MEASURING DUAL TASK COST USING THE PERFORMANCE OPERATING CHARACTERISTIC: THE EFFECT OF EMOTIONAL WORDS ON ONE'S FUNCTIONAL FIELD OF VIEW

BY

Jeffrey Dressel

Submitted to the graduate degree program in Psychology
and the Graduate Faculty of the University of Kansas
in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Chairperson

Committee members

Date defended

The Dissertation Committee for Jeffrey Dressel certifies
that this is the approved version of the following dissertation:

MEASURING DUAL TASK COST USING THE PERFORMANCE OPERATING CHARACTERISTIC: THE EFFECT OF EMOTIONAL WORDS ON ONE'S FUNCTIONAL FIELD OF VIEW

Chairperson

Committee members

Date accepted

Abstract

Jeffrey Dressel, M.A.

Department of Psychology, May 2008

University of Kansas

This dissertation discusses design elements that should be utilized for optimal measurement of dual task performance, and reviews literature suggesting that these elements are underutilized. Participants seem to be able to effectively “tune out” one or the other task in a dual task paradigm, though traditional analyses and POC analyses converge to inform us that under these experimental conditions (which may not require adequate cognitive load), UFOV performance is not as greatly impacted by concurrent verbal tasks as pilot data and theory suggest. While smaller than expected, these dual task costs have implications in an applied setting, as 19% of subjects exhibited UFOV scores under dual task conditions that would predict more than double the risk of injurious accident. Finally, highly arousing negatively valent verbal stimuli may lead to greatest interference with visual attention performance.

Measuring Dual Task Cost Using the Performance Operating Characteristic: The Effect of Emotional Words on One's Functional Field of View

Table of Contents:

This dissertation makes contributions to both the applied and basic science fields. As a basic science goal, the experiments investigate to what degree the emotional content of verbal stimuli might change patterns of attentional resource allocation and salience of visual stimuli. As an applied science goal, the same experiments investigate the role this diminished visual attention performance may play in accident risk to drivers conversing on cellular phones.

There is compelling evidence to support the notion that emotional stimuli, especially when negatively emotionally valent, may be especially capable of diminishing visual attention performance. Thus, it is important to investigate this interference effect to inform both applied and basic science.

In order to investigate these effects thoroughly, certain methodological protocol should be adopted. Performance operating characteristics (POC curves) describe the ability to perform two tasks concurrently in greater statistical resolution than the static, equal priority dual-task methods we have used to collect pilot data. Additionally, the POC paradigm allows one's ability to trade-off prioritization of performance between the two tasks to be measured. This is important because it answers the common question of a driver's ability to simply "prioritize" the driving task (i.e., visual attention). It may be the case that some verbal information is especially difficult to attenuate, or to "set" to a low level of priority; this research explores that possibility.

Rationale

The Useful Field of View (UFOV)

Atchley, P. and Dressel (2004) described that the functional, or useful, field of view (UFOV) is a cognitive mechanism that may be capable of accounting for increased accident rates for drivers conversing on a cellular phone. The functional field of view is a subset of the entire visual field present at any given eye fixation that can be processed to the level of ability to guide action. The functional field of view is assessed using a dual-task paradigm of visual search for a target within a large region of visual space, with the addition of a concurrent visual task at the foveal location. For example, an observer is typically asked to make a judgment (e.g., the detection of a visual feature) about a target that appears at the center of the visual field while concurrently detecting the location (radial direction) of a second peripheral target. In addition, distracting information can be presented at peripheral locations to compete with the peripheral target for selection. As the central task becomes more demanding, as distractors are added, and as the duration of presentation is decreased, some observers show a decline in the ability to accurately detect the peripheral target, indicating that the field of view within which they can accurately extract information has decreased.

Reductions in the functional field of view have been shown to occur in some persons at an advanced age (Ball & Owsley, 1991; Ball, Owsley, Stalvey, Roenker, Sloane & Graves, 1998). These reductions have been shown to reliably predict an increase in accident risk, with a 40% reduction in the functional field of view effectively doubling accident risk (Owsley, et al. (1998). These results were replicated by Ball, et al. (2006), who demonstrated that low performance on subtest 2 of the UFOV test (divided attention task, 353ms or longer) was associated a relative accident risk of 2.02. This finding was comparable to other predictors that were investigated by Ball, et al. (2006). These other predictors included the Closure subtest of the Motor Free Visual Perception Test (MVPT), which involves choosing from a set of incomplete pictures which exemplar would match a given complete picture, and part B of the Trail Making Test, which involves using a pencil to connect a sequence of alternating integers and letters. Further support for the usefulness of the UFOV in predicting accident rate is provided by Hoffman, et al. (2005), who found in their sample of 155 adults that UFOV (subtest 2 and 3, divided attention and selective attention) performance was unrelated to on-road accident risk, but a modest predictor of driving simulator performance. Receiver operating characteristics (ROC curves) revealed greatest sensitivity of 85% classification as "at risk" with 45% false positive rate. The UFOV has been demonstrated to be a consistent, reliable predictor of motor vehicle accidents of older adults. Edwards, et al. (2005) report test-retest reliability for PC versions of the UFOV test as high (r= 0.735 for a touch-screen interface, as we employ in our research), and as having high correlation with standard, "professional" versions of the test (r= .0746 for a touch-screen interface).

Although the UFOV appears to be a valid and reliable predictor of accident risk in older adults, there is debate as to the cognitive or perceptual mechanism employed during UFOV performance. Whereas Sekuler and Ball (1986) and Ball et al., (1988) have interpreted decreased UFOV performance as a *constriction* of the UFOV, adopting a spatial attention, or "spotlight" metaphor, this position may not be readily adopted by all researchers using the UFOV test. For example, Seiple et al. (1996) suggested that UFOV performance could be related to other stimulus conditions than merely eccentricity of the peripheral targets, such as backward masking, presence of distractors, and luminance. In a series of experiments parametrically varying these factors, these researchers found that masking and distractor presence affected UFOV performance (consistent with the intended design of the UFOV test), but that luminance and target eccentricity had no effect on test performance. That is, the eccentricity of the peripheral target (in UFOV subtests 2 and 3) had no effect on target detection, while presence of distractors did have an effect on target detection. Thus, Seiple et al. (1996) conclude that UFOV performance decrements do not reflect a constriction of a region of visual attention, but rather, a higher-order cognitive performance decrement, such as executive function decline, or decrease in complex visual search performance. Furthermore, Sekuler, Bennet, and Mamelak (2000) have reached similar conclusions. These researchers found no eccentricity-dependent performance decrements, and conclude that UFOV decline is best conceptualized as a decrease in efficiency with which older subjects are able to extract information from a cluttered scene, rather than strictly a constriction of visual spatial attention.

The notion that UFOV performance decrements in older adults may be explained as an inability to extract relevant information from a field of irrelevant information, or an inability to effectively utilize an executive function to prioritize relevant information, is consistent with notions of dual-task performance couched in a general resource theory (see below). That is, if the UFOV test is viewed as a complex *cognitive* task, and not a lower-order *perceptual* task, it follows that the UFOV may more readily be interfered with by cognitively demanding tasks, such as concurrent conversations (or verbal tasks). To preview, it may be the case that cognitive demands of conversation-like tasks deplete general cognitive resources such that adequate resources are not available to perform the UFOV task at short presentation durations and thus, attend to visual information while driving.

This rationale led Atchley, P. and Dressel (2004) to an investigation of the impact of verbal tasks on the functional field of view. Using a functional field of view testing apparatus and a highly controlled verbal task, it was found that none of the younger adults yielded UFOV scores predicting an increased risk category when performing the functional field of view task in isolation. However, adding a concurrent verbal task yielded a considerable reduction in the functional field of view in the 18-25 year old subjects. All participants at least were categorized as showing “some difficulty with divided attention” category while few were categorized as “extremely impaired”. Applying the accident risk norms to the functional field of view performance of the younger observers in the concurrent verbal task condition led to the prediction that 6% of the participants were about 16 times more likely to have an injurious accident.

Subsequent unpublished studies by Dressel and Atchley, P. have investigated the relative effects of different verbal tasks (i.e., different categories of words, different memory and linguistic processing demands) on the UFOV. Directional verbal tasks have yielded compelling results which, given the directional aspect of the UFOV task, is consistent with a theory of common code interference proposed by Wickens’ (2002) multiple resource model of attention. Wickens’ (2002) model suggests that processing signals using similar “codes”, or related information, may produce greater cross-talk interference than processing signals carrying unrelated information. However, the largest interference effects observed to date appear to be the product of proto-conversational content consisting of negative emotional valence. The theoretical implications of these findings as well as the findings themselves, are described in greater detail below.

Emotion and Attention

Researchers attempting to investigate emotion may face unique challenges. For instance, some researchers suggest that there is a vague understanding of what is meant by “emotion”, and that the terms commonly used to label emotional experiences do not adequately and exclusively map onto physiological and psychological phenomena (e.g., Davidson & van Reekum, 2005). That is, an emotion commonly referred to as “fear” may include such different psychological states as apprehension, panic, a startle response, feelings of inferiority, and more. Thus, it is important to note that any discussion of emotion in research literatures may be subject to this vague interpretation.

Research investigating “cue utilization”, conducted in the 1950’s suggests that as stress increases, the use of cues for performing tasks decreases. For example, Callaway and Thompson (1953) evoked in participants a negative emotional state and subsequently asked these participants to perform a number of size and distance matching tasks. Reduced accuracy on these tasks while experiencing negative emotion was suggested to be due to decreased use of peripheral cues or, to use current terminology, a reduction of attention. In a review, Easterbrook (1959) notes a number of studies suggesting that stress results in a “shrinkage of the perceptive field” (page 180). While these studies predate the classic notions of spatial attention, they do suggest that experiencing negative emotion may have a negative impact on attention.

More recently, research has suggested a distinction between two dimensions of emotional stimuli: arousal and valence. These dimensions have unique effects, both physiologically and behaviorally (Heller & Nitschke, 1998; Lang, Greenwald, Bradley & Hamm, 1993). Arousal refers to the perceived intensity of the stimulus, and is typically rated from “low” to “high”. Valence refers to a general pleasantness dimension and is rated from the polar “pleasant” or “positive” to “unpleasant” or “negative” (e.g., for most individuals, “rainbow” is a positively valent image or word, and “malaria” is a negatively valent image or word).

The valence dimension of emotional stimuli has received great attention by the scientific community. Heller and colleagues (e.g., Heller, 1990; Heller, Nitschke, & Lindsay, 1997) have suggested that valence and arousal dimensions may be processed in different regions of the brain. Specifically, these researchers have found that valence seems to be processed in the anterior regions, with positive valence displaying more activity in right than left hemispheres, and negative valence displaying more activity in left than right hemispheres. Furthermore, the arousal dimension appears to be processed in the right parietotemporal region, in addition to being processed in subcortical structures.

Some specific regions of the brain that have been associated with emotional information processing have been demonstrated to be active during visual attention tasks. Thus, shared neurophysiological resources suggest a biological foundation for the large interference effect of emotional stimuli on visual attention. The amygdala, which projects to visual areas V1 and V2, is more active when participants view emotional faces than faces expressing no, or less, emotion (Pessoa & Ungerleider, 2002). Also, the anterior cingulate gyrus has been shown to integrate emotion and attention, and thus may be a source of processing limitations (Yamasaki, LaBar, & McCarthy, 2002).

Bradley, et al (2003), using fMRI, have found elevated BOLD levels in the occipital lobe while participants view emotion-laden scenes, such as those depicting physical violence and threat, relative to scenes with equal visual complexity, but lower subjective ratings of emotional content. Thus, vision centers in the brain may experience greater activity, or readiness, when emotional content is viewed.

The effect of emotion on attention may be especially strong in clinical populations. Atchley, R., Ilardi, and Enloe (2003), using a divided visual field paradigm, found that depressed and previously depressed subjects showed a speed and accuracy benefit for making valence judgments of negatively valent words primarily processed by the right hemisphere. Non-depressed control subjects showed the opposite effect, displaying an advantage for positively valent words. Thus, experience of a depressed mood (here, in a clinical context, not an induced context as in other studies) may increase a readiness to process negative emotional information.

In addition to a neurophysiological foundation for the interaction of emotion and attention, there is support for a theoretical perspective that describes the interaction of emotional experience and visual attention. Davidson (2003) refers to this emotion-contingent activity as “motivated attention”. Motivated attention involves the orienting, or priming, of our perceptual systems to process threat-related (i.e., emotional) information. Ellis (2005) describes emotion as an early mechanism that is capable of motivating individuals to have a propensity to process information in the environment that suggests something is risky, or “off” (p. 28). This motivation mechanism allows information relevant to survival to be prioritized above other information, and to enter our conscious awareness more readily than information not related to threats. Some researchers (e.g., Davison, 2003; Levenson, 2003) suggest that such a mechanism could have an evolutionary advantage, as heightening awareness of threats in the environment would promote survival.

Motivated attention has been observed in a variety of tasks in which emotion modulates cognition. Bower (1981) employed a flanker task in which a to-be-remembered emotional (pleasant, unpleasant) target word was presented centrally, with emotional flanker words surrounding the target. Subjects showed less ability to inhibit the processing of flankers matching the mood they had been induced to experience (i.e., subjects in a pleasant mood were less likely to inhibit pleasant words). This is consistent with a theory of motivated attention suggesting that the experience of emotion can "prime" the processing emotional content presented to an individual, or orient an attentional system to emotional information.

Similarly, Calvo, Castillo, and Fuentes (2006) employed a lexical decision task in which central and parafoveal probes/primes varied in emotional valence scores. Emotional state of the participants was manipulated by interleaving the display of emotional scenes between trials (in a within-subject design, thus, a single participant would be induced into either a positive or threat-laden negative mood). Results indicated greater interference effects of the parafoveal probe if the probe *matched* the emotional state of the participant. That is, participants were less likely to attenuate negative stimuli if they had been induced to feel negative emotion, as if their attentional mechanisms had been oriented to process negative information.

Further support for the effect of emotion on attention has been demonstrated by Schupp, et al. (2004). In these studies, subjects viewed pictures from the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 2001). Content of the pictures was varied, with subjects viewing 25 pleasant pictures, 10 neutral pictures, and 25 unpleasant pictures. Additionally, a white noise startle probe was presented with one of two stimulus onset asynchronies (SOA; 2.5s or 4.5s). Researchers found that P3 event-related potentials (ERPs), thought to be an indicator of processing “oddball”, or startling information, were smaller for subjects who viewed emotional pictures than for those who viewed neutral pictures. Furthermore, P3 amplitudes were smallest when participants viewed negative emotional pictures. Thus, these findings are also consistent with a theory of motivated attention suggesting that emotional information, especially negatively valent emotional information, may be especially salient, and individuals may both favor attending to such information and may have difficulty shifting attention away from it.

In its general form, a theory of motivated attention has been supported by a robust collection of experimental paradigms. Robinson, et al. (2004) have refined this general theory, and suggest a two-stage model of the interaction of emotion and attention. Emotional arousal, these authors suggest, may be processed quickly, and at a preattentive level. Emotional valence, while processed quickly, may be processed (i.e., evaluated, discriminated, and responded to) later than arousal. Furthermore, Robinson et al. (2004) have shown that arousal and valence interact, such that low arousal negative stimuli may be processed slowly (nearly 1100ms) and high arousal positive stimuli may likewise be processed slowly (1120 ms), whereas high arousal negative stimuli may be processed in the least amount of time (nearly 1030ms) and low arousal positive stimuli may be processed moderately quickly (over 1040ms). Thus, it seems that participants are best able, or are perhaps most prepared or “motivated”, to process arousing negative stimuli, while they are least able, or least prepared or motivated, to process arousing positive stimuli. Convergent evidence for this hypothesis is provided by Atchley, R., et al. (in press). Atchley, R., et al. (in press) used a similar paradigm to Atchley, R., Ilardi, and

Enloe (2003, above), recruiting depressed, formerly depressed, and non-depressed individuals, and employing a divided visual field paradigm requiring subjects to make evaluations of emotional words. However, Atchley, R., Ilardi and Enloe (2003) manipulated both arousal and valence dimensions of emotion. In a pattern of results consistent with Robinson et al. (2004), findings indicated that all subjects demonstrated an advantage of evaluating highly arousing negatively valent words, and *low* arousing positively valent words.

These studies suggest that the greatest impact on an attentional system may then come from highly arousing negative stimuli, followed by positive stimuli with low arousal, negative stimuli with low arousal, and finally positive stimuli with high arousal. These findings will be discussed in greater detail in the hypotheses section, below.

With specific regard to the effect of emotional word processing on UFOV (and thus driving) performance, our lab has collected data from three pilot experiments, demonstrating the magnitude of emotional word processing (especially negatively valent emotional words). The following three experiments demonstrate the strength of the distracting effect of emotional word processing relative to alternative language processing, such as mere semantic association, and directional translation.

Pilot Experiment 1

Method

Participants

Twenty undergraduate students at the University of Kansas participated for course credit. All participants reported English to be their primary language and had normal or corrected to normal visual acuity.

Materials

Images were presented on a 17-inch EloTouch monitor, which allowed for touch-screen response inputs. Three visual tasks from the functional field of view software developed by Visual Awareness, Inc., IL, were used. The details of these tasks will be discussed in more detail in the procedures section. Conversational stimuli consisted of moderate frequency words, chosen from the University of South Florida Word Association Norms (Nelson, et al., 1998). Correct responses were coded as those comprising the 5 most frequent responses listed by these norms. Conversational stimuli were recorded using Sound Studio 2.2 software (Felt Tip Software) for later analysis.

Procedure

Visual tasks.

The functional field of view software presents three tasks. The first visual task consisted of a sequence of stimuli in which a center fixation point was presented, followed by a car or truck icon inside a white box. Presentation times varied from 16ms – 500ms, determined by an adaptive staircase procedure (described below). The presentation of the icon was followed by a 1 second random-dot mask. The mask was followed by a response screen that displayed both the car and truck icon (always in the same positions, to the right and left of fixation). Participants were required to make a discrimination judgment between the two possible targets. They were allowed as much time as needed to respond. Responses were made by touching the icon of their choice on the response screen.

The second visual task consisted of the same central discrimination task as task one, with the addition of an item (a car) simultaneously presented in the perimeter at one of eight locations (top, bottom, left, right, or diagonal positions) 12 deg from center simultaneously with the centrally located car or truck icon. Participants responded first with a discrimination judgment to which item they had seen presented in the center, then with a localization judgment for where they had seen an item presented in the perimeter. The response screen for this judgment consisted of boxes at the eight possible target locations connected to the central location by radii. Participants were required to indicate only the location of the second target. The identity of this second target was unimportant to their response selection.

The third visual task duplicated the second, with the addition of triangle-shaped distractors arranged in the remainder of the visual field. The distractors were presented at all possible target locations in addition to the target, as well as at two locations along the radii from the center to the target location and possible target locations. The first ring of distractors was 4 deg from center. The second ring of distractors was 8 deg from center. Participants responded as in the second visual task with a discrimination judgment for the central task and a localization judgment for the secondary task.

Conversational task.

Participants listened to words read by a single (common to each participant) male experimenter from the list mentioned above. Words were presented via a common telephone’s speakerphone function, placed less that two feet from the participant on the same desk as the monitor for the visual task. Participants were to listen to a given word and respond with the first “meaningfully related” word that came to mind. Subjects were informed that this could include antonyms (e.g., light : dark), or words related in context (e.g., desk : chair). Participants were told to respond to both auditory and visual stimuli as quickly and accurately as possible.

Results

Conversational task performance

In all pilot experiments, participants performed at high levels of accuracy (90% or above). This leads us to believe that significant care was taken on the part of the participants to perform the secondary task with appropriate effort. That is, participants did not merely "say anything" in response to our verbal prompts.

Visual task performance

For all three experiments, effects of conversation are estimated by comparing to a common "no conversation" baseline. These baseline data were extracted from Atchley, P., and Dressel (2004), Experiment 2. For these baseline data, the average threshold for the first functional field of view task was 16ms, the average threshold for the second functional field of view task was 18ms, and the average threshold for the final functional field of view task was 51ms. These baseline data remained stable across five experiments in which the baseline data were measured, suggesting continued measurement of the baseline was not necessary. The three functional field of view tasks comprised a within-subjects repeated measure, and the effect of conversational dual-task was analyzed as a between-subjects factor.

Mixed-design ANOVA (with visual tasks as a repeated measure, conversational task presence as a between-subjects variable) revealed that the effect of conversation on visual task performance was significant, $F(1, 48) = 26.3$, $p < .001$. Overall, participants performed the three functional field of view tasks with an average baseline threshold of

28 milliseconds; when conversing, they performed the tasks with an average threshold of 89 milliseconds. This was a moderate effect ($\eta^2 = .35$). Clearly, even the demands of a rudimentary conversational task involving simple free association can affect visual attention.

As was expected by the design of the functional field of view test, there was a significant main effect of visual task complexity as well, $F(2, 96) = 55.9$, $p < .001$. The effect size for this main effect was large ($\eta^2 = .54$). Post-hoc paired comparison t-tests revealed that the simplest functional field of view task yielded the smallest thresholds (33 ms), the more complex functional field of view task yielded significantly greater thresholds (81 ms, $t(19) = 2.9$, $p < .05$), and the most complex yielded still greater thresholds (154 ms, $t(19) = 4.5$, $p < .05$).

The interaction of the effects of visual task complexity and conversational dual task was also significant, $F(2, 96) = 16.1$, $p < .001$. This moderate effect ($\eta^2 = .25$) represents the increase in amount of time needed to correctly perform the central functional field of view task with the addition of a second target, and larger increase when distractors are added, when a conversational is also being processed. That is, the demands of the conversation seem to impede the processing of a second visual target, moreso a search for a second target among distractors, thus requiring longer presentation times to complete the tasks accurately.

To determine if the semantic task had less of an impact on the functional field of view than the task used in Atchley, P., and Dressel (2004), we compared mean thresholds in this experiment (33, 81 and 154 ms for the three functional field of view tasks, respectively) to mean thresholds in Experiment 2 of Atchley, P., and Dressel (56, 159, and 202 ms for the three tasks) using t-tests. The means for all three subtests were significantly lower (t (48) = 1.7, 3.1, 1.7, p <.05 for the first, second, and third functional field of view tasks, respectively) in the current experiment, indicating the semantic conversational task was significantly less demanding than the task in which participants produced a word beginning with the last letter of the word they heard. This supports the central hypothesis that conversational demand modulates attentional reductions, and suggests accident risk may be a function of conversation type. We will re-examine this question of increased risk in the general discussion.

Pilot Experiment 2

In the following set of experiments we examine the role of emotional conversations on visual attention. Anecdotally, anyone that has ever driven while in an angered mood state can attest that it is a more difficult task than when in a calm mood state. The question, then, is whether emotional conversations reduce performance by reducing attention. One possibility is that they do, but only through a mechanism of arousal. If that is true, we would expect both positive and negative emotional conversations to produce the same effect on attention. However, if being a negative mood state is particularly problematic, consistent with previous research and anecdotal evidence, then negative conversational tasks should lead to the greatest reduction in attention. To examine these hypotheses, we presented positively and negatively valent emotional words that were equivalently arousing, and measured the effect on the functional field of view performance of the participants.

Method

Participants

Twenty undergraduate students at the University of Kansas participated for course credit. All participants reported English to be their primary language and had normal or corrected to normal visual acuity.

Materials

All materials were the same as those used in Pilot Experiment 1 with the exception of the wordlist. In this third experiment, the wordlist consisted of words that had been rated by Bradley and Lang (1999) as having the top 25% emotional valence rating (positive) (e.g., jewel, infant). The words were equivalent in arousal rating, frequency, and length with the words in Pilot Experiment 3.

Procedure

Visual tasks.

The procedures for the three visual tasks were the same as those in Experiment 1.

Conversational task.

All procedures and dependent measures from the conversational task were identical to those in Pilot Experiment 1. However, in Pilot Experiment 2, participants were asked to reply to these now “emotional” words by stating a word that began with the last letter of the given word, as was the task employed in Atchley, P., and Dressel (2004), experiments 1 and 2. It was not a requirement of the pilot experiment that the response word be “emotional”.

Results

The effect of conversation on visual task performance was once more significant, $F(1, 48) = 45.6, p < .001$. Overall, participants performed the three functional field of view tasks with an average baseline threshold of 28 milliseconds; when conversing, they performed the tasks with an average threshold of 167 milliseconds. This was a moderate to large effect ($\eta^2 = .49$).

Once more, consistent with the design of the functional field of view test, there was a significant main effect of visual task complexity as well, F (2, 96) = 40.3, $p < .001$. The effect size for this main effect was similar to that reported in Pilot Experiment 1 ($\eta^2 = .46$). Post-hoc paired comparison t-tests revealed that the simplest functional field of view task yielded the smallest thresholds (94 ms), the more complex functional field of view task yielded significantly greater thresholds (188 ms, $t(19) = 3.8$, $p < .05$), and the most complex yielded still greater thresholds (219 ms, $t(19) = 1.9$, $p < .05$).

As in Experiment 1, the interaction of the effects of visual task complexity and conversational dual task was also significant, $F(2, 96) = 17.2$, $p < .001$. As mentioned above, this moderate effect ($\eta^2 = .26$) is interpreted as the *larger* increase in amount of time needed to correctly perform the increasingly complex functional field of view tasks when a conversational dual-task is also being processed.

Pilot Experiment 3

Method

Participants

Twenty undergraduate students at the University of Kansas participated for course credit. All participants reported English to be their primary language and had normal or corrected to normal visual acuity.

Materials

All materials were the same as those used in Experiment 1 with the exception of the wordlist. In this third experiment, the wordlist consisted of words that had been rated by Bradley and Lang (1999) as having the bottom 25% emotional valence rating (negative) emotional valence rating (e.g., malaria, terrorist). The words were equivalent in arousal rating, frequency, and length with the words used in Pilot Experiment 2.

Procedure

Visual and Conversational tasks

The procedures for the three visual tasks and conversation task were the same as those in Pilot Experiment 2.

Results

The effect of conversation on visual task performance was significant, $F(1, 48) = 96.6$, $p < .001$. Overall, participants performed the three functional field of view tasks with an average baseline threshold of 28 milliseconds; when conversing, they performed the tasks with an average threshold of 207 milliseconds. This was a large effect ($\eta^2 = .67$). This large effect of negative-emotion evoking conversation is consistent with early work on the effect of emotion on visual attention (see Easterbrook, 1959).

Again, consistent with the design of the functional field of view test, there was a significant main effect of visual task complexity as well, $F(2, 96) = 66.1$, $p < .001$. The effect size for this main effect was similar to those reported in Pilot Experiments 1 and 2, and 3a ($\eta^2 = .58$). Post-hoc paired comparison t-tests revealed that the simplest functional field of view task yielded the smallest thresholds (97 ms), the more complex functional field of view task yielded significantly greater thresholds (230 ms, $t(19) = 5.2$, $p <.05$), and the most complex yielded still greater thresholds (295 ms, $t(19) = 3.1$, $p < .05$).

As in Pilot Experiments 1 and 2, the interaction of the effects of visual task complexity and conversational dual task was also significant, $F(2, 96) = 35.9$, $p < .001$. As mentioned above, this moderate effect ($\eta^2 = .43$) is interpreted as the *larger* increase in amount of time needed to correctly perform the increasingly complex functional field of view tasks when a conversational dual-task is also being processed.

Emotional Valence Analysis

A difference in effect size was observed between the effect of positive emotional conversation ($\eta^2 = .49$) and negative emotional conversation ($\eta^2 = .67$) on functional field of view thresholds. This difference can be interpreted as an ability to predict 18% more of the variability in visual attention performance when the conversation is negatively emotional than when it is positively emotional. Consistent with prior research (Easterbrook, 1959), thresholds in the third, most complex, functional field of view task were higher for the negative emotional word case (mean presentation time = 295 ms) than the positive word condition (mean presentation time = 219 ms, $t(38) = 1.7$, $p <.05$).

Summary and conclusions of pilot studies

To summarize these findings, the risk data are consistent with the analyses of the individual experiments and they succinctly show how the different conversational tasks impact attention. A good proportion of participants (75%) were able to perform the simplest task with deficit to attention. The emotional conversations yielded UFOV scores predicting the highest risk group (10% of participants), and the negative emotional conversations seemed most problematic because they left fewer participants spared of any risk (10%) and resulted in more people in the two highest risk categories (40%). That any participants would still be in a non-risk category with this conversational task is interesting and will be the subject of future research.

To conclude, these pilot experiments demonstrate, with specific regard to the UFOV measure, not only that verbal tasks may affect visual attention mechanisms, but that emotional verbal tasks may be especially detrimental to visual task performance. UFOV scores were observed to be most affected by negatively valent emotional verbal tasks. This offers compelling evidence that emotional verbal processing may be particularly distracting to drivers, and is an important area of further study.

However, dual task studies may require special consideration. There are at least four design elements that are important to consider when designing an experiment to achieve optimal, informative investigation of dual tasks costs. These design elements are discussed in the following section.

Dual Task Methodology

The nature of the ability for an individual to process emotional verbal stimuli while processing visual information, or any attentional bottleneck which may impede performance on two such tasks, is best investigated by the performance operating characteristic, or POC paradigm. An alternative to traditional dual-task methods, in which one task alone is fully prioritized, was described by Norman and Bobrow (1975), who described a method of measuring dual-task performance that takes task prioritization into consideration. This methodology describes the mathematical function of performance under various conditions of tradeoff between task prioritization.

A POC for a given pair of tasks and for a particular subject may be the function of all combinations of task performance measures (P_{T1}, P_{T2}) that arise from splitting total resources between two tasks into all possible task prioritization ratios. That is, some metric of performance for Task 1 is plotted along the X-axis, and some metric of performance for Task 2 is plotted along the Y-axis. When subjects are instructed to prioritize one task at various weights relative to the other task, a function relating the task tradeoff can be plotted.

If perfect tradeoff between tasks occurs (i.e., a one unit decrease in performance on Task 1 performance leading to a one unit increase in Task 2), the plot of this function is a straight line with slope of -1. Such a function may be an indication of a shared resource pool, or central capacity. The graph of such a tradeoff appears as in the plot in Figure 1, (see below),

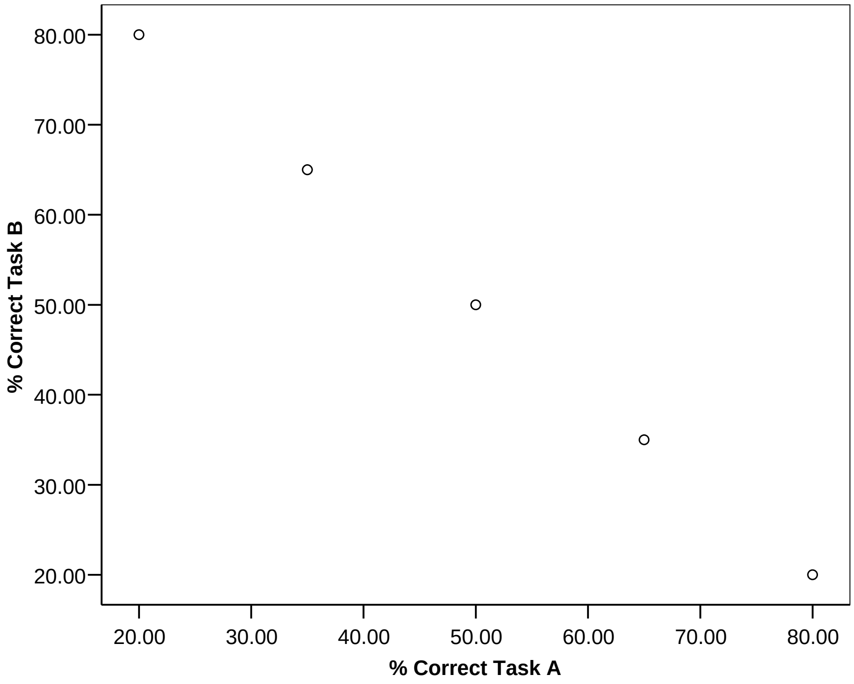


Figure 1. Hypothetical POC curve depicting tradeoff between task performance, or a “resource-limited process”.

which demonstrates a scenario in which performance improvement on one task cannot occur without an equal performance decrease on another task. This is what Norman & Bobrow (1975) refer to as a "resource-limited process".

Alternatively, it is possible that performance on Task 1 can improve without a decrease in performance on Task 2. In the extreme example, performance on a given task could remain perfect while performance on the other task could be adjusted freely in response to experimenter requests of prioritization. If performance levels of both tasks were capable of being adjusted with no related tradeoff, the plot of such a function appears as Figure 2, (see below).

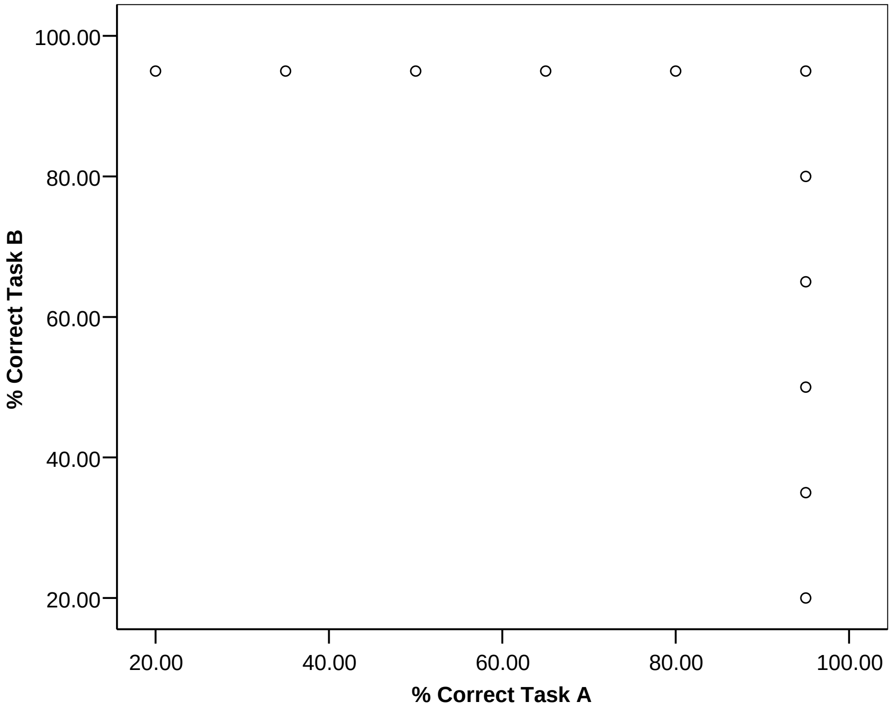


Figure 2. Hypothetical POC curve depicting perfect resource sharing between task performance.

Thus, two such tasks would be thought to share no processing resources or capacity limits, suggesting the presence of multiple attentional resource pools (e.g., Wickens, 2002).

A function that approaches a straight line would be more congruent with central capacity theories of performance (e.g., Kahneman, 1973); a function that approaches the outer bounds of the plot would be more congruent with multiple-resource pool theories of performance (e.g., Wickens, 2002). In addition to testing the nature of the resource pools available for performing two concurrent tasks, the POC curve paradigm is more informative, containing data for prioritization levels other than 100% for one or the other task (i.e., data are collected for 50%-50% sharing of prioritization of the two tasks, for 10%-90% sharing of prioritization, etc.). Thus, the most complete description of concurrent attentional performance can be accomplished via a POC methodology.

The Performance Operating Characteristic (POC) methodology is couched in a theory of general cognitive resources. Support for such a notion is offered by Navon (1984), who suggests that a general resource theory is sufficient to create complex predictions of performance patterns given different concurrent task loads. For example, Lavie (2005) has made an appeal to a general resource pool theory, terming the task demands “cognitive load”. Lavie has shown that whereas increasing *perceptual* load (e.g., decreasing signal-to-noise ratio) may decrease the likelihood that distractors are processed, thus interfering with the processing of a target, increasing *cognitive* load (e.g., introducing a concurrent memory task) appears to increase the likelihood that distractors are processed, increasing interference. Lavie suggests that this may be the overloading of a cognitive control mechanism, or central executive.

Furthermore, Lavie, et al. (2004) have found that two selective attention mechanisms exist: (1) a perceptual selection mechanism which may be employed to reduce distraction from noise-laden displays, exhausting perceptual resources such that distractors are not processed to a level that allows for interference, and (2) a cognitive control mechanism (which we claim can be affected by emotion, as motivated attention theory suggests) that reduces interference from distractors *as long as sufficient cognitive resources are available.* Lavie et al. (2004) found that under conditions involving high cognitive demand (i.e., including a secondary working memory probe task in addition to a primary selective attention task), a larger interference effect of visual distractors in the selective attention task was observed.

Similar findings have been found using cross-modal designs. Tellinghuisen and Nowak (2003), using a modification of a within-modality paradigm used by Lavie and Cox (1997) that increasing complexity of a concurrent auditory task *increases* the distraction effect of high visual load in a visual search task. Tellinghuisen and Nowak (2003) extended the cognitive load theory to involve cross-modal distractors. That is, distractors to visual targets could be compatible, neutral, or incompatible, as in the paradigm used by Lavie and Cox (1997), but instead were presented in the auditory modality. Tellinghuisen and Nowak found interference for auditory distractors in *both* high and low visual perceptual load conditions, and *greater* distraction (in contradiction to Lavie’s finding of less distraction) in high perceptual load conditions.

These resource allocation theories suggest that the additional processing demands of concurrent verbal tasks may impede processing of a visual attention task (i.e., cellular phone conversation may impede attending to relevant events in the driving environment). Processing language may be a demanding enough cognitive task that it interferes with the processing of visual information, with insufficient general resources to perform both tasks at high levels.

In addition to POC methodology, subjective reports are an important research method to employ when investigating dual-task performance. Gopher and Donchin (1986) suggest that measurement of dual-task performance can benefit from the addition of first-person subjective data. That is, the individual performing given tasks is capable of producing data regarding how the tasks “felt”, in addition to her or his performance data. If performance data suggests that two tasks were easily performed (i.e., low errors and short response times), yet the participant reports that the combination of tasks was more difficult than a combination for which the participant produced more erroneous data, something must account for this discrepancy. Research demonstrates that subjective reporting of cognitive load can be a reliable measure of task performance, yielding consistent results (Gopher & Braune, 1984; Tsang & Velaquez, 1996; Paas, et al., 2003). Drivers are largely unaware of the risks associated with cellular phone use while driving (White, Eiser, & Harris, 2004). Thus, use of subjective reports may be especially appropriate for the measurement of dual-task costs to driving performance so that not only driving performance decrements, but also drivers’ *awareness* of these decrements may be investigated.

Although the benefits of employing POC and subjective report methodologies in dual task studies are clear, they are not often included in an experimental design, especially in the applied field. The following section samples studies from the driver distraction literature, and discusses the limitations of existing experimental designs.

Review of the Cellular Phone Related Driving Accident Risk Literature

To date, research investigating dual-task interference as it relates to cellular phone use while driving can be described as belonging to one of three categories. “Correlational” (or regression) studies have attempted to demonstrate the relationship between the frequency or co-occurrence of cellular phone use with traffic accident frequency. Simulator studies (or studies simulating a roadway on a closed track) have attempted to demonstrate a decrement in driving performance (e.g., increase in collisions) while conversing on a cellular phone relative to driving performance in silence. Third, some studies have attempted to demonstrate a similar decrement in performance of tests of processes thought to be important to driving (e.g., visual attention tasks) while conversing on a cellular phone relative to performance in silence. Evidence from these three categories of research provide converging evidence that cellular phone use while driving increases accident risk, however, each category suffers distinct limitations due to experimental design issues.

Correlational studies

Correlational studies of the effect of cellular phone conversation on driving accident risk have, at times, yielded mixed results. For example, using a survey method, Violanti and Marshall (1996) obtained data from 100 drivers in New York who had been involved in accidents causing $1000 or more in damage within two years preceding their study. Also, they obtained the same data from a random sample of 100 drivers in New York who had not been in any accidents within ten years preceding their study. Their results showed that drivers who spoke on a cellular phone more than 50 minutes each month were 5.59 times as likely to have been in an accident. However, a limitation of this finding is that only 14 participants (7 percent of the sample) reported cell phone use.

Supporting this finding, Redelmeier and Tibshirani (1997) used a case-crossover design, obtaining accident reports and cellular phone billing records from 699 drivers who had recently had an auto accident. By comparing the billing record statement report of on-air phone usage with the time of collision reported in the accident report, they found that drivers had 4.3 times as many accidents during times when they were conversing on a cellular phone relative to times they were not using a cellular phone. Furthermore, accident risk was not significantly different for drivers using “hands free” phone devices and those using handheld devices (relative risks 5.9 and 3.9, respectively, actually greater for hands-free).

Contrarily, Sullman and Baas (2004) used hierarchical logistic regression to analyze a survey of 861 New Zealand drivers. They found that after six demographic variables: age, gender, mileage driven, make and model of the automobile, “centre size”, and engine size were partialled out of a predictive equation, self-reported cellular phone use (as measured by a Likert-type scale of frequency) was no longer predictive of self-reported accident involvement. However, this finding has the limitation of the possibility that after using six covariates, there may not have been sufficient statistical power to detect a meaningful effect of cellular phone conversation. Additionally, a limitation of this technique is that self-report studies may be susceptible to inaccurate or false reporting. A further limitation of this approach is that participants filling out surveys are aware they are being studied, they may (whether they are aware of it or not) alter their responses to align with their own beliefs regarding cellular phone use.

Another study suggesting cellular phone use has little to no impact on accident risk was provided by Stutts, Reinfurt, Staplin, & Rodgman (2001). This epidemiological study analyzes data from the Crashworthiness Data System (CDS) from 1995-1999. The CDS samples 5,000 police reported accidents yearly in which at least one car has been towed from the accident scene. Data are collected by "trained, professional crash investigation teams" (Stutts, et. al, 2001, p 3). These data suggest that 8.3 percent of accidents were attributable to driver distraction. This is in contrast to the National Highway Traffic Safety Administration's (NHTSA) estimate that 20 to 30 percent of accidents are attributable to driver distraction, during a comparable timeframe (Wang, Knipling, & Goodman, 1996). Specifically, the authors report that of those 8.3 percent of drivers who were distracted, 1.5 percent were distracted by using or dialing a cellular phone. Initially, this suggests the impact of cellular phone use is a miniscule effect.

However, there are problems with such an interpretation. Stutts, et. al, (2001) also report that 5.4 percent of accidents were of the category "looked, but didn't see" (p. 9), a category that seems to describe a cognitive distraction (i.e., driver distraction, in this study seems to describe overt attention, while "looked but didn't see" may include covert attention). It is possible that some drivers in the "looked but didn't see" driver attention status category were *also* conversing on a cellular phone, but this cross-tabulation is not reported. Thus, to the extent that “looked but didn’t see” drivers were also “distracted”, perhaps more than 1.5 percent of drivers were distracted by a cellular phone.

Another intriguing finding reported by Stutts, et al. (2001) is that the second-most frequently observed driver attention status was the “unknown / no driver” category, which accounted for 35.9 percent of the accidents (“attentive” was the most frequent driver attention status, accounting for 48.6 percent of accidents). Disregarding for the moment the accidents involving no driver (oddly), the “unknown” frequency counts in this driver attention status essentially deflate the percentage of “known” attention status accidents for each of the other categories. That is, greater than 8.3 percent of drivers *whose attention status was known, or who were present at the time of the accident* were distracted. Perhaps Stutts, et al. (2001) say it best when they state that “it is recognized that the CDS underestimates the role of driver inattention and distraction in crashes (p. 35).

There are inherent limitations to a correlational, epidemiological approach to the issue of driver distraction. Perhaps most importantly, correlational (or regression) studies, each of these authors note, do not imply causation. Thus, it is possible for a third variable to be the causal factor contributing to higher accident rates among those who drive while conversing on a cellular phone. For example, frequency of risk-taking behavior could feasibly contribute to individuals engaging in both cellular phone conversations *and* hazardous maneuvers while driving. If such were the case, it would not be helpful for legislative bodies to ban cellular phone use while driving. In this example, risk-taking drivers would still take risks and get into accidents, merely without talking on a cellular phone at the same time. So, while correlational or co-occurrence studies relating real-world driving accidents to cellular phone use are intriguing and suggest a causal relationship, scientific experimental studies are necessary to empirically demonstrate such a causal relationship with a random sample of potential drivers.

Simulation studies

Simulation studies of the impact of cellular phone use have resulted in more consistent results than their correlational counterparts. All simulation studies described below describe decrements in driving performance as a result of a conversational dual-task. For example, an early study by Brown, Tickner, and Simmonds (1969) used a “real” automobile on a closed course. Drivers were to complete a circuit on a course, occasionally making judgments as to whether or not a gap between temporary obstacles was wide enough for the car to drive through. Gaps ranged from 3 inches smaller than the width of the vehicle, to 3 inches wider than the vehicle, thus difficulty of the driving task was adjusted. In the event of a narrow gap, the participant was to drive around the obstacles. Performance on the driving task with no concurrent conversation was compared to driving performance while drivers responded verbally to a series of spoken questions similar to those used by Baddeley (1968). In this task, participants were to listen to an assertion regarding two letters of the alphabet followed by those two letters of the alphabet, and were then to respond as to whether the assertion was true or false (e.g., A follows B – BA – True). Drivers made significantly more errors in gap judgment (45 percent errors) and took longer to make those judgments when conversing, compared to when driving alone. Also, drivers slowed their rate of driving when conversing, and speakers slowed their rate of replying while driving (and permitted more errors than in baseline) to compensate for task demands.

Horswill and McKenna (1999) used video driving simulators to assess risk-taking judgments made by 121 university students with license to drive. In this study, participants were asked to prioritize a verbal task, which consisted of responding to each example in an auditory letter string. Letters were played via audiotape, at a pace of one letter per second, and participants were to reply “yes” if the letter was ‘K’, and “no” if it were any other. Results showed decrements in driving related risk perceptions such as accepting a size of gap in cross-traffic in which to pull out, reacting to hazards, and following distance, to the magnitude of an average z-score of .4, or nearly half a standard deviation.

Furthermore, decrements in driving performance measures while participants were engaged in a concurrent conversational task while driving a simulated vehicle relative to their performance in silence were demonstrated by Horberry, Anderson, Regan, Triggs, and Brown (2006). In the conversation condition, participants were to respond to general knowledge questions by choosing an answer from two verbally presented alternatives. While engaged in conversation, participants displayed greater amounts of risky driving relative to when they were not engaged in conversation. These risky driving behaviors included greater deviation from posted speed limits, and less reduction in speed in response to hazards such as a pedestrian standing in the roadway, a car backing into the roadway from a driveway, and a pedestrian crossing the road. Furthermore, subjective reports of perceived workload were congruent with behavioral data: participants described sensing greater mental workload in dual-task conditions than in single-task conditions. Also, participants correctly described still greater mental workload (i.e., correctly related to reduced performance) during another secondary task which yielded the greatest magnitude of dual-task costs, a manual task involving manipulating an in-vehicle information and entertainment system.

Jamson, Westerman, Hockey, and Carsten (2004) demonstrated that such interference effects are not limited to cellular phones, but to increasingly common in-vehicle computers. In this study, a computerized voice (based on Microsoft's .NET Speech Software Development Kit Version 1.0) "read aloud" statements based on Baddeley's (1968) grammatical reasoning task (similar to the task discussed above used by Brown, et al. (1969)). These statements occurred in one of two schedules: at the participant's request (i.e., at a time when the roadway required low attentional demand) or at a predetermined pace, cued by a tone. Participants were to prioritize the driving task. With this design, rather than a block design, this study involved silence and conversation within one experimental session. Results demonstrated a decrement in performance when pacing of the e-mail delivery was not participant controlled relative to when delivery was participant controlled. Participants made more errors in judging the statements made in the e-mails, and were slower to reply when e-mails were delivered at a predetermined pace. Intriguingly, the section of the intermittent cellular phone experimental block during which a message was not being played (nor manipulated in any way) also yielded decrease performance in driving tasks. That is, merely *anticipating* a message (or to generalize, a cellular phone call) may interfere with driving.

Furthermore, Jamson, et al. (2004) showed decrements in driving related measures as well. For example, participants were approximately 800 milliseconds slower to “anticipate” a lead car braking in response to visible traffic cues such as a red traffic light (i.e., participants responded to the environment and/or the lead car braking by applying the brakes in the simulator vehicle) when a predetermined e-mail was read than when no e-mail was being read, and approximately 400 milliseconds slower to anticipate a lead car braking when a participant-determined e-mail was read. It should be noted that simulated traffic and road conditions may not have been the same in these two conditions; nevertheless, significant and meaningful dual-task costs were observed.

Converging evidence for increased reaction times to other vehicles braking is provided by Strayer, Drews, and Johnston (2003). This study required participants to drive through a realistic and multifaceted simulation course either in silence or while engaging in casual conversation. Conversations consisted of discussions of topics that had been reported to be of interest by each individual participant. The conversational dual-tasks were found to increase reaction time to a lead car braking, an effect that interacted with increased traffic density (in another lane of traffic) to create still greater reaction times.

In another experiment, Strayer et al. (2003) found that recognition memory for billboards that had appeared along the simulated driving route was greater when the billboards were encountered under single-task conditions than when they had been encountered under dual-task, conversation conditions. This improvement in memory was not better accounted for by gaze fixation differences. Using an eyetracker, Strayer, et al. (2003) found the conditional probability of recognizing a billboard given that it had been fixated (i.e., looked directly at) in silence was double that of billboards fixated while conversing.

Later, Strayer and Drews (2004) used similar methodology to describe other deficits in driving performance that can be attributed to concurrent cellular phone use, and to compare these deficits between younger and older drivers. Using a similar task as Strayer et al. (2003), conversations consisted of casual, topical conversations of interest to the participants. No differences in the effect of cellular phone conversation on driving performance measures between younger and older drivers were found. However, statistically significant and large decrements in performance measures such as braking reaction time, speed, following distance, and recovery of speed following braking were found for all participants.

In a recent study, Strayer, Drews, and Crouch (2004) compared the deleterious effects of conversing on a cellular phone while driving to the effects of driving legally intoxicated. Using a driving simulator, and help from the Utah Highway Patrol, one group of subjects served as a baseline of driving performance, a second group of subjects performed a conversational task similar to those described above, and another group drank a vodka and orange juice mixture in calculated doses to reach a blood alcohol concentration level of .08. The data suggest that the driver conversing on a cellular phone drove in a more impaired manner than the legally intoxicated driver. While intoxicated drivers drove more aggressively than baseline drivers, the only significant difference between baseline and intoxicated drivers was the amount of maximum brake force applied when braking (intoxicated drivers used greater force). Conversing drivers, contrarily, were 70 milliseconds slower to respond to a lead car braking than were the baseline drivers (intoxicated drivers were actually faster to respond than baseline drivers, suggesting an increased anticipation of braking). Also, conversing drivers demonstrated

2.3 seconds greater standard deviation in their following distance than baseline drivers. It is compelling that drivers conversing on a cellular phone (for which there are currently no complete bans) drive more dangerously than drivers who are legally intoxicated (for which all states have some legislation to reduce).

While Strayer and his colleagues demonstrate reliable, meaningful decrements in a number of driving performance measures, the effects of prioritization remain unknown, as well as the relative interference caused by different conversational demands.

However, the question of whether different conversational demands yield different dual-task costs has been investigated by others. Some researchers have included variations of the complexity of a conversational task. For example, McKnight and McKnight (1993) used a video recording of driving vignettes as a form of driving simulator. Participants were provided with accelerator and brake pedals and a steering wheel, and were instructed to respond to instances in the video that required adjustment of speed or direction to avoid an accident, that is an evasive action. A variety of dual-tasks were implemented, including tuning a radio to match a signal radio, "casual conversation" regarding demographic data and pastimes, and "intense conversations" regarding mathematical problems and short-term memory tests of digits. While likely atypical and artificial, these conversational tasks serve well to measure interference effects resulting from conversations of different levels of complexity, or cognitive effort. Response time to these questions was purposely not measured, but the proportion of critical events requiring evasive action that were missed by participants was 7 to 10 percent greater in all distraction conditions, with intense conversations yielding significantly more missed events than casual conversations.

Further evidence of the effect of conversational complexity on the magnitude of dual-task costs to driving measures is described by Shinar, Tractinsky, and Compton (2005). Ten young drivers with less than 6 months driving experience, ten experienced drivers with 8 to 15 years driving experience, and ten older drivers with an average of 35 years of driving experience participated in the study. All participants were to drive a simulated vehicle through a course in either a moderate speed (50 mph), fast speed (65 mph), or lead car (varying between 55 an 65 mph) condition for five experimental sessions, in which effects of practice were to be measured. Participants drove under three conditions: (1) silence, (2) an arithmetic questioning conversation, and (3) an emotionally involving conversation. The arithmetic conversation consisted of sequences of digits and operations were spoken before a prompt for the solution. The emotional conversation consisted of participants being challenged about topics reported to be of interest to themselves (e.g., a fan of a sports team was told that said team wasn't performing well, and made to rebut).

A significant effect of conversational complexity was observed. Greater decrements in driving performance measures such as driving slower than the appropriate speed and driving with sporadic speed fluctuations demonstrated during the difficult, arithmetic conversations than during the less complex, emotional conversations. However, other driving performance measures such as steering deviations and reaction time to peripherally presented signals yielded no effects of conversational complexity. For this reason, Shinar, et al. (2005) suggest that conversations involving arithmetic, which they contend are commonly used in experimental paradigms, may overestimate conversational interference.

Contrarily, Lee, Caven, Haake, & Brown (2001) provided evidence that dual-task costs from conversations may *not* be affected by conversational complexity. Specifically, Lee, et al. demonstrated the distraction a driver might face when interacting with a speech-based e-mail, “reading” system. Participants were to operate both a simple and complex e-mail delivery system. The simple system required decisions between two options at each of three levels of operation menus; the complex system required decisions between four to seven options at the three levels of operation menus. After practicing driving the simulator, participants followed a lead car, which would remain 1.8 seconds ahead of the driver, programmed to travel approximately 40 to 45 miles/h. This lead car would brake at random intervals (with given deceleration), and the participant was to respond by braking as soon as possible. Subjects were implicitly told to prioritize the driving task. The e-mail navigation dual-task yielded significantly longer reaction times to the lead car braking, by an average of 300 milliseconds. Furthermore, this study demonstrated additive effects of road complexity and the use of the e-mail system, yet failed to demonstrate an effect of the complexity of the menu systems in the e-mail task. However, a self-report scale of perceived distraction showed participants felt more distracted when using the complex system than the simple system, indicating a dissociation between perceived performance and behavioral data.

Evidence for the lack of an effect of conversational complexity on dual-task costs to driving measures is also provided by Rakauskas, Gugerty, and Ward (2004). In this study, 24 young adult licensed drivers drove through a simulated circuit track. This track included parked car distractors, and hazardous events, such as a parked car pulling out in front of the driver or an oncoming car swerving in front of the driver programmed to occur at given distances from the driver. In addition to driving during a conversation-free baseline, participants drove while responding to questions that had been rated during pilot testing as easy or difficult. None of these questions were thought to be visual/spatial in nature. Easy questions included demographic questions as well as questions about participants' real-life schedules. Difficult questions included philosophical and moral dilemmas. Additionally, participants rated their perceived mental effort during the experiment as a whole via a single-dimension Likert-type scale.

Small but statistically significant differences in driving performance were observed. Participants drove more slowly (1 mph), and displayed greater variability in their speed (.2 mph) when engaged in conversation relative to when they drove in silence. However, there was no effect of conversational complexity (i.e., question difficulty) on any of seven driving performance measures. Intriguingly, subjects reported an increased mental workload when conversing while driving relative to driving alone, but did not report a difference in workload between levels conversational complexity, mirroring their own behavioral data.

Of course, findings such as Lee, et al. (2001) and Rakauskas, et al. (2004) do not falsify the hypothesis that conversational complexity can affect the magnitude of dual-task interference, they merely suggest that some perceived differences in conversational complexity may not be sufficient to produce differences in driving performance measures. Furthermore, Shinar et al.'s (2005) contention that arithmetic tasks should be avoided in experiments suggests further reason to consider conversational task demands. Thus, researchers should investigate carefully potential conversational tasks to determine aspects of conversation that affect dual-task costs with specific driving performance measures.

To summarize, simulator studies demonstrate the robust nature of dual-task interference as a result of conversing while driving. These studies demonstrate this effect in a causal manner. That is, samples of individuals who are randomly placed into a condition of driving while conversing show reliable decrements in a host of driving performance measures relative to individuals randomly placed into a condition of driving in silence. This decrement in performance occurs under identical driving conditions, other extensive controls, and often within-subject. Applied science approaches might consider these findings of causality to be pragmatic and useful. However, basic science approaches take interest in a question that is left unanswered by simulator studies: the identification of specific cognitive mechanisms that can account for these observed performance decrements.

Studies of cognitive mechanisms used in driving

While basic science approaches have produced a great number of studies investigating cross-modal attention effects, a manageable number of studies have framed cross-modal attention specifically within a driving context. For example, an effect of conversation on simple reaction time is investigated in a study conducted by Consiglio, Driscoll, Witte, and Berg (2003). In this study, simple response time to the illumination of a red lamp (appropriately sized and shaped to reflect the visual angle to be obtended by an automobile’s brake lamp 12 m in front of a driver) situated in front of the participant was measured under several distraction conditions. Participants responded by releasing a mock accelerator and depressing a mock brake pedal in each of five conditions: (1) silence, (2) passive listening to music played on a radio, (3) conversing with a "passenger" seated to the right of the participant in a casual manner "as one might have with a new acquaintance" (p 496), (4) a similar conversation with the same individual via hands-held phone, and (5) a similar conversation with the same individual via hands-free phone. Orders of these conditions were presented in counterbalanced order. Results suggested no difference in response time between silence and passive listening conditions, but significant increase in response time relative to these conditions for each of the three conversation conditions, by an average of 60 milliseconds. No differences between conversation conditions were observed. The study makes no direct citation of the accident risk increase associated with a 60 ms increase in response time; however, the notion that simple detection of the onset of a fixed-location target can be impaired by casual conversation is compelling.

Another mechanism capable of accounting for conversation's negative effect on driving performance is that of planning eye fixations. Recarte and Nunes (2000) discussed differences in scanning and fixating on items in the driving environment under conditions of various conversations and a silent baseline. In an on-road, "real-life" automobile, an eyetracker recorded gaze positions and fixation locations and durations. Participants were to drive a planned route while: (1) not conversing, (2) replying to a given letter with a list of words beginning with that letter for 30 seconds per letter, or (3) replying to a given letter by answering an imagery-related question regarding the letter's shape. Imagery-related questions consisted of answering whether the given letter remained the same if "flipped" across a horizontal, or vertical, axis, and whether the letter was “closed” (i.e., contained a “circular” area, as in the letter ‘b’ or the number ‘4’) or “open” (i.e., did not contain such an area, as in the letter ‘s’ or the number ‘7’).

Eyetracking data showed that verbally creating word lists decreased fixation times relative to baseline, while verbal imagery tasks increased fixation times relative to baseline. Also, during baseline, participants exhibited large “visual inspection windows”, or regions in which gaze fixations were likely to occur. This inspection window was significantly smaller while verbally creating word lists, and smallest while performing verbal imagery tasks. Furthermore, participants fixated on mirrors and the speedometer less frequently when verbally creating word lists than during baseline, and least frequently when performing verbal imagery tasks (under most driving conditions).

In addition to cellular phone induced differences in fixation patterns, reduced efficacy of visual attention may increase accident risk. Supporting this notion, Amado and Ulupinar (2005) posited that decrements in performance on attentional measures could account for higher accident rates of drivers conversing on a cellular phone. Forty-eight undergraduate students demonstrated decreased performance on two attentional tasks, the “Cognitrone” task and the “Peripheral Detection and Dual-processing Task (PDDpT)”, when conversing relative to performance in silence. The Cognitrone task consisted of a computer-based test that uses a variation of a match-to-sample task, combining error rates and reaction time to create a composite score. The PDDpT task consisted of a driving simulator in which drivers were instructed to avoid other cars on a straight path, and respond to illumination of an LED panel on each side of the simulator (at an unreported degree of eccentricity). Participants were to make a speeded response when either or both LED panels were illuminated, while at the same time driving the simulated vehicle. In addition to a conversation-free baseline, the conversational dual-task demands were administered at two levels of complexity. These conversational tasks consisted of general knowledge questions and arithmetic problems divided into two levels of difficulty. Questions were either read aloud next to participants or played on a CD player. Amado and Ulupinar found reliable decrements in both the Cognitrone and PDDpT attentional measures. No effect of question delivery mechanism (i.e., reading aloud vs. CD player) was found. However, there was an effect of question difficulty. Participants yielded poorer, slower performance on both attentional measures when complex questions were being asked relative to when simple questions were being asked. Also, conversing drivers were 2 to 4 times as likely to collide with a median (center line) as drivers who were not conversing. Thus, attentional mechanisms may suffer from reliably poorer performance when a driver is conversing.

Further evidence for the hypothesis that visual attentional mechanisms may suffer as a result of concurrent conversation was provided by Atchley, P., and Dressel (2004). In this study, participants' Useful Field of View (UFOV), or functional field of view (Sanders, 1970), was measured in both dual-task and single-task conditions. The UFOV has been established as having a relationship with older drivers' accident risk was shown to decline in dual-task conditions relative to single-task conditions (Goode, et al. 1998). Conversations consisted of, in two experiments, simple and memory load imposing single word responses to single given words. When engaged in conversation, participants yielded thresholds to detect targets many times greater relative to thresholds when no conversational dual-task was required. Thresholds of the magnitude exhibited by some participants might predict them to be as much as 16.3 times more likely to be in an injurious accident as a result of concurrent conversation.

Evidence that another aspect of visual attention can be negatively affected by conversation is offered by McCarley, Vais, Pringle, Kramer, Irwin, and Strayer (2004). In two experiments, these researchers demonstrated that performance in a change detection task was reduced under the demands of conversation, but not under demands of attentive listening. Change detection involves noticing, in a flicker display paradigm (Rensink, O'Regan, & Clark, 1997), the change in an element of the scene (e.g., the vanishing and reappearance of an object). This measure is thought to tap endogenous attention, and is often a difficult task; the time it takes to notice the change in the scene (or the failure to notice it altogether) has been termed "change blindness" by Simons and Levin (1997). Participants, especially a group of older participants, made more errors in change detection when engaged in a topical conversation regarding television shows and hobbies, relative to a single-task condition.

In addition to effects related to eyetracking and specific attentional mechanisms, studies have demonstrated that performance on a continuous attentional task not unlike simulated driving may be negatively impacted by concurrent conversation, a pursuit tracking task. In a pursuit tracking task, participants must use some manual interface to control an on-screen target object such that it remains within specified (moving) boundaries. Briem and Hedman (1995) used a pursuit tracking task using an accelerator pedal and steering wheel. Conversations took two forms: simple and complex. During "simple" conversations, participants engaged in two minute conversations regarding topics such as the war in Bosnia and the unemployment situation in Sweden. During complex conversations, participants were to first indicate whether each given sentence was *logical* or *illogical* (e.g., “The driver ate a car” is illogical), and then repeat to the experimenter the first words of the past four sentences, thus requiring decisions to be made while imposing a memory load. Briem and Hedman demonstrated that participants emitted both more speed violation errors and greater position deviation errors in a pursuit tracking task when conversing relative to when they were not. Interference was observed during both concurrent conversational tasks, however greater interference on many measures was observed during more complex conversations.

Strayer and Johnston (2001) used a similar pursuit tracking task to assess the dual-task cost of concurrent conversation. While participants conversed about topics such as the impeachment of President Clinton and other (then) current events, participants made greater tracking error on difficult “courses”. Control of the “car” cursor in this study was performed with a joystick. Additionally, participants were to click a joystick button in response to the target (i.e., to-be-tracked) object changing color to red. Participants were more likely to miss a “red light”, and were slower to respond, when conversing relative to a single-task condition. This pattern of data was not observed, it should be noted, when participants listened to a radio broadcast of their choosing.

In conclusion, pursuit tracking studies indicate that drivers may likewise regulate speed less effectively and maneuver an automobile with less precision while conversing on a cellular phone. Such parallels in driving performance seem intuitively related to greater accident risk. Furthermore, drivers may scan the roadway less effectively and completely while conversing. Compounding this problem, drivers may process visual information within these diminished scanning patterns less effectively while conversing.

Thus, several cognitive mechanisms have been identified that can offer specific accounts for the impaired driving performance demonstrated explicitly by simulator studies, and suggested by correlational studies.

Discussion of Driver Distraction Literature Review

In this sample of 17 (not counting correlational) studies investigating the effect of cellular phone conversations on driving accident risk, designs are less than optimal. Most notably, none of the studies discussed in this review utilized the powerful Performance Operating Characteristic paradigm discussed above. Thus, it remains unknown what function describes a driver's ability to "trade-off" performance from the driving task to the conversation task. Many studies utilize a traditional dual-task method, assigning prioritization schedules for the participants to adhere to. Without explicitly measuring priority tradeoff, a researcher cannot be certain that prioritization is occurring in the intended manner. That is, while driving was (often) to be prioritized in these studies, the lack of potential harm, and lack of explicit instructions to "de-prioritize" conversational tasks may enable the participant to also prioritize conversation. Also, traditional dual-task methods leave the possibility decrements in driving (or driving related) performance only exist at the assigned levels of prioritization. That is, while the current evidence suggests that talking drivers are prone to miss hazards, for example, it could be that a real-world driver would prioritize the "life or death situation" of driving an automobile *higher* than the artificial prioritization of driving a "safe" simulator. By prioritizing real-world driving at a higher level (or conversation at a lower level), perhaps decrements in performance would not be seen. Correlational studies mentioned above suggest this prioritization explanation is not the case; POC studies could scientifically demonstrate that this is not the case.

Second, eight studies (nearly half) fail to report performance data for both “driving” and conversational tasks. As mentioned above, such reporting is necessary to insure that the secondary task has its intended effect. It could be the case, in these studies, that participants merely “tuned out” the conversation, and provided incoherent or unrelated responses to their conversational prompts. While conversations that can be scored tend to deviate from typical conversational material and format (e.g., mathematical problem solving or recalling the first word of each sentence), some aspect of the conversation should be measured. Doing so means that experimenters can measure whether participants are paying thoughtful attention and planning appropriate responses to the conversational prompts. Efforts to do so while preserving ecological validity of the conversation task will be most fruitful. As discussed above, identifying such conversational tasks is an important area for future research.

Finally, only four studies utilized a self-report measure of subjective task demands. Interestingly, of these four studies, two reported congruent patterns of results for subjective ratings and behavioral data (Rakauskas, et al., 2004; and Horberry, et al., 2006), and two reported incongruent patterns of results (Lee, et al., 2001; and Shinar, et al., 2005). That is, a limited number of studies demonstrate both that participants are aware of the difficulty of dual-tasks and the decrement in performance they impart, and that participants are unaware of the difficulty of dual-tasks and are unaware of any decrement in performance.

Clearly, more research is needed to address the issue of the utility of self-reports in dual task studies. Understanding the relationship between subjective and behavioral data is of urgent importance in this applied setting. In a study investigating drivers' ratings of the danger involved in doing various tasks while driving, White, Eiser, and Harris (2004) found that making and receiving hands-free cell phone calls were rated as relatively harmless, falling between sneezing and smoking in rankings (Table I, p. 326). Thus, it seems that drivers may underestimate the interfering effects of concurrent cellular phone conversation while driving; any understanding as to how these errors are made may be used to help calibrate these judgments.

To summarize, studies investigating the impact of cellular phone conversations on driving performance can be grouped into three categories: correlational studies, simulator studies, and studies of specific cognitive mechanisms. Simulator studies provide the causal relationship that correlational studies lack at the possible expense of some ecological validity. Studies of specific cognitive mechanisms used in driving provide more specific explanations for driving performance decrement than simulator studies can provide, yet may sacrifice still more ecological validity. Together, these studies converge to report reliable, valid findings that cellular phone use while driving increases accident risk. Table 2 summarizes the effects of driver distraction reported by each study.

However, existing studies may leave gaps in our knowledge of this specific dual-task due to methodological limitations. Future studies would benefit from (1) measuring performance on both tasks, to insure (or at least to measure) any task trade-offs, (2) using a Performance Operating Characteristic paradigm to systematically induce these trade-offs, and to find the function defining how drivers may feasibly trade-off between tasks of driving and conversing (3) using multiple levels of complexity of both tasks, to measure not just the raw outcome of difficulty manipulations, but also how the *processing* of the two tasks may change as a result of difficulty manipulations and (4) measuring subjective ratings of cognitive load or perceived task difficulty, particularly to identify any dissociation between actual and perceived performance. The prevalence of these four design elements in the literature reviewed above is displayed in Table 1.

Thus, careful consideration must be given to particular design elements when researching dual tasks. The four design elements discussed above have been employed in the current investigation into the distracting effect of emotional verbal tasks on visual attention performance.

Dissertation Experiment

Method

Subjects

One hundred sixteen participants were recruited from Introductory Psychology and Research Methodology classes at the University of Kansas. Thirty participants did not complete the full experiment due to misunderstanding directions, resulting in unusable data. Three participants were not included in analyses because of equipment failure (battery expiration in devices). Two participants were not included because they did not demonstrate 20/20 vision for both eyes in a prescreening acuity test. One participant did not complete the study, reporting that the given words were not audible. Data obtained from the remaining 80 participants are reported below. Mean age of these participants at the time of their participation was 20.5 years.

Materials

A computerized version of the UFOV visual attention test (Visual Awareness, Inc., Chicago, IL) was implemented via EloTouch touch-screen monitor. This test consists of three subtests: (1) central target discrimination, (2) central target discrimination and simultaneous peripheral target localization, and (3) task two, above, in the presence of a field of triangle distracters, (see Figure 3, below, for an example of the third UFOV task stimulus screen).

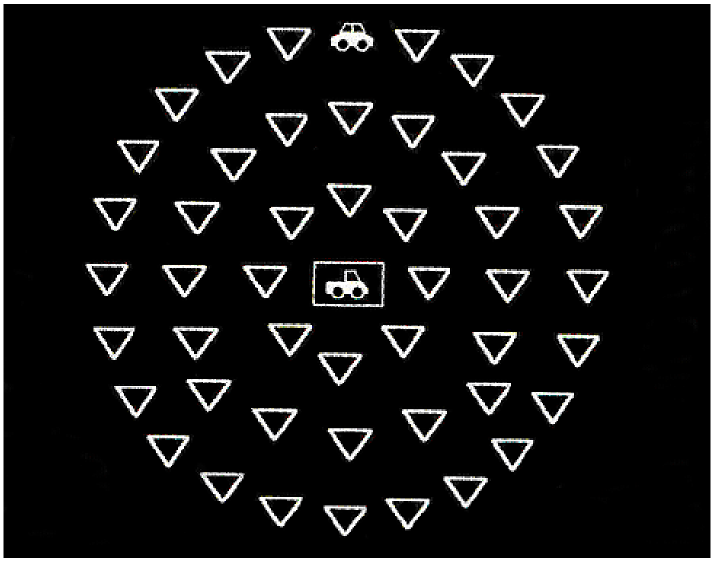
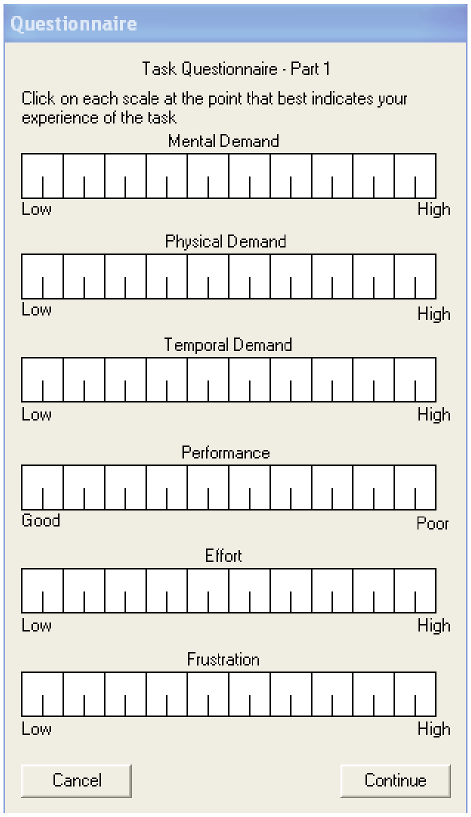


Figure 3. Example of the third UFOV task stimulus screen.

Dependent measures for the UFOV task are the presentation time threshold, determined via adaptive staircase design, for participants to detect or discriminate the single target (in subtest 1) or both targets (in subtests 2 and 3) correctly 75% of the time. Three UFOV scores are reported, one for each subtest, however, a composite of these three scores determine a categorical odds-ratio prediction for driving accident risk.

Verbal tasks consisted of words read one at a time, alternating stimulus presentation (i.e., the experimenter reading the word) and response (i.e., the subject speaking his or her single word response) via E-prime software. Words were presented at a constant rate with interstimulus-interval (ISI) of four seconds, a rate that pilot data reveals to be a sufficient time to allow subjects to generate a response. Four word lists have been created, manipulating emotional valence and arousal ratings in a factorial design: (1) positive valence / high arousal, (2) positive valence / low arousal, and (3) negative valence / high arousal, and (4) negative valence / low arousal. All lists are controlled for word frequency (mean frequency = 38.7 appearances per 100,000 words, Kucera & Francis, 1967). Valence and arousal levels of the stimulus words were manipulated using the Affective Norms for English Words (ANEW) Database (Bradley & Lang, 1999), which includes ratings of 1034 words on a 9-point Likert-type scale. Valence has been manipulated such that words on the positive valence list have ratings greater than 6 on a 9-point scale, and words on the negative valence list have ratings less than 4. Arousal has been manipulated such that words on the high arousal list have ratings greater than 6 on a 9-point scale, and words on the low arousal list have ratings less than 6. Dependent measures for the verbal task include latency to respond, measured by a voice key and E-prime software, and the sum of error rates for errors of commission (i.e., words that are rated as being unrelated to the given word), and errors of omission (i.e., a failure to respond to the given word in the four second response allowance time).

A computerized version of the NASA Task Load Index (TLX) was administered immediately following each of the three experimental (dual-task) phase, that is, after each UFOV subtest. Hart and Staveland (1988) demonstrated the utility of the NASA TLX in measuring six workload related factors to create a reliable and useful subjective rating of cognitive load. The TLX measures subjective workload in two stages. First, participants report the strength of each of six workload dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration on a 20 point Likert-type scale, ranging from "low" to "high", (see Figure 4, below).



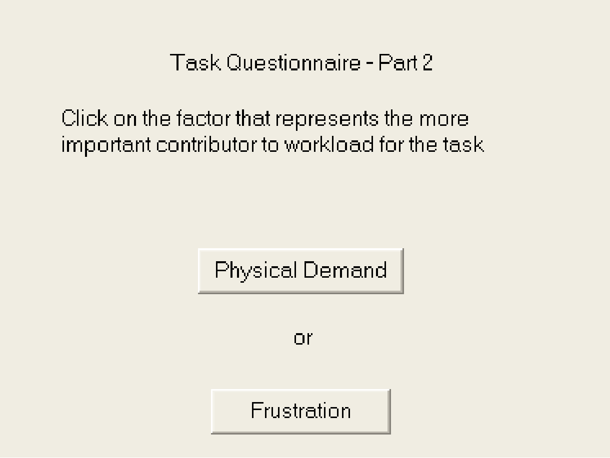
Questionnaire

Task Questionnaire - Part 1

Click on each scale at the point that best indicates your experience of the task

Figure 4. Screenshot of the first stage of TLX subjective reporting.

Next, participants respond to 15 two-alternative-forced-choice options, selecting which of two factors was a greater influence on workload, (see Figure 5, below).



Task Questionnaire - Part 2

Click on the factor that represents the more important contributor to workload for the task

Figure 5. Screenshot of the second stage of TLX subjective reporting.

These stages combine to create an overall workload score and scores describing the impact of each of the six dimensions on this overall workload score.

Experimental Design and Procedure

Subjects were randomly assigned to one of four experimental conditions, in a 3 (prioritization schedule) x 4 (emotion condition) mixed factorial design: subjects were instructed to prioritize the visual attention task (i.e., they were told they could “skip” as many responses to the verbal task as they wished, including all of them, or they could make “unrelated” or repeated responses), as 50% of their dual-task performance goal (i.e., they were told to try to perform both tasks as well as they could), or to focus on the verbal task (i.e., they were told to only pay enough attention to the UFOV tests to advance through the response screens, and that they “didn’t have to try to be correct”), a within-subjects measure. Also, subjects heard words from one of the 4 emotion condition lists (high arousal negative valence, low arousal negative valence, high arousal positive valence, and low arousal positive valence), a between-subjects measure. Thus, each cell in the factorial design included data from 20 subjects.

Subjects were first screened for visual acuity; only data from subjects with normal or corrected-to-normal vision were collected for analysis. Subjects were then informed of the experimental procedures (above). Subjects were asked if they comprehend the prioritization schedule and experimental tasks. A minimum of four practice trials of each UFOV subtest were administered before the tests began; subjects were allowed to repeat blocks of four practice trials until they reported an understanding of the subtest. The concurrent UFOV tasks and verbal task were administered; verbal responses were recorded with a digital voice recorder into digital sound files for later analysis of error rate (see below). Subjects were asked to respond to the words they heard by stating the first meaningfully related word that comes to their mind, though not necessarily a true synonym (e.g., “chair” in response to the given word “desk”). Responses were be judged by two research assistants on a 7-point Likert-type scale (1 = not at all related in meaning, 7 = very closely related in meaning).

Following each stage of the UFOV task during the experiment proper, the NASA TLX was administered, measuring subjects’ subjective reports for the *dual-task* workload. That is, rather than assess the visual and conversational tasks independently, the dual-task itself was assessed.

Hypotheses

Anticipated Dual Task Cost

While much of the current research is exploratory, there are reasons to predict a specific pattern of results. First, given the wealth of dual-task interference effects described in basic and applied science literatures, specifically prior research using similar methods to those proposed (e.g., Atchley, P., & Dressel, 2004), we hypothesize a main effect of dual-task demand. That is, we expect to observe greater UFOV thresholds in the visual task and both longer response latencies and a greater number of errors in the verbal task under dual-task conditions, overall, than in "single task" (i.e., focused attention) conditions. This hypothesis is additionally supported by theories, mentioned above, which view the UFOV task as not merely a low-level perceptual task, but as a higher level cognitive task (e.g., Sekuler, Bennett, and Mamelak, 2000). If the UFOV task involves executive cognitive control processes, then it may more readily be interfered with by a cognitively demanding task such as proto-conversation (e.g., Lavie et al., 2004; Tellinghuisen & Nowak, 2003). It is anticipated that both visual attention and verbal task performance will decline in the presence of dual-task demands relative to focused attention.

Anticipated POC Curves

Additionally, it is hypothesized that the *shape* of the POC curves (i.e., the ability for subjects to trade off prioritization) will be dependent on the emotional valence and arousal of the stimuli words. As discussed above, the interfering effects of negative emotional valent stimuli have been theoretically, neurophysiologically, and empirically (i.e., behaviorally) supported. Furthermore, specific directional hypotheses can be expected from the findings of Atchley, R., et al. (in press), and Robinson, et al. (2004). It is expected that the average POC curve exhibited by subjects in the high arousal negatively valent emotion condition will show greatest linearity (i.e., more consistent with a resource limitation) than the average POC curve exhibited by subjects in the high arousal positively valent emotion condition (i.e., suggesting availability of excess resources), which is expected to show greatest curvature (i.e., worst fit to linearity). Between these two poles are expected to lie the POC curves for the low arousal positively valent condition (expected to yield the second best fit to linearity), and for the low arousal negatively valent condition (expected to yield the third best fit to linearity). Because of the relative salience of these stimuli, and how these emotional conditions map to a theory of motivated attention, this pattern of results is to be expected.

Anticipated Relationship of Task Performance and Subjective Awareness

Finally, subjective reports are hypothesized to be indicative of overall dual-task performance, though to a moderate extent (i.e., with substantial error). That is, positive, yet weak correlations are expected to be obtained between NASA TLX ratings of task difficulty and task performance. Specific relations between subjective reports and performance on single tasks are uncertain. That is, we do not have reason to firmly hypothesize the relative impacts of task prioritization, arousal and emotional valence on subjective ratings. However, anecdotal evidence from informally questioning pilot subjects suggests that subjects are aware of the increasing difficulty of the three UFOV tasks, which suggests that there may be a relationship between subjective reports and UFOV task performance. Contrarily, drivers report a lack of awareness of the risks of cellular phone conversations while driving (White, Eiser, & Harris, 2004). These conflicting findings make a more systematic assessment of the relationship between subjective awareness of task demands and dual-task performance a necessary and important venture.

Dissertation Experiment Results

Useful Field of View Data

As could be expected from the intended design of the UFOV test, there was an effect of UFOV subtest type (i.e., difficulty) on UFOV threshold scores. Repeated-measures ANOVA revealed that this was a significant effect, $F(2, 158) = 150.2$, $p < .01$, $\eta^2 = .66$. Furthermore, post-hoc t-tests revealed that the first UFOV subtest yielded the lowest thresholds ($M = 74$ ms), the second UFOV subtest yielded larger thresholds ($M = 106$ ms), and the third UFOV subtest yielded the largest thresholds ($M = 143$ ms); each Bonferroni adjusted pairwise comparison showed these differences to be significant ($p < .01$).

Useful Field of View thresholds were also significantly affected by attentional prioritization, or how much effort was devoted to performing the UFOV task by the participants, $F(2, 158) = 82.9$, $p < .01$, $\eta^2 = .51$. Bonferroni adjusted pairwise comparisons revealed that lowest mean UFOV thresholds (i.e., the average of the three subtest thresholds) were exhibited when participants focused attention on the UFOV task ($M = 35$ ms), larger thresholds were exhibited when participants devoted “equal” attention to the UFOV and verbal tasks ($M = 64$ ms), and the largest thresholds were exhibited, not surprisingly, when participants “ignored” the UFOV task (i.e., focused attention on verbal task performance) ($M = 223$ ms); each pairwise comparison was again significant ($p < .05$).

Additionally, there was a significant interaction of the effects of attention allocation and UFOV subtest difficulty. When focusing on the UFOV task, participants exhibited mean UFOV thresholds of 17, 22, and 68 milliseconds for the first, second, and third UFOV subtests, respectively (all comparisons significant, $p < .05$). The effect of UFOV difficulty under the "focused" condition was significant, $F(2, 158) = 92.4$, $p < .01$, $\eta^2 = .54$, (see Figure 6, below).

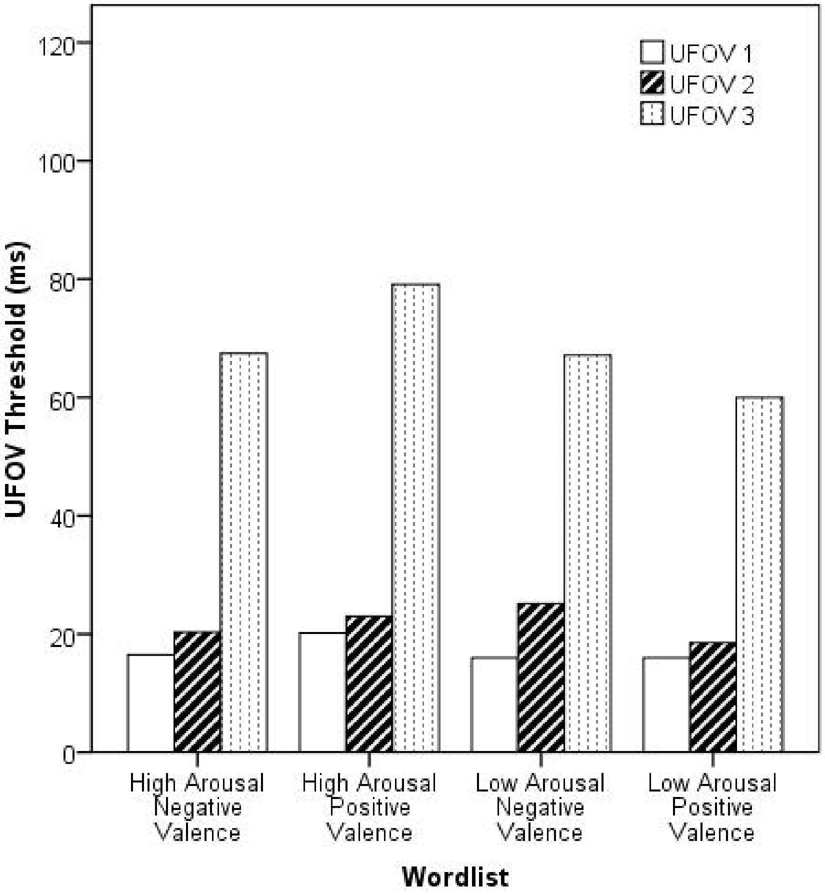


Figure 6. Mean UFOV thresholds observed under the "Focus on the UFOV" prioritization condition.

When attending to the UFOV and verbal task "equally", participants exhibited mean UFOV thresholds of 35, 58, and 97 milliseconds for the first, second, and third

UFOV subtests, respectively (again, all comparisons significant, p <.05). The effect of UFOV difficulty under the "equal emphasis" condition was significant, F(2, 158) = 82.3, $p < .01$, $\eta^2 = .51$, (see Figure 7, below).

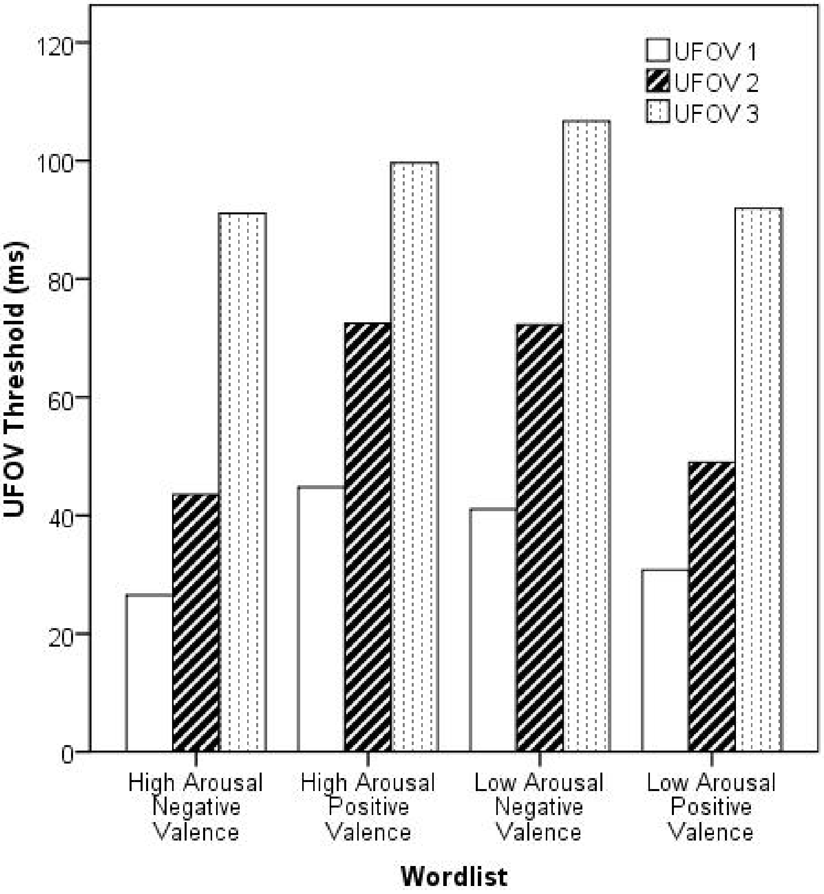


Figure 7. Mean UFOV thresholds observed under the "equal emphasis" prioritization condition.

Finally, when "ignoring" the UFOV task, participants exhibited mean UFOV thresholds of 169, 236, and 264, for the first, second, and third UFOV subtests, respectively (again, all comparisons significant, p <.05). The effect of UFOV difficulty under the "ignore the UFOV" condition was significant, F(2, 158) = 53.1, $p < .01$, $\eta^2 =$ .40. Thus, while a significant effect of UFOV task difficulty was observed under each of the three attentional prioritization conditions, this effect was smaller when the UFOV test was "ignored", (see Figure 8, below).

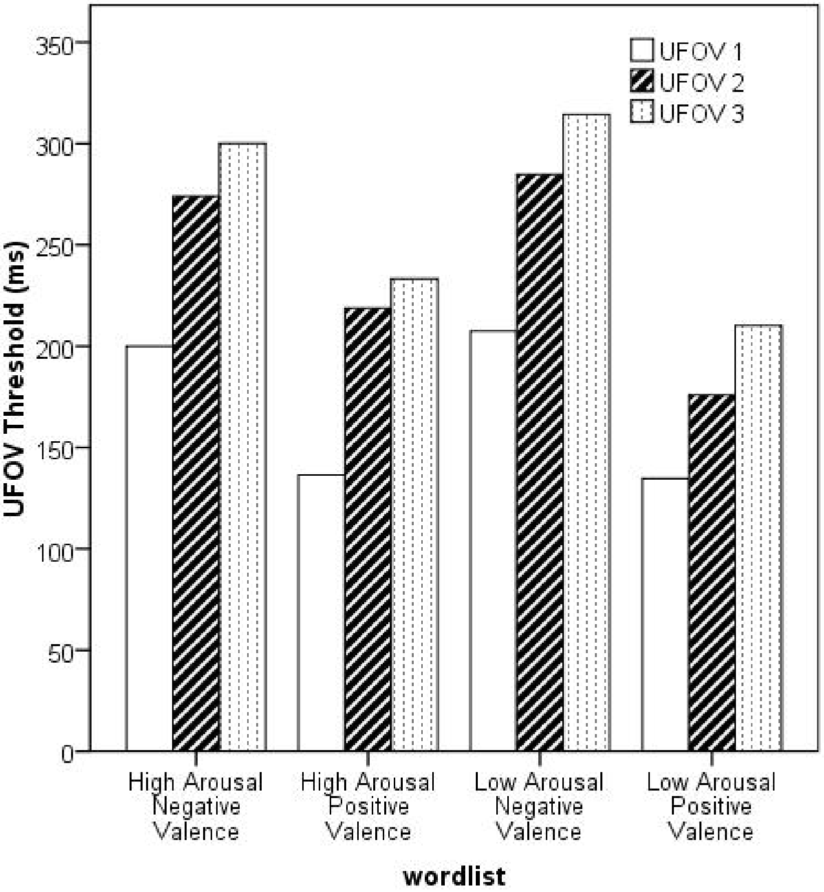


Figure 8. Mean UFOV thresholds observed under the "ignore the UFOV" prioritization condition.

No effects of the emotional qualities of the given words in the verbal task were observed on the UFOV thresholds. Neither emotional valence, *F* (1, 76) = 1.4, *p* = .24, nor emotional arousal level, *F* (1, 76) = .06, *p* = .8, demonstrated significant effects. No interaction effects with UFOV subtest or attentional allocation level (or both) were observed (all *p* >.05). However, within the first UFOV subtest, there was a significant, yet small, interaction between emotional valence and attentional focus, *F* (2,158) = 3.4, *p* < .05, η^2 = .04. Post-hoc analyses revealed that, within the first UFOV subtest, participants yielded UFOV thresholds of 16, 27, and 199 milliseconds under conditions of focusing on the UFOV, placing “equal” emphasis on both tasks, and “ignoring” the UFOV, respectively, when responding to negatively valent given words (of both high and low arousal levels). When responding to *positively* valent given words (of both high and low arousal levels) within the first UFOV subtest, participants yielded UFOV thresholds of 18, 43, and 140 milliseconds under conditions of focusing on the UFOV, placing “equal” emphasis on both tasks, and “ignoring” the UFOV, respectively. Pairwise comparisons revealed that longer thresholds were observed when ignoring negatively valent words (M = 199 ms) than when ignoring positively valent words (M = 140 ms, $p < .05$). Thus, it may be the case that it was more difficult for participants to “tune out” negatively valent verbal stimuli, resulting in larger UFOV thresholds during the first UFOV subtest. However, as the UFOV task became more complex, this effect diminished. Within the second UFOV task, this interaction was not significant, F (2,158) = 2.7, p = .07. Within the second UFOV task, this interaction was not significant, F (2,158) = 1.9, p = .15. That is, the increasing cognitive demands of a more complex concurrent UFOV task may have attenuated the effect of the emotional properties of the given verbal stimuli.

Verbal Task Data – Reaction Time

As described above, reaction time measures were measured from the offset of the given recorded word to the beginning of the participant’s utterance. From these reaction time measures, only reaction times for “related” responses were analyzed. That is, reaction times for *only* those responses with a mean similarity rating of 3.5 or greater were included in these analyses.

Unlike the pattern of data observed for the UFOV task performance, there was no significant effect of concurrent UFOV task difficulty on verbal task reaction time, *F* (2, 34) = 2.1, $p > .05$, $\eta^2 = .11$. However, a significant main effect of attentional focus was observed, *F* (2, 34) = 3.9, $p < .05$, $\eta^2 = .19$. Participants responded in the least amount of time when "ignoring" on the UFOV task (i.e., focused on the verbal task), *M* = 1120 milliseconds, significantly less time than when devoting "equal" attention to both tasks, *M* = 1241 ms, $p < .01$, yet not significantly less time than when focused the UFOV task (i.e., "ignoring" the verbal task), *M* = 1221 ms, $p > .05$, (see Figures 9, 10, and 11, below).

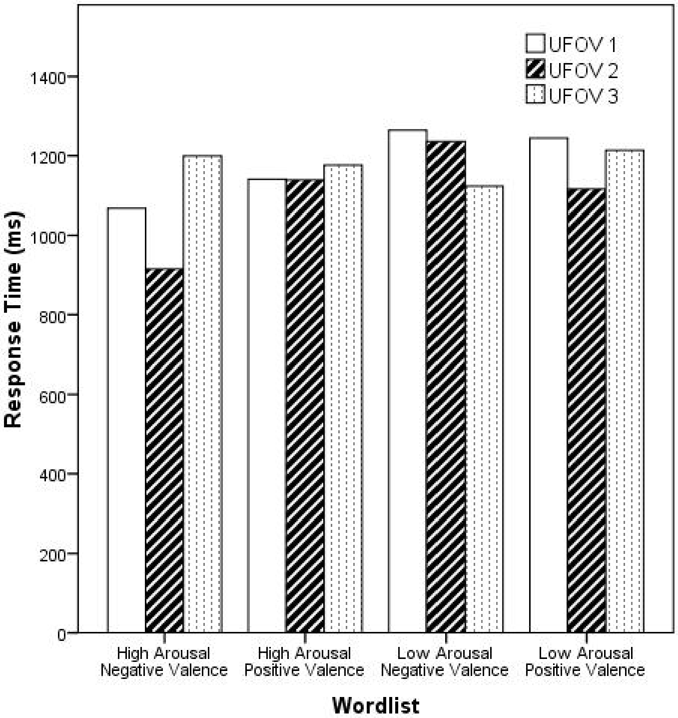


Figure 9. Reaction time of verbal responses under the "ignore the UFOV" prioritization condition.

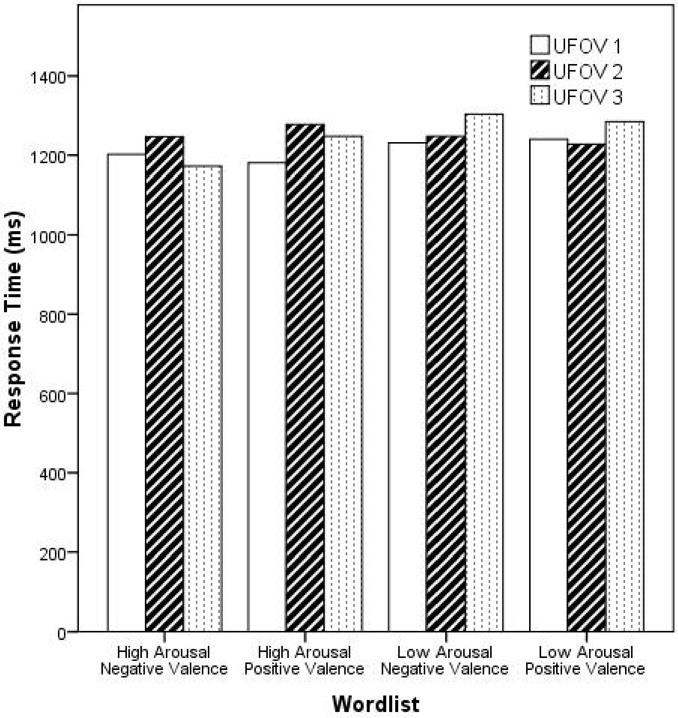


Figure 10. *Reaction time of verbal responses under the "equal emphasis" prioritization condition.*

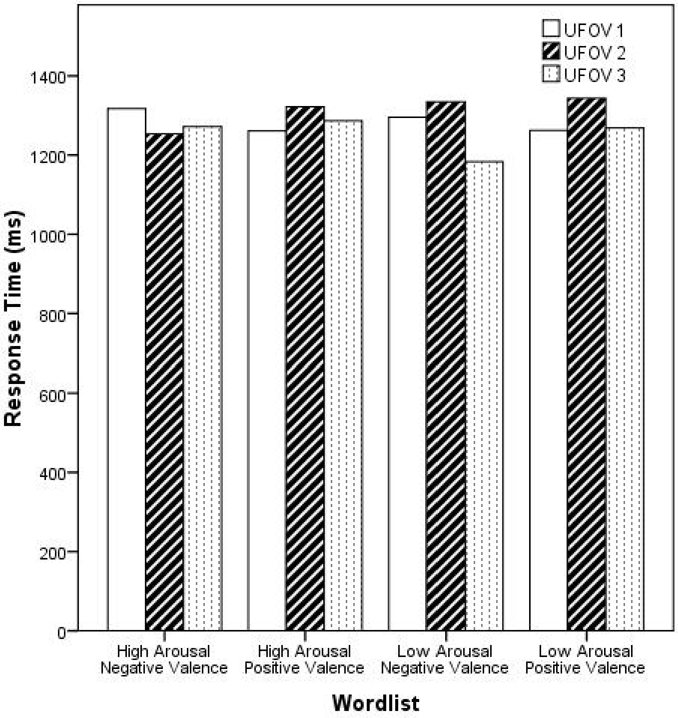


Figure 11. *Reaction time of verbal responses under the "focus on the UFOV" prioritization condition.*

That is, while placing additional attention on the verbal task improved performance relative to placing equal attention on both tasks or ignoring the verbal task, there was no difference in performance between participants ignoring the verbal task and placing equal emphasis on both tasks. In short, participants could improve performance by allocating more attention to the verbal task.

No main effect of the emotional qualities of the given words in the verbal task were observed on verbal task response times. That is, there was no main effect of the valence or arousal levels of the words presented to participants. Neither emotional valence, F (1, 64) = .16, $p = .69$, nor emotional arousal level, F (1, 64) = .10, $p = .75$, demonstrated significant effects. However, there was a significant interaction between prioritization level and the emotional category of the stimulus words. When participants placed equal emphasis on both tasks, high arousal negatively valent words yielded the longest mean response time (M = 1478ms, all $p < .05$). When participants focused on the UFOV task (i.e., “ignored” the verbal task), these same high arousal negatively valent words yielded the shortest mean response time (M = 970ms, all $p < .05$).

Thus, the greatest dual-task cost was observed when participants heard high arousal negatively valent words, yet these same words were responded to in the least amount of time when being “ignored”. Further discussion of these findings appears below.

Correlations of Verbal Task Reaction Time and UFOV Task Threshold

Correlations between UFOV task performance and verbal task reaction time are listed in Table 2, below. Correlations between performance on the concurrent tasks were small, and in only two cases, significant. Correlations ranged from -.04 to .29; the only significant correlations were observed during the second and third UFOV subtests under the equal emphasis prioritization condition (*r* = .27and .29, respectively, both *p* < .05). These positive correlations, though weak, suggest that under true dual task demands, performance on one task was not compromised for the sake of performance on the other. That is, as performance on one task increased, performance on the other task increased as well. Correlations were weakest when ignoring the UFOV test, suggesting that performance on the two concurrent tasks most independent under this prioritization schedule, again supporting the notion that participants were able to effectively "tune out" the visual attention task.

POC Curves for UFOV Threshold and Verbal Task Reaction Time

The first level of analysis (i.e., ANOVA and mean comparison of UFOV scores and verbal task RT in isolation) discussed above, found effects of both prioritization level and an interaction effect of prioritization level with the emotional properties of the given words. This interaction revealed that high arousal negatively valent words were responded to with the greatest reaction time under true dual task conditions, yet were most effectively ignored. In addition to these findings, the POC method of analysis describes profound differences in participants' ability to *allocate attention* to one or the other task depending on the emotional properties of the given words.

As mentioned above, the POC method is deigned to reveal the nature of the ability to allocate attention, or effort, to one or the other task in a dual task paradigm.

POC curves for UFOV threshold plotted against verbal task reaction time during the first, second, and third UFOV subtests appear as Figures 12, 13, and 14, below.

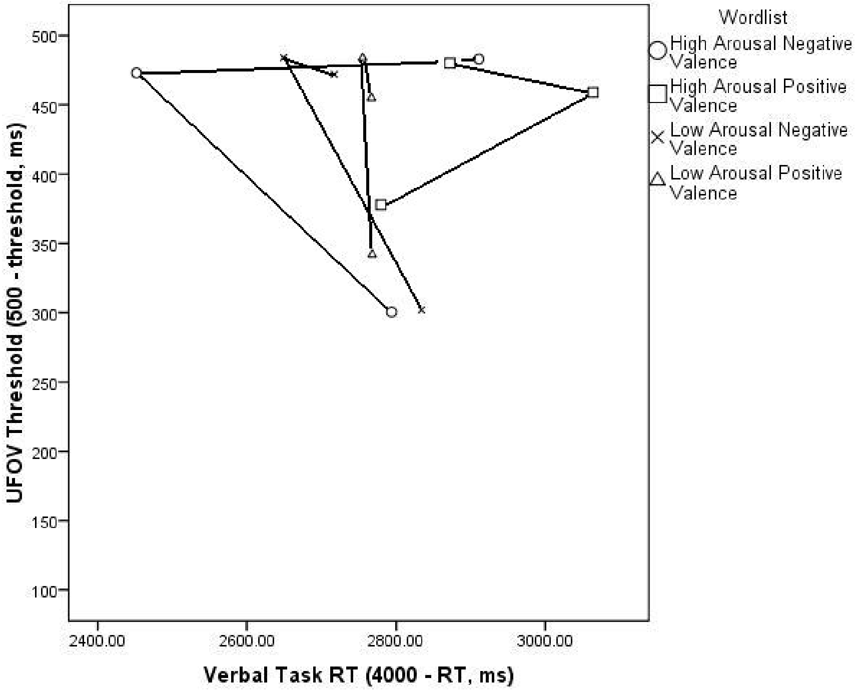


Figure 12. POC curve plotting UFOV threshold against verbal task reaction time, during the first UFOV subtest.

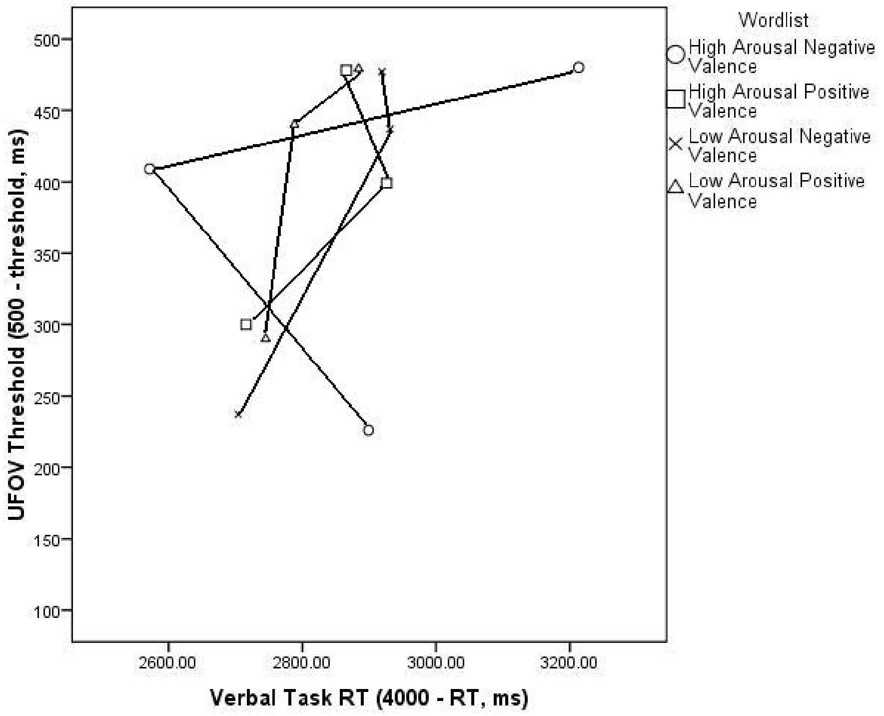


Figure 13. *POC curve plotting UFOV threshold against verbal task reaction time, during the second UFOV subtest.*

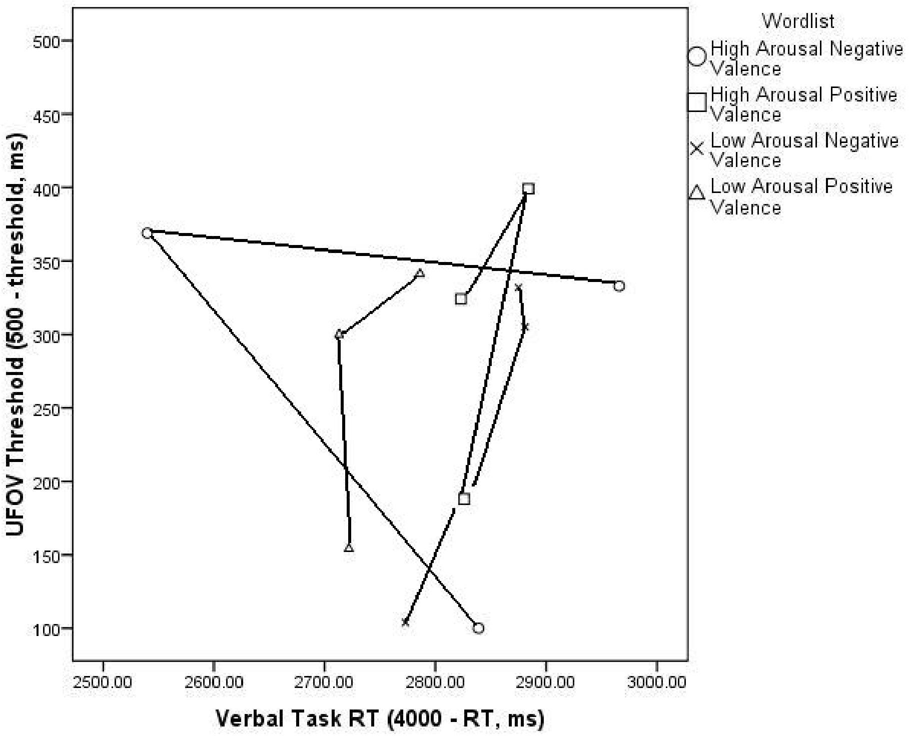


Figure 14. *POC curve plotting UFOV threshold against verbal task reaction time, during the third UFOV subtest.*

Goodness-of-fit tests (R^2) to a regression line for each of the curves can serve as a way to quantify the shape of the curve. However, for these patterns of data, which yielded atypical POC curves, goodness-of-fit tests is inappropriate and perhaps misleading. That is, the line that will be fit to the data does not adequately reflect the shape of these unique, and in some cases, vertical, POC curves. Thus, the shape of the POC curves will be discussed qualitatively rather than quantitatively.

As can be noted from the figures above, one robust pattern that emerges is the extreme *concave* shape of the POC curve yielded by participants who heard the high arousal negatively valent given words. That is, for each subtest of the UFOV, these POC curves appear in a “reverse numeral 7” shape. This shape suggests high performance on both tasks when ignoring the UFOV, low performance on the verbal task, yet high performance on the UFOV task when placing equal emphasis on both tasks, and reduced performance on the UFOV task with moderate performance on the verbal task when focusing on the UFOV task. This means that for this wordlist, participants did not ignore the UFOV when instructed to do so, yet focused on the UFOV when instructed to do so. Furthermore, these data mean that the dual task cost was absorbed to a greater level by the verbal task than the UFOV task.

These data partially support our hypothesis that high arousal negatively valent words would be most distracting to UFOV performance. In order to maintain a high level of performance on the UFOV task (as was observed in all other wordlist conditions, see above), reaction time to high arousal negatively valent words was compromised.

Verbal Task Data – Errors of Omission

Similar to the pattern of data observed for the UFOV task performance, there was a significant effect of concurrent UFOV task difficulty on verbal task errors of omission (expressed as the percent of possible responses that were omitted), F (2,128)= 8.0, p <.01, $\eta^2 = .11$. However, post-hoc paired comparisons revealed that this effect was driven by frequent errors of omission during the third (i.e., most complex) UFOV subtest. Participants responded least often while performing the third UFOV subtest concurrently (M = 44% errors of omission), significantly less often than when performing the second UFOV subtest (M = 39.7% errors of omission, $p < .01$), and first UFOV subtest (M = 38.3% errors of omission, $p < .01$). Verbal task performance during the first and second UFOV subtests did not significantly differ, $p = .26$. That is, the effect of concurrent UFOV task complexity on verbal task reaction time was only observed during the most complex UFOV subtest, just as was the case when considering verbal task response rate (though it should be noted that response rates of 4000 ms were entered for non-responses, making these data not independent).

Additionally, a significant main effect of attentional focus was observed, F (2, 128) = 67.4, $p < .01$, $\eta^2 = .51$. Participants responded least often when focused on the UFOV task (i.e., "ignoring" the verbal task), M = 61.6% errors of omission, see (Figure 15, below),

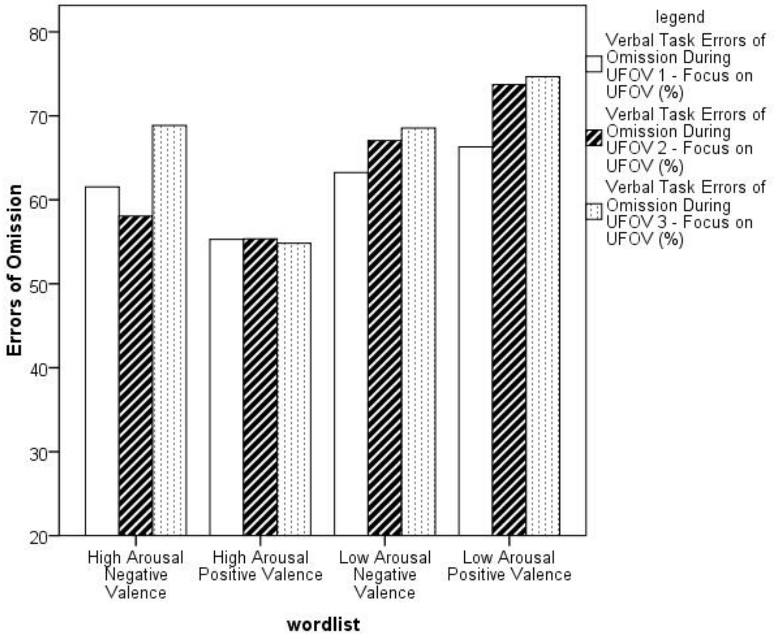


Figure 15. Percentage of possible responses omitted under the "focus on the UFOV" prioritization condition.

significantly less often than when devoting "equal" attention to both tasks, *M* = 29.3% errors of omission, *p* < .01, (see Figure 16, below),

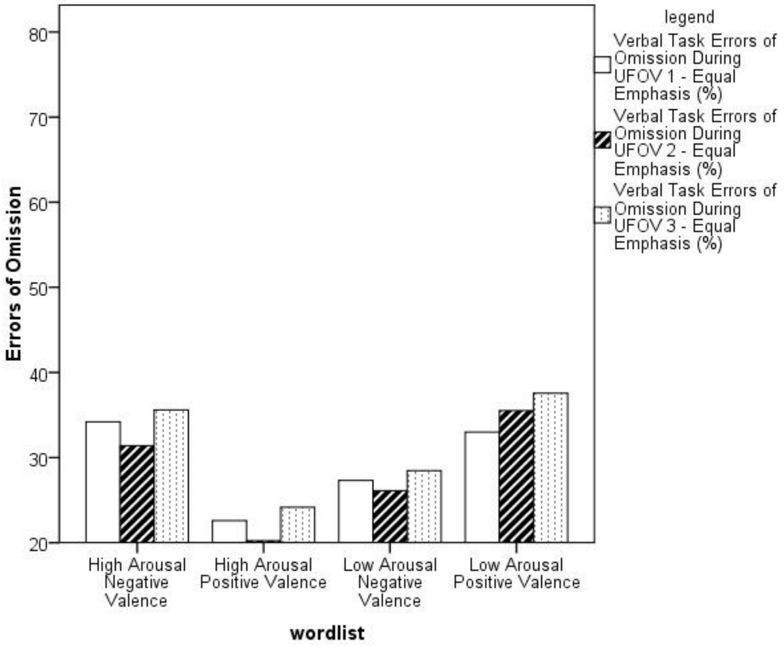


Figure 16. Percentage of possible responses omitted under the "equal emphasis" prioritization condition.

and when “ignoring” the UFOV task (i.e., focusing on the verbal task), *M* = 31% errors of omission, *p* < .01, (see Figure 17, below).

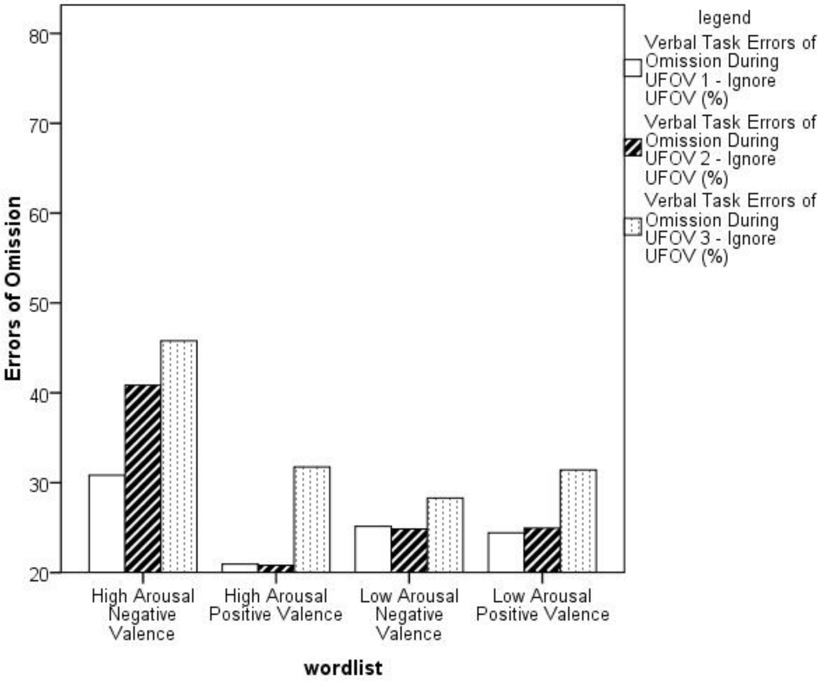


Figure 17. Percentage of possible responses omitted under the “ignore UFOV” prioritization condition.

Response times under conditions of “ignoring” the UFOV and placing “equal” emphasis on both tasks did not differ significantly, *p* = .39. That is, again, while placing additional attention on the verbal task did not improve performance relative to placing equal attention on both tasks, participants *were* able to effectively “tune out” the verbal task, and diminish verbal task performance.

No effects of the emotional qualities of the given words in the verbal task were observed on verbal task response times. Neither emotional valence, *F* (1, 64) = .37, *p* = .55, nor emotional arousal level, *F* (1, 64) = .03, *p* = .87, demonstrated significant effects.

No interaction effects with UFOV subtest or attentional allocation level (or both) were observed (all p >.05).

Correlations of Verbal Task Errors of Omission and UFOV Task Threshold

Correlations between UFOV task performance and verbal task errors of omission are listed in Table 3, below. Correlations between performance on the concurrent tasks were small, and in only one case, significant. Correlations ranged from -.15 to .2. As was the case with reaction time data, correlations were weakest when focusing on the UFOV test, suggesting that performance on the two concurrent tasks most independent under this prioritization schedule, again supporting the notion that participants were able to effectively "tune out" the verbal task.

Verbal Task Data – Accuracy

Accuracy scores were, as mentioned above, assessed by to raters who judges the similarity in meaning of the given word and the verbal response. Similarity was judged on a 7 point Likert scale, with a score of 1 reflecting "strongly unrelated" and a score of 7 reflecting "strongly related". Correlations between the mean ratings for participants within levels of the UFOV subtest and prioritization schedule provided by the two raters appear in Table 5. All correlations were significant (p < .01), and all had values of +.90 or greater. For each given word / response pair, the mean rating was determined. This rating was then dichotomized: responses with a mean rating of 3.5 or greater were judged to be related, responses with a mean rating of less than 3.5 were judged to be unrelated. Finally, the percentage of all responses that were judged to be related was determined.

Again, there was not a significant effect of concurrent UFOV task difficulty on verbal accuracy, *F* (2, 212)= .13, $p = .88$, $\eta^2 < .01$. Thus, the complexity of the concurrent UFOV task did not affect the accuracy of the words produced by participants. However, a significant, and notably large, main effect of attentional focus was observed, *F* (2, 212) = 42.5, $p < .001$, $\eta^2 = .6$. As might be expected, participants responded least accurately when focusing on the UFOV task (i.e., ignoring the verbal task) (*M* = 41.3%), (see Figure 18, below),

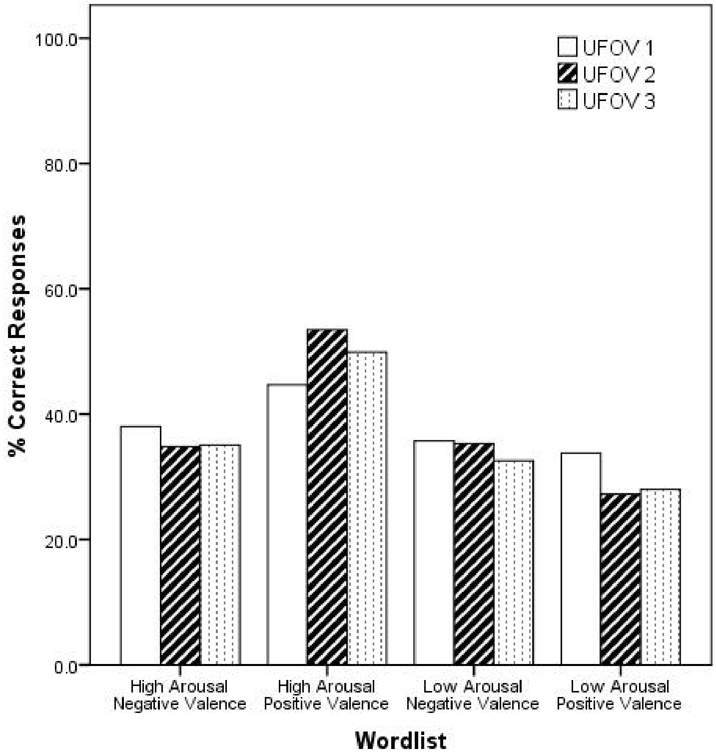


Figure 18. Percent of responses made that were judged to be "related" during the "ignore the UFOV" prioritization schedule.

significantly less accurate than when placing equal emphasis on both tasks (*M* = 85%), see Figure 19, below),

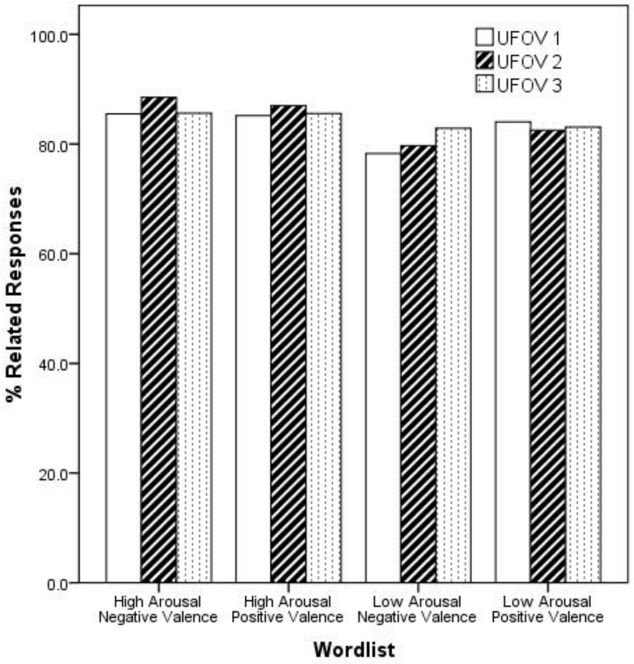


Figure 19. Percent of responses made that were judged to be "related" during the "equal emphasis" prioritization schedule.

and when focusing on the verbal task (*M* = 86.1%), (see Figure 20, below).

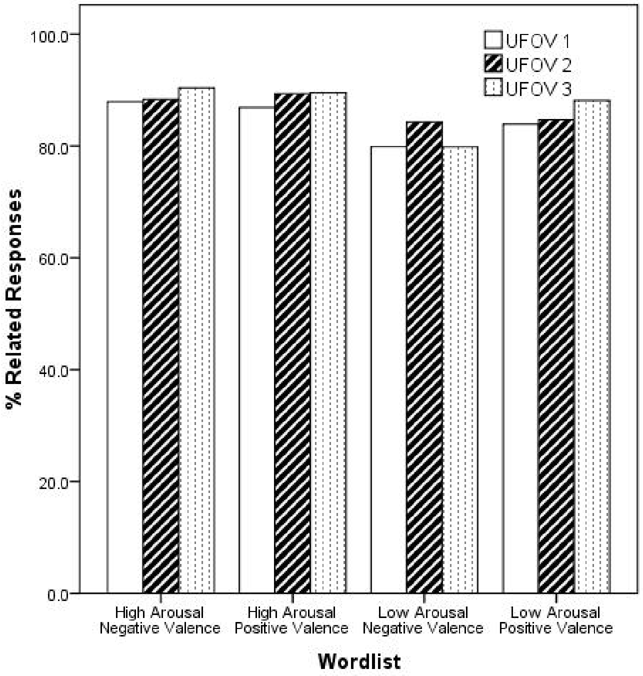


Figure 20. Percent of responses made that were judged to be "related" during the "ignore the UFOV" prioritization schedule.

Accuracy did not significantly differ between conditions of equal emphasis and focusing on the verbal task, $p > .05$. Again, participants appear to be able to “tune out” the verbal task effectively, but do not experience a direct overall cost in performance when dividing attention relative to focusing attention on the verbal task.

No main effects of emotion were observed. There was no main effect of the emotional category of words presented to participants, $F(3, 53) = 2.4$, $p = .07$, $\eta^2 = .12$. That is, participants responded with equivalent accuracy (or relatedness) regardless of the emotional properties of the given words.

Correlations of Verbal Task Accuracy and UFOV Task Threshold

Correlations between UFOV task performance and verbal task accuracy are listed in Table 4, below. Ranged from -.291 to .184, and were significant in two instances. Under the prioritization condition of placing equal emphasis on both tasks, and during the first and second UFOV tasks, correlations were -.291 and -.329, respectively (both $p < .01$). Thus, whereas verbal task *reaction time* yielded *positive* correlations with UFOV threshold, verbal task accuracy yielded *negative* correlations with UFOV performance. However, this is again consistent with a lack of tradeoff in performance from one task to another, as higher accuracy scores and *lower* thresholds are both indicative of better task performance. That is, as UFOV thresholds increased (i.e., poor performance), the percentage of accurate (or related) responses decreased (i.e., poor performance).

POC Curves for UFOV Threshold and Verbal Task Accuracy Ratings

The first level of analysis (i.e., ANOVA and mean comparison of UFOV scores and verbal task accuracy in isolation), as discussed above, found *no main effects* of the emotional properties of the given words on verbal task accuracy (i.e., relatedness of the responses). Furthermore, the only significant main effect, found for the prioritization level of the two tasks, was such that ignoring the UFOV (i.e., focusing on the verbal task) improved performance, but there was no significant difference between placing equal emphasis on both tasks and focusing on the UFOV (i.e., ignoring the verbal task). That is, no dual task cost was noted when considering verbal task accuracy. This lack of a dual task cost created unique POC curves.

Figures 21, 22, and 23 (see below) display the POC curves obtained by plotting UFOV thresholds for the first, second, and third UFOV subtests, respectively, against accuracy ratings in the verbal task. Again, due to the atypical shape of these POC curves, goodness-of-fit tests to quantify linearity are inappropriate and misleading.

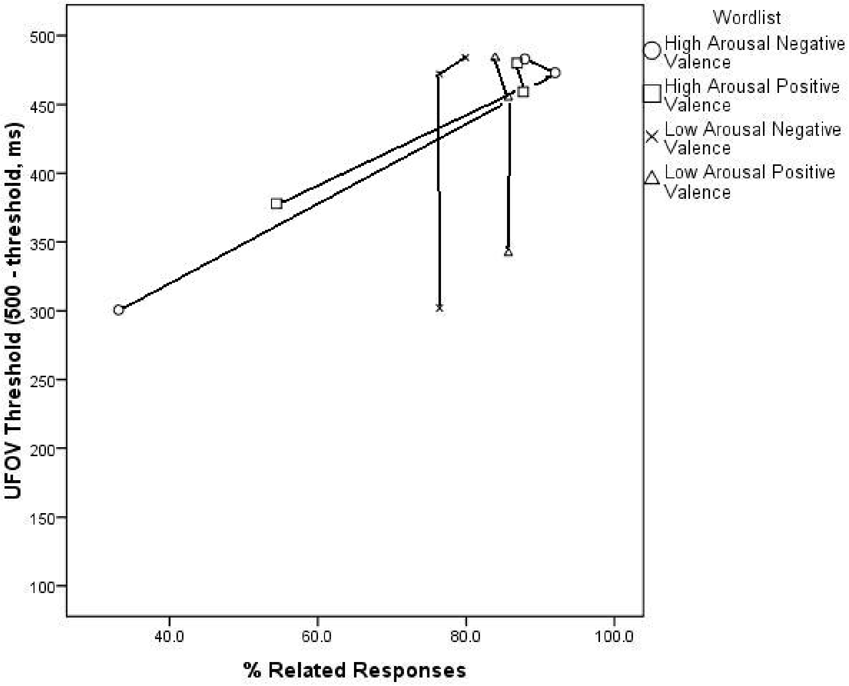


Figure 21. *POC curve plotting UFOV threshold against verbal task accuracy ratings, during the first UFOV subtest.*

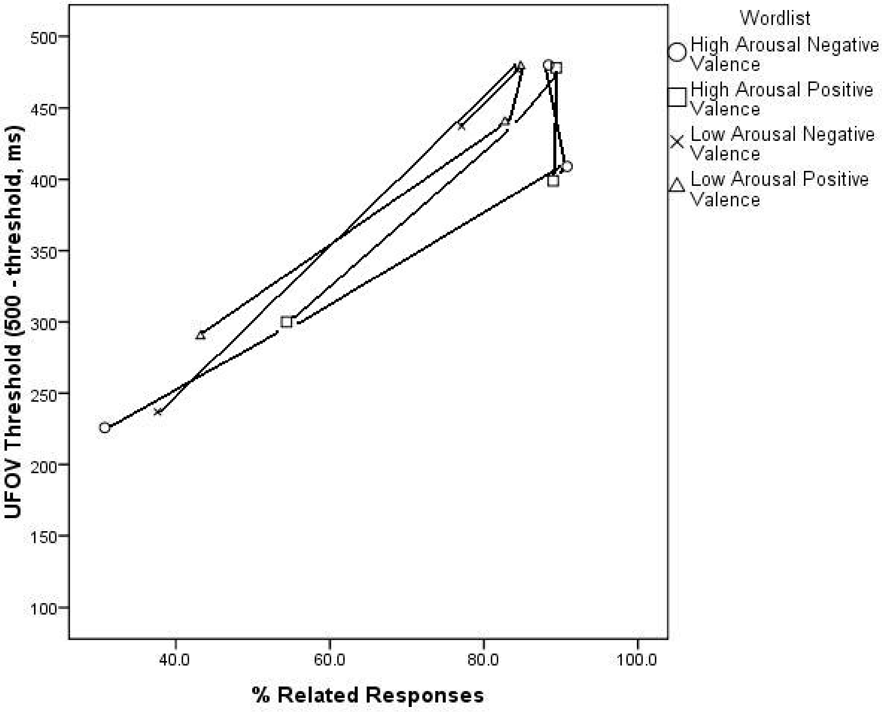


Figure 22. *POC curve plotting UFOV threshold against verbal task accuracy ratings, during the second UFOV subtest.*

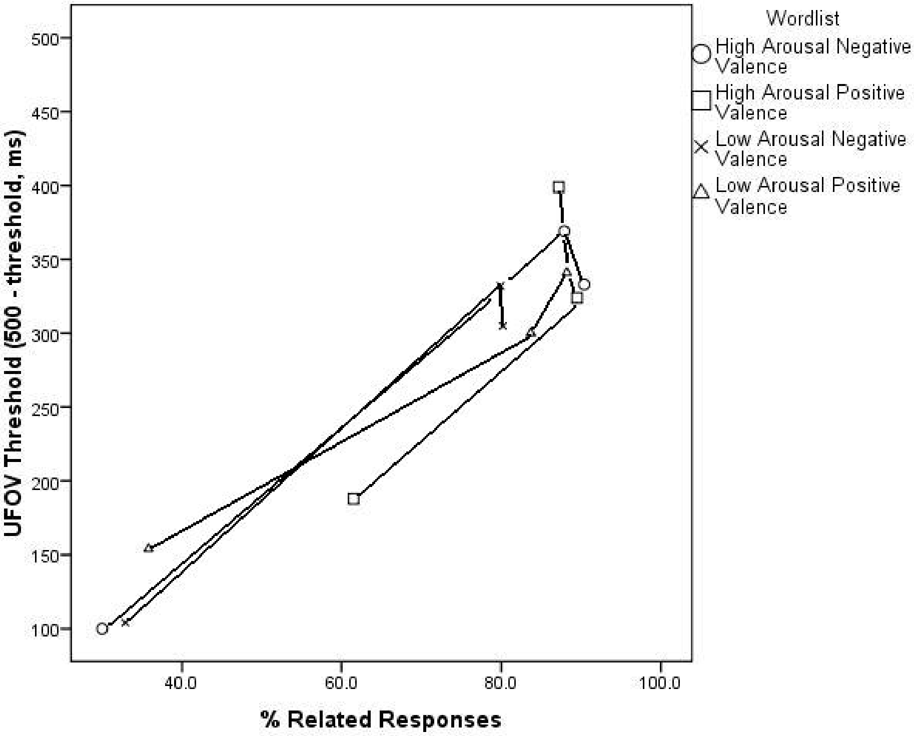


Figure 23. POC curve plotting UFOV threshold against verbal task accuracy ratings, during the third UFOV subtest.

These POC curves can be summarized, clearly, as linear, but in the *opposite* direction than is typically observed. The lack of a dual task cost in verbal task accuracy created POC curves with essentially *two* points, rather than the intended *three*. In each of the graphs above, there is a cluster of points for the "focus on the UFOV" prioritization condition, and a single cluster of points for both the "place equal emphasis on both tasks" and "ignore the UFOV" condition. Thus, the POC curves connect, in essence only two points, resulting in a line, rather than a curve. Nonetheless, these POC curves reflect the lack of a performance gain with regard to verbal task accuracy when focusing on the verbal task.

NASA TLX Data

Overall TLX Scores

There was no effect of attentional prioritization on TLX index scores, F (2, 152) = 1.3, $p = .28$, $\eta^2 = .02$. Participants did not rate the task load of the dual task significantly differently under conditions of focusing on the UFOV (M = 69.1), placing equal emphasis on both tasks (M = 67.3), or ignoring the UFOV (M = 59.7). Furthermore, there was no effect of wordlist category on TLX index scores, F (3, 76) = .78, $p = .51$, $\eta^2 = .03$. That is, participants did not rate the task load of the dual task significantly differently when hearing high arousal negatively valent words (M = 72), high arousal positively valent words (M = 65.4), low arousal negatively valent words (M = 64.6), or low arousal positively valent words (M = 59.5). The dual-task scenario was thus rated as having essentially equal task load (or cognitive demand) regardless of the attentional prioritization or wordlist employed.

TLX Performance Scores

Similarly, there was no effect of attentional prioritization on TLX “Performance” scores, F (2, 152) = .93, $p = .4$, $\eta^2 = .01$. Participants did not rate their overall performance of the dual task significantly differently under conditions of focusing on the UFOV (M = 51.6), placing equal emphasis on both tasks (M = 48), or ignoring the UFOV (M = 48.6). Furthermore, there was no effect of wordlist category on TLX index scores, F (3, 76) = .55, $p = .65$, $\eta^2 = .02$. That is, participants did not rate the task load of the dual task significantly differently when hearing high arousal negatively valent words (M = 46.9), high arousal positively valent words (M = 49.5), low arousal negatively valent words (*M* = 52.7), or low arousal positively valent words (*M* = 48.5). Participants felt they performed equally well under each of these conditions.

Correlations between TLX Performance Scores and UFOV Thresholds

Under the condition of ignoring the UFOV task, correlations between TLX performance ratings and UFOV thresholds were small, negative, and nonsignificant. TLX performance scores correlated with the first, second, and third UFOV subtest thresholds at -.18, -.15, and -.18, respectively.

When focusing on the UFOV task, a similar, yet weaker pattern of correlations was observed. TLX performance scores correlated with the first, second, and third UFOV subtest thresholds at -.05, -.02, and -.07, respectively, in this condition.

When placing equal emphasis on both tasks, as was hypothesized, correlations were again weak, but now positive. TLX performance scores correlated with the first, second, and third UFOV subtest thresholds at .12, .08, and .09, respectively, in this condition. While interpreting small, nonsignificant correlations should be done with caution, it is compelling that in all other instances, a negative correlation between subjective rating of performance and UFOV performance was exhibited. It may be the case that participants are (to a small degree) best able to assess their UFOV task performance under conditions of emphasizing both tasks.

Correlations between TLX Performance Scores and Verbal Task Accuracy

Under the condition of ignoring the UFOV task, correlations between TLX performance ratings and verbal task accuracy were weak, negative, and nonsignificant. At this prioritization schedule, TLX performance scores correlated with verbal task accuracy during the first, second, and third concurrent UFOV subtests at -.16, -.07, and -.05, respectively. Thus, a weak relationship between perceived performance and actual performance existed when ignoring the UFOV task.

When placing equal emphasis on both tasks, correlations were notably weaker, and not significant. TLX performance scores correlated with the first, second, and third UFOV subtest thresholds at -.05, -.005, and -.05, respectively, in this condition. As was the case during the “ignore the UFOV” prioritization schedule, no relationship between perceived performance and actual performance was observed during the “equal emphasis” prioritization schedule.

However, when focusing on the UFOV task, a moderate and significant pattern of correlations was exhibited. TLX performance scores correlated with verbal task accuracy ratings during the first, second, and third concurrent UFOV subtests at .41, .48, and .49, respectively, in this condition. These were the strongest correlations with the TLX performance ratings in any category, and each were significant ($p < .01$). Thus, it may be the case that participants are able to most accurately judge their verbal task performance under conditions of ignoring the verbal task (i.e., when ignoring the words, they are aware of their lack of performance).

Discussion

Support for hypotheses

Although the effects of negatively valent emotional verbal tasks were not found in a robust manner, the current study demonstrated compelling effects, and support for many of our hypotheses, mentioned above. Specifically, there is evidence to support the hypothesis of a dual-task cost, the hypothesis of verbal task interference with visual attention, the hypothesis that differing emotional characteristics of the given words will affect the shape of POC curves, and the hypothesis that weak positive correlations will exist between TLX ratings and task performance.

The hypothesis of a dual task cost was supported by a significant effect of attentional prioritization on UFOV thresholds. Though smaller than some dual-task costs we've induced in our lab, and smaller than observed in the pilot data, participants exhibited larger UFOV thresholds (*M* = 64 ms) when devoting "equal" attention to both tasks than when focusing on the UFOV (*M* = 35 ms), meaning that the UFOV stimuli had to remain on the screen nearly twice as long under "true" dual task conditions relative to, essentially, single task conditions. The additional finding that much larger UFOV thresholds were found when participants were told to "ignore" the UFOV (*M* = 223 ms), suggests that participants were effective at *attenuating* processing of the visual stimuli. That is, they showed a greater difference between focusing on both tasks and *ignoring* the UFOV task than between focusing on both tasks and *focusing on* the UFOV task.

Likewise, the hypothesis of a dual task cost was supported by a significant effect of attentional prioritization on verbal task reaction time and errors of omission. Participants took longer to respond when dividing attention between tasks equally than when focusing on the verbal task (*M*s = 1230 and 1123 ms, respectively), responded less often when dividing attention than when focusing on the verbal task (*M*s = 61.6% errors and 29.3% errors of omission, respectively). Additionally, participants responded less accurately when ignoring the verbal task than when dividing attention (*M*s = 41.3% and 85%, respectively), suggesting that participants were able to "tune out" the verbal task, and diminish performance, perhaps allowing for improved performance on the UFOV task.

The hypothesis that the shape of the POC curves would be affected by emotional properties of the given words in the verbal task was supported as well, in a similar manner to that which was hypothesized. It was hypothesized that high arousal negatively valent words would yield the most linear POC curves, and that high arousal positively valent words would yield the most convex POC curves. Instead, high arousal negatively valent words yielded a *concave* POC curve in the case of reaction time POCs. That is, there was a unique ability for subjects in this condition to maintain UFOV task performance at a high level by sacrificing verbal task performance, at a dramatically greater level than for the other emotional conditions. This, coupled with the traditional analysis findings that high arousal negatively valent words yielded reaction times that were greatest of any emotional category during the “equal emphasis” prioritization condition support the hypothesis that high arousal negatively valent words are especially distracting to visual attention performance.

However, the unique shape of the reaction time POC curves, and the linear shape of the accuracy POC curves, suggest that for these particular task measures, the Performance Operating Characteristic method may not be suitable. That is, the sensitivity of these tasks and their measures may not yield typical POC curves. The range of scores, particularly for the verbal task dependent measures, was not large. Furthermore, little difference was observed between various combinations of the prioritization levels. If POC curves are to yield fruitful interpretations, more sensitive measures which differentiate performance at these prioritization levels must be employed.

Additionally, the hypothesis that NASA TLX ratings of task difficulty and task performance would be weakly, yet positively correlated with actual performance was partially supported. Participants demonstrated strongest correlations between TLX ratings and UFOV threshold under the condition of placing equal emphasis on both verbal and UFOV tasks. Though the correlations were weak, it could mean, counterintuitively, that a better estimate of UFOV performance can be made under dual task conditions than when focusing on the UFOV. This may be due to participants understanding that UFOV performance suffers under dual task demands relative to focusing on the UFOV, and accurately reducing their subjective performance ratings (though this difference was not significant, it may have affected these weak correlations). Furthermore, stronger positive correlations between subjective NASA TLX ratings and verbal task accuracy were observed when participants focused on the UFOV task (i.e., when they ignored the verbal task). That is, participants most accurately rated their overall performance when they could “factor out” verbal task performance, or account for poor verbal task performance. When participants ignored the verbal task, they accurately described their performance decrease. The fact that these correlations were larger may indicate that participants’ performance ratings were driven more by verbal task performance than by UFOV task performance. This may have important implications for drivers conversing on a cellular phone, as they may judge their overall performance based largely on the quality of their conversation.

Implications

The current study demonstrated, in agreement with a wealth of dual task literature, that there is a notable effect of verbal tasks on visual attention. The Useful Field of View, or the extent (either in terms of spatial attention or efficiency) to which information can be extracted from a visual scene and used to guide action, is substantially impacted by the presence of a concurrent verbal task. Whereas *none* of the participants yielded UFOV scores that predict increased accident risk while focusing on the UFOV task, 19% of these *same* participants yielded UFOV scores that predict greater than double the risk of getting into an injurious accident when performing a concurrent verbal task. Fifteen subjects demonstrated this increased risk prediction, six in the low arousal negative valence wordlist condition, five in the high arousal positive valence wordlist condition, three in the low arousal positive valence wordlist condition, and one in the high arousal negative valence wordlist condition.

These data add support to a growing corpus of research that suggests that the *type* of conversation (or verbal task) carried out concurrently with a driving (or visual attention) task can affect visual attention performance. Clearly, there are differences in the dual task demands created by the different wordlists used in this experiment.

In summary, participants seem to be able to effectively “tune out” one or the other task in a dual task paradigm, though traditional analyses and POC analyses converge to inform us that under these experimental conditions (which may not require adequate cognitive load), UFOV performance is not as greatly impacted by concurrent verbal tasks as pilot data and theory suggest. While smaller than expected, these dual task costs have implications in an applied setting, as 19% of subjects exhibited UFOV scores under dual task conditions that would predict more than double the risk of injurious accident. Finally, highly arousing negatively valent verbal stimuli may be the most distracting category of emotional verbal tasks (i.e., proto-conversations) to have while driving. This is congruent with behavioral (e.g., Easterbrook, 1959), neurophysiological (e.g., Bradley, et al., 2003), and evolutionary (e.g., Davidson, 2003) theories regarding the nature of emotional stimuli and attention.

Directions for further study

As mentioned above, the manipulations of task load via prioritization schedules implemented in this study may not have been sufficient to tease apart robust differences found in task performance between emotional properties of the given words. In addition to adjusting prioritization schedules, verbal task load can be manipulated in many other ways. Potential manipulations include, but are not limited to: adjusting the concreteness or abstractness of the given words and intended responses, adjusting frequency of occurrence of the given words (i.e., participants could be asked to respond to either very obscure words or very common words), adjusting the semantic neighborhood set size of the given words, or adding various levels of auditory noise to the given word signal.

By using additional methods of manipulating task load of the verbal task, not only can the robustness of the interference effects discussed in this dissertation be assessed, but more specific mechanisms of language processing can be weighed as contributing to the interference with visual attention performance. Furthermore, tasks that more closely resemble real, natural conversations can be used (still employing varying degrees of experimental control and manipulation) can and should be employed, making the link between cellular phone use and driving accidents more valid. The current study employed a set of “verbal tasks”, but lacked the planning, the syntax, the predictability and “cloze” features of naturalistic conversation. Investigation to cognitive mechanisms of language processing which interfere with visual attention would do well to include these features in their design and manipulations.

References

Atchley, P. & Dressel, J. (2004). Conversation limits the functional field of view. *Human Factors, 46*, 664-673.

Atchley, R., Ilardi, S., & Enloe, A. (2003). The effect of current and past depression. *Brain and Language, 84,* 105-119.

Atchley, R., Stringer, R., Mathias, E., Ilardi, S., & Minatrea, A. (in press). The right hemisphere's contribution to emotional word processing in currently depressed, remitted depressed, and never-depressed individuals.

Ball, K. K. and Owsley, C. (1991). Identifying correlates of accident involvement for the older driver. *Human Factors, 33,* 583-595.

Ball, K. K., Owsley, C., Stalvey, B. T., Roenker, D. L., Sloane, M. E., and Graves, M. (1998). Driving avoidance and functional impairment in older drivers. *Accident Analysis and Prevention, 30,* 313-322.

Ball, K.K., Roenker, D.L., Wadley, V.G., Edwards, J.D., Roth, D.L., McGwin, G., Raleigh, R., Joyce, J.J., Cissell, G.M., & Dube, T. (2006). Can high-risk older drivers be identified through performance-based measures in a department of motor vehicles setting? *Journal of the American Geriatrics Society, 54,* 77-84.

Bradley, M., & Lang, P. (1999). Affective norms for English words (ANEW). Gainesville, FL. The NIMH Center for the Study of Emotion and Attention, University of Florida.

Bradley, M., Sabatinelli, D., Lang, P., Fitzsimmons, J., King, W., & Desai, P. (2003). Activation of the visual cortex in motivated attention. *Behavioral Neuroscience,* *117,* 369-380.

Bower, G.H. (1981). Mood and memory. *American Psychologist, 36,* 129-148.

Calvo, M.G., Castillo, M.D., & Fuentes, L.J. (2006). Processing of "unattended" threat-related information: Role of emotional content and context. *Cognition and Emotion, 20,* 1049-1074.

Cowan, N., Wood, N.L., & Borne, D. N. (1994). Reconfirmation of the short-term storage concept. *Psychological Science, 5,* 103-106.

Davidson, R.J. (2003). Affective neuroscience and psychophysiology: Toward a synthesis. *Psychophysiology, 40,* 655-665.

Davidson, R.J., van Reekum, C.M. (2005). Emotion is not one thing. *Psychological Inquiry,* 11-15.

Easterbrook, J.A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review, 66,* 183-201.

Edwards, J.D., Vance, D.E., Wadley, V.G., Cissell, G.M., Roenker, D.L., & Ball, K.K. (2005). Reliability and validity of useful field of view test scores as administered by personal computer. *Journal of Clinical and Experimental Neuropsychology, 27,* 529-543.

Ellis, R.D. (2005). *Curious Emotions: Roots of consciousness and personality in motivated action. Advances in consciousness research, v61.* Amsterdam, Netherlands: John Benjamins Publishing Company.

Gopher, D., and Donchin, E. (1986). Workload – an examination of the concept. In: K.R. Boff, L. Kaufman, & J.P. Thomas, eds. *Handbook of Human Perception and*

Performance: Volume II, Cognitive Processes and Performance. New York: John Wiley & Sons.

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

Hart, & Staveland. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. In: P.A. Hancock & N. Meshkati (Eds.), *Human Mental Workload,* (pp. 239-250). Amsterdam: North Holland Press.

Heller, W., & Nitschke, J.B. (1998). The puzzle of regional brain activity in depression and anxiety: The importance of subtypes and comorbidity. *Cognition and Emotion, 12,* 421-447.

Hoffman, L., McDowd, J., Atchley, P., & Dubinsky, R. (2005). The role of visual attention in predicting driving impairment in older adults. *Psychology and Aging, 20,* 610-622.

Kahneman, D. (1973). *Attention and Effort*. New Jersey: Prentice Hall.

Kucera, H. and Francis, W.N. (1967). *Computational Analysis of Present-Day American English*. Brown University Press, Providence, RI.

Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in Cognitive Sciences, 9,* 75-82.

Lavie, N., & Cox, S. (1997). On the efficiency of attentional selection: Efficient visual search results in inefficient rejection of distraction. *Psychological Science, 8,* 395-398.

Lavie, N., Hirst, A., de Fockert, J., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General, 133,* 339-354.

Lang, P.J., Bradley, M.M., & Cuthbert, B.N. (2001). *International affective picture system (IAPS): Instruction manual and affective ratings.* Technical report A-5, The Center for Research in Psychophysiology, University of Florida.

Lang, P.J., Greenwald, M.K., Bradley, M.M., & Hamm, A.O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology, 30,* 137-159.

Levenson, R.W. (2003). Blood, sweat, and fears: The autonomic architecture of emotion. *Annals of the New York Academy of Sciences, 1000,* 348-366.

Navon, D. (1984). Resources – a theoretical soup stone? *Psychological Review, 91,* 216-234.

Norman, D.A., and Bobrow, D.G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology, 7,* 502-507.

Norman, D.A., and Bobrow, D.G. (1976). On the analysis of performance operating characteristics. *Psychological Review, 83,* 508-510.

Owsley, C., Ball, K., McGwin, G., Sloane, M., Roenker, D., White, M., & Overley, T. (1998). Visual impairment and risk of motor vehicle crash among older adults. *Journal of the American Medical Association, 279,* 1083-1088.

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

Pessoa, L., & Ungerleider, L.G. (2005). Visual attention and emotional perception. In L.

Itti, G. Rees, and J.K. Tsotsos (Eds.), *Neurobiology of attention*. San Diego, CA: Elsevier.

Robinson., M.D. (1998). Running from William James' bear: A review of preattentive mechanisms and their contributions to emotional experience. *Cognition and Emotion, 12,* 667-696.

Robinson, M.D., Storbeck, J., Meier, B.P., & Kirkeby, B.S. (2004). Watch out! That could be dangerous: Valence-arousal interactions in evaluative processing. *Personality and Social Psychology Bulletin, 30,* 1472-1484.

Seiple, W., Szlyk, J.P., Yang, S., & Holopigian, K. (1996). Age-related functional field losses are not eccentricity dependent. *Vision Research, 36,* 1859-1866.

Sekuler, A.B., Bennett, P.J., & Mamelak, M. (2000). Effects of aging on the useful field of view. *Experimental Aging Research, 26,* 103-120.

Schupp. H.T., Cuthbert, B.N., Bradley, M.M., Hillman, C.H., Hamm, A.O., & Lang, P.J. (2004). Brain processes in emotional perception: Motivated attention. *Cognition and Emotion, 18,* 593-611.

Tellinguisen, D., & Nowak, E. (2003). The inability to ignore auditory distractors as a function of visual task perceptual load. *Perception and Psychophysics, 65,* 817-828.

Tsang, P.S., and Velaquez, V.L. (1996). Diagnosticity and multidimensional subjective workload ratings. *Ergonomics, 39*, 358-381.

White, M.P., Eiser, J.R., & Harris, P.R. (2004). Risk perceptions of mobile phone use while driving. *Risk Analysis, 24,* 323-334.

Wickens, C.D. (2002). Multiple resources and performance prediction. *Theoretical Issues* *in Ergonomics Science, 3,* 159-177.

Yamasaki, H., LaBar, K., & McCarthy, G. (2002). Dissociable prefrontal brain systems for emotion and cognition. *Proceedings of the National Academy of Sciences of the United States of America, 99,* 11447-11451.

Table 1.

Appearance of Four Key Design Elements in Articles Cited in Literature Review

Author	"Driving" task	1	2	3	4
Atchley, P. & Dressel (2004)	UFOV	Yes	No	Both tasks	No
Amado & Ulpinar (2005)	Attention tasks	Yes	No	Conversation task	No
Briem & Hedman (1995)	Pursuit-tracking	No	No	Both tasks	No
Brown, et al. (1969)	On-road	Yes	No	Driving task	No
Consiglio, et al. (2003)	Simple RT	No	No	Both tasks	No
Horberry, et al. (2006)	Simulator	Yes	No	Driving task	Yes
Horswill & McKenna (1999)	Simulator	Yes	No	Driving task	No
Jamson, et al. (2004)	Simulator	Yes	No	Driving task	No
Lee, et al. (2001)	Simulator	Yes	No	Both tasks	Yes
McCarley, et al. (2004)	Change blindness	No	No	Conversation task	No
McKnight & McKnight (1993)	Simulator	No	No	Both tasks	No
Rakauskas, et al. (2004)	Simulator	Yes	No	Both tasks	Yes
Recarte & Nunes (2000)	Eyetracking	No	No	Both tasks	No
Shinar, et al. (2005)	Simulator	Yes	No	Both tasks	Yes
Strayer & Drews (2004)	Simulator	No	No	Driving task	No
Strayer & Johnston (2001)	Pursuit-tracking	No	No	Driving task	No
Strayer, et al. (2003)	Simulator	No	No	Driving task	No

Table 2.

Correlations between UFOV task thresholds and verbal task reaction time (separated by UFOV subtest, such that concurrent tasks are displayed together).

Ignore UFOV Condition

	RT During UFOV 1	RT During UFOV 2	RT During UFOV 3
UFOV 1 Threshold	.009		
UFOV 2 Threshold		.046	
UFOV 3 Threshold			-.039

Equal Emphasis Condition

	RT During UFOV 1	RT During UFOV 2	RT During UFOV 3
UFOV 1 Threshold	.226		
UFOV 2 Threshold		.273*	
UFOV 3 Threshold			.29*

Focus on UFOV Condition

	RT During UFOV 1	RT During UFOV 2	RT During UFOV 3
UFOV 1 Threshold	.129		
UFOV 2 Threshold		.281	
UFOV 3 Threshold			-.026

Table 3.

Correlations between UFOV task thresholds and verbal task errors of omission (separated by UFOV subtest, such that concurrent tasks are displayed together).

Ignore UFOV Condition

	Omit During UFOV 1	Omit During UFOV 2	Omit During UFOV 3
UFOV 1 Threshold	-0.11		
UFOV 2 Threshold		-0.09	
UFOV 3 Threshold			-0.15

Equal Emphasis Condition

	Omit During UFOV 1	Omit During UFOV 2	Omit During UFOV 3
UFOV 1 Threshold	-0.05		
UFOV 2 Threshold		0.02	
UFOV 3 Threshold			0.20

Focus on UFOV Condition

	Omit During UFOV 1	Omit During UFOV 2	Omit During UFOV 3
UFOV 1 Threshold	0.08		
UFOV 2 Threshold		-0.11	
UFOV 3 Threshold			0.03

Table 4.

Correlations between UFOV task thresholds and verbal task accuracy ratings (separated by UFOV subtest, such that concurrent tasks are displayed together).

<u>Ignore UFOV Condition</u>

	Acc. During UFOV 1	Acc. During UFOV 2	Acc. During UFOV 3
UFOV 1 Threshold	-.025		
UFOV 2 Threshold		.02	
UFOV 3 Threshold			-.004

<u>Equal Emphasis Condition</u>

	Acc. During UFOV 1	Acc. During UFOV 2	Acc. During UFOV 3
UFOV 1 Threshold	-.291**		
UFOV 2 Threshold		-.329**	
UFOV 3 Threshold			-.146

<u>Focus on UFOV Condition</u>

	Acc. During UFOV 1	Acc. During UFOV 2	Acc. During UFOV 3
UFOV 1 Threshold	.038		
UFOV 2 Threshold		.184	
UFOV 3 Threshold			.031

Table 5.

Correlations between mean similarity scores for participants, within UFOV condition and attentional prioritization condition.

Focus on UFOV Condition

	Rater 2		
	UFOV 1	UFOV 2	UFOV 3
Rater 1			
UFOV 1	.993**		
UFOV 2		.997**	
UFOV 3			0.968**

Equal Emphasis Condition

	Rater 2		
	UFOV 1	UFOV 2	UFOV 3
Rater 1			
UFOV 1	.935**		
UFOV 2		.939**	
UFOV 3			.948**

Ignore UFOV Condition

	Rater 2		
	UFOV 1	UFOV 2	UFOV 3
Rater 1			
UFOV 1	.917**		
UFOV 2		.901**	
UFOV 3			.955**

Appendix

Negative Valence High Arousal

- abuse
- accident
- afraid
- agony
- ambulance
- anger
- angry
- annoy
- assault
- bankrupt
- betray
- blackmail
- bloody
- bomb
- brutal
- burn
- cancer
- cockroach
- crash
- crucify
- danger
- demon
- despise
- destroy
- detest
- devil
- disaster
- disloyal
- distressed
- divorce
- drown
- enraged
- fear
- fearful
- guillotine
- guilty
- hate
- hatred
- horror
- hostage
- hostile
- humiliate
- insult
- intruder
- jealousy
- killer
- leprosy
- lie
- mad
- murderer
- mutilate
- nightmare
- pain
- pervert
- poison
- pollute
- punishment
- quarrel
- rabies
- rage
- rape
- rejected
- riot
- roach
- rude
- scalding
- scared
- sinful
- slap
- slaughter
- slave
- stress
- suffocate
- surgery
- terrible
- terrified
- terrorist
- thief
- tornado
- torture
- toxic
- tragedy
- trauma
- troubled
- tumor
- ulcer
- unfaithful
- vandal
- venom
- victim
- violent
- war
- wicked

Negative Valence Low Arousal

- ache
- addicted
- alone
- anguished
- blister
- bored
- burial
- cemetery
- coffin
- corpse
- coward
- crime
- criminal
- crisis
- death
- defeated
- deformed
- depressed
- depression
- disappoint
- discomfort
- disgusted
- failure
- fat
- fever
- filth
- flabby
- foul
- funeral
- garbage
- germs
- gloom
- grief
- hardship
- headache
- hell
- helpless
- illness
- impotent
- infection
- lice
- loneliness
- lonely
- loser
- louse
- maggot
- malaria
- massacre
- measles
- misery
- mistake
- morbid
- morgue
- mosquito
- neglect
- obesity
- paralysis
- penalty
- poverty
- pus
- rotten
- sad
- scum
- shamed
- sick
- slime
- slum
- stench
- stupid
- tomb
- trash
- ugly
- unhappy
- useless
- waste

Positive Valence High Arousal

- admired
- adventure
- affection
- aroused
- beautiful
- birthday
- brave
- car
- cash
- cheer
- christmas
- confident
- couple
- dazzle
- desire
- dollar
- ecstasy
- elated
- engaged
- erotic
- excitement
- exercise
- fame
- festive
- fireworks
- flirt
- fun
- gift
- graduate
- happy
- heart
- holiday
- intercourse
- intimate
- joke
- joy
- kiss
- laughter
- leader
- love
- loved
- lucky
- lust
- memories
- millionaire
- miracle
- mother
- orgasm
- outstanding
- party
- passion
- profit
- promotion
- rescue
- riches
- rollercoaster
- romantic
- sex
- sexy
- skijump
- song
- success
- sunlight
- surprised
- talent
- terrific
- thrill
- travel
- treasure
- triumphant
- valentine
- victory
- win

Positive Valence Low Arousal

- adorable
- agreement
- angel
- bath
- beauty
- bed
- bird
- bless
- blossom
- brother
- bunny
- butterfly
- cake
- capable
- carefree
- caress
- color
- comfort
- cozy
- cuddle
- devoted
- dignified
- earth
- easy
- easygoing
- elegant
- enjoyment
- family
- fantasy
- free
- friendly
- gentle
- grateful
- grin
- heal
- home
- honest
- house
- kind
- kindness
- loyal
- luxury
- masterful
- melody
- music
- nature
- ocean
- palace
- paradise
- peace
- pillow
- politeness
- protected
- rainbow
- refreshment
- respect
- respectful
- reward
- safe
- sailboat
- satisfied
- scholar
- secure
- sky
- sleep
- snuggle
- soft
- soothe
- spouse
- sun
- sunrise
- sunset
- truth
- twilight
- untroubled
- useful
- warmth
- wise
- wish