	High mental effort
	8	Very high mental effort
	9	Very, very high mental effort