EFFECTS OF SIGNAL PROBABILITY ON MULTITASKING-BASED DISTRACTION IN DRIVING, CYBERATTACK & BATTLEFIELD SIMULATION

by

BEN D. SAWYER
B.S. Colorado State University, 2010
M.S.I.E. University of Central Florida, 2014

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Psychology
in the College of Science
at the University of Central Florida
Orlando, Florida

Fall Term
2015

Major Professor: P. A. Hancock

ABSTRACT

Multitasking-based failures of perception and action are the focus of much research in driving, where they are attributed to distraction. Similar failures occur in contexts where the construct of distraction is little used. Such narrow application was attributed to methodology which cannot precisely account for experimental variables in time and space, limiting distraction's conceptual portability to other contexts. An approach based upon vigilance methodology was forwarded as a solution, and highlighted a fundamental human performance question: Would increasing the signal probability (SP) of a secondary task increase associated performance, as is seen in the prevalence effect associated with vigilance tasks? Would it reduce associated performance, as is seen in driving distraction tasks? A series of experiments weighed these competing assumptions. In the first, a psychophysical task, analysis of accuracy and response data revealed an interaction between the number of concurrent tasks and SP of presented targets. The question was further tested in the applied contexts of driving, cyberattack and battlefield target decision-making. In line with previous prevalence effect inquiry, presentation of stimuli at higher SP led to higher accuracy. In line with existing distraction work, performance of higher numbers of concurrent tasks tended to elicit slower response times. In all experiments raising either number of concurrent tasks or SP of targets resulted in greater subjective workload, as measured by the NASA TLX, even when accompanied by improved accuracy. It would seem that "distraction" in previous experiments has been an aggregate effect including both delayed response time and prevalence-based accuracy effects. These findings support the view that superior experimental control of SP reveals nomothetic patterns of performance that allow better understanding and wider application of the distraction construct both within and in diverse contexts beyond driving.

This dissertation is dedicated to my father, Dr. John Michael Sawyer, Ph.D.,
and to my mother, Mrs. Marian White Sawyer, M.A.,
who taught me to love reading.

Thank you for your patience and perseverance in answering the endless questions
which naturally followed.

ACKNOWLEDGMENTS

I have had the kind assistance of so many.

The Sawyer Team of Research Assistants at the MIT^2 Laboratory
learned and operated complex systems, which allowed a project of this scope.

The United States Air Force Research Laboratories
The 711th Human Performance Wing appointed me twice a Repperger Research Scholar,
shaping these ideas. Battlefield Acoustics' BATMAN Laboratory provided financial support.

The United States Air Force Academy Behavioral Sciences Program
provided space, subjects and support for a dedicated battlefield attention simulator.

Dr. Victor Finomore
provided friendly and insightful mentorship throughout.

Dr. Joel Warm
impressed upon me the importance of vigilance, both as a research topic and my own life.

Drs. Mouloua, Szalma and Matthews, my doctoral committee members,
each, in person or writing, had guided me so well in the past that
I asked them to guide me here.

Dr. Peter Hancock, my doctoral advisor and chair of my doctoral committee,
at every step provided intellectual, editorial, material, and moral support.

My wife, Mrs. Szuhui Wu
learned to be a scientist alongside me, checked everything twice, and kept me laughing.

These and many others have made this dissertation a pleasure and a success.

Thank you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS (or) ACRONYMS

3D	Three-dimensional
ADD	Attention Deficit Disorder
AF	Air Force
ANOVA	Analysis of Variance
ANSI	American National Standards Institute
DSR	Device Status Reporting
EEG	Electroencephalogram
ER	Event Rate
ERP	Event Related Potential
ET	Email Testbed
FWSP	Familywise Signal Probability
GB	Gigabyte
GPS	Global Positioning System
HUD	Heads-Up Display
IED	Improvised Explosive Devices
ISI	Inter-stimuli Interval
KR	Knowledge of Results
LCD	Liquid Crystal Display
M	Mean
MANOVA	Multivariate Analysis of Variance
MPH	Miles per Hour

MRT	Multiple Resource Theory
NASA	National Aeronautics and Space Administration
PDF	Printer Definition File
RAM	Random Access Memory
RMSE	Root Mean Square Error
ROC	Receiver Operating Characteristic
SD	Standard Deviation
SMS	Short Message Service
SP	Signal Probability
SPS	Self-paced Search
SSD	Solid State Drive
TCD	Trans-Cranial Doppler
TLX	Task Load Index
VDL	Video Downlink

CHAPTER ONE: AN INTRODUCTION TO DISTRACTION

When, in the course of multitasking, performance decrements in one task are due to attentional allocation to a concurrent task, the condition can be viewed as a distraction from the first task. Current research into distraction tends to center on a single context: that is, distraction while driving (for a recent review see Caird, Johnston, Willness, Asbridge, & Steel, 2014, as well as Hancock, Mouloua, & Senders, 2008; Young, Lee, & Regan, 2008). From abundant findings which detail the dangers of roadway multitasking with mobile devices, the term 'distracted driving' has migrated into traffic law and popular use (Agnes, 2004). Scientific and popular attention to distracted drivers is easily justified, given the quantifiable risks incurred by engaging in a variety of in-vehicle tasks (Jerome, Ganey, Mouloua, & Hancock, 2002; Fitch et al., 2013; Sawyer, Finomore, Calvo, & Hancock, 2014a). What remains puzzling is the rather narrow focus of applied multitasking-based distraction research on this one context. Constructs such as situational awareness (Eriksen, 1995; Smith & Hancock, 1995) and workload (Hancock & Meshkati, 1988; Moray, 1979) are applied in a wide variety of contexts. It seems unreasonable that distraction should exist only in the task of driving.

There are some findings of effects outside driving that might arguably fall under the umbrella of distraction. Aviation research into pilot multitask use of heads-up displays while landing (Fischer & Haines, 1980; Wickens & Long, 1994) has revealed costs to, for example, detection of unexpected obstacles on the runway. Battlefield target detection (as in Yeh, Wickens, & Seagull, 1999) involves similar multitasking decrements; as more information is displayed to a dismounted combatant they become less likely to detect critical signals. In both of these situations performing multiple tasks leads to poorer performance, but none are consistently referred to as ‘distraction’ within each of the respective associated literatures. If distraction has the potential to occur in any multi-task situation, it so remains as yet unexplored in many domains.

Defining Distraction in the Context of Driving

Although not the seminal driving distraction work, ‘Driven to distraction’ (Strayer & Johnston, 2001) remains one of the most cited papers. In this work’s first experiment, of two conducted, these authors made use of a driving-representative tracking-task to compare baseline single-task driving to dual-task driving. Dual-task conditions included listening to the radio or engaging in a conversation on a cell-phone. In terms of response time, no difference was found between the baseline and the radio listening condition, however, a significant lengthening of response times were recorded for participants conversing on a cell phone. These findings suggested that something about responding to another human in conversation and/or generating language incurred greater driving detriment than simply listening. In Strayer and Johnston’s second experiment (2001), baseline single-task driving was compared to dual-task cell phone use while driving, with participants either generating their own conversation, or simply repeating the words of a confederate (i.e. ‘shadowing’). Participants drove continuously, and so rather than a discrete measure interrupting driving (such as the response time to a brake car used in experiment I), the continuous measure of root-mean-square error (RMSE) of steering was used. The experiment was performed on both an easy and difficult driving course. Over the easy course, the shadowing condition showed no significant differences from baseline driving. In contrast, over the difficult course all three conditions, baseline, shadowing and generation, illustrated a significant continuum of increasing impairment. These finding, it was suggested, showed that a limited resource was being directed at two tasks. Further, in the phone task the generation of language required more of that resource than listening alone. Performance in the driving task was suffering due to attentional allocation to these concurrent tasks, a state reported as "distraction from driving".

The design of most distraction studies allows a statistical assessment as to whether driving while using a device differs from baseline driving. In studies with multiple devices such analysis may reveal differences that allow a rank ordering of relative disturbance by dependent variable. This basic methodological strategy, the comparison of baseline driving performance to multi-task performance with one or more devices or circumstances, has served as the template for many of the additional studies that have followed. Exemplar findings from this body of research include the effect's extension to multitasking with devices beyond vocal use of cellular phones, including in-vehicle entertainment (Chisholm, Caird, & Lockhart, 2008; Mouloua, Hancock, Rinalducci, & Brill, 2003), food consumption (Young, Mahfoud, Walker, Jenkins, & Stanton, 2008), text messaging (Drews, Yazdani, Godfrey, Cooper, & Strayer, 2009; Sawyer, 2010; Sawyer & Hancock, 2012), and even the use of next-gen heads-up displays (HUDs, Sawyer et al., 2014a). Comparisons of the severity of detriment from individual devices have also been conducted (as in Jerome et al., 2002; Strayer et al., 2013; Strayer, Turrill, Coleman, Ortiz, & Cooper, 2014). The applied implication has been, and continues to be, that cellular phones represent only one example in a diverse set of roadway distractions.

"Thieves of attention" (Hancock et al., 2008) transiently disengage the operator of a vehicle from solely engaging in the role of 'driver' into other social roles. Billboards, for example, invite the role of 'consumer', children invite the role of 'parent', and digital communications devices, invite a veritable Pandora's Box of roles, social and otherwise, into the vehicle. Of course, even ‘driver’, as a role, entails a number of important tasks beyond driving itself. These include ‘navigator’, ‘climate control specialist’, and ‘he who must remember milk’. However most, if not essentially all, driving research identifies driving itself as the ‘primary’ task, and so compares baseline single-task performance to performance while engaged in one or more ‘secondary’ tasks. This primacy, ontologically speaking, is a state assigned by the researcher. Explicit instructions may be desirable, as primacy may not be obvious to participants in the laboratory (Dressel & Atchley, 2008) and indeed on the road (Fitch et al., 2013). Such transitions may even be induced; some attentional ‘thieves’ rely on a driver’s propensity to switch roles. Roadside billboards, for example, would not exist if drivers never engaged with them (Hancock et al., 2008), and so distraction they cause (as in Dukic, Ahlstrom, Patten, Kettwich, & Kircher, 2013) can be said to be willfully engineered. Given that it is presently both legal and common to build such attractive distractive tasks, how is the word “distraction” to be understood?

Distraction etymologically derives from the Latin dis traho, to pull apart, and its meanings in English center upon the undesirable pulling of attention (as compared to, for example, the French distraction: a pleasurable diversion. See Room, Buchanan-Brown & Pratchett, 2000). “Distraction” thus identifies an objectionable channel for attention, and presupposes one preferable. The differentiator of these channels is clearly identified in the above scientific literature on driving distraction: potential resultant harm. The distracted driver, by attending to a distracting channel, is more likely to drive in a manner which risks a collision. Collateral negative impacts range from disruption of the traffic stream through financial damage up to the deaths of the driver and their victims. This argument is strongest in situations where no reasonable virtue can be assigned to the task (or tasks) beyond driving, with social phone and messaging use the most common example. Contrasting situations certainly exist. The idea of competing risk is well and humorously illustrated in Hancock, Mouloua, and Senders' (2008) with the example of the discovery of a venomous snake in the car. Weighing potential resultant harm in such a situation, a driver might rationally take up the task of "snake fighting" despite its obvious distractive influence on the driving task. An epidemiological argument for virtue in non-driving tasks can be found in the logic of Swedish opponents of legislation banning in-vehicle phone use, as the country became the last in the European Union to adopt such laws. These rightly point out that driving itself is a hazardous activity and that 'social messaging' can shorten overall driving time by, for example, sending a reminder that eliminates the need to drive back out for milk. In Sweden, a country with long rural routes, this reduction in exposure (as described in Kircher, Ahlström, Gregersen & Patten, 2013) is argued to outweigh other factors. Finally, situations can be found where driving is the distraction. In desert military campaigns roads can be long, straight, devoid of traffic, and vehicles driven highly robust to roadway departures. Arial reconnaissance means lethal hazards, in the form of improvised explosive devices (IED) or enemy combatants, may be relayed to drivers in advance, and specialized forces dispatched to resolve the situation. The communication channel in such situations is of primary importance, and any attention diverted from it increases the potential for resultant harm. In each of these situations, an assessment of potential risks is necessary to determine which task is primary and which secondary, which is distraction and which the human must not be distracted from. Therefore, it can be said that the definition of distraction is fluid by context, and that arguments based in an actuarial view of primacy are necessary for a researcher to apply the term.

Design, training, and legislation in driving, as in so many contexts, are filled with decisions in which life is lost on both sides of a balance. Such Faustian pacts are difficult to resolve scientifically, and so belong to the philosophical field of ethics, where they are often considered within a framework known as “the trolley problem” (for a recent treatment see Skulmowski, Bunge, Kaspar & Pipa, 2014). This thought experiment unfolds as follows: “a vehicle (trolley) is moving toward a group of helpless people (tied to the tracks), and you, a bystander, may pull a lever to divert it onto a track in which there are fewer helpless people (a single unaware worker).” Those that choose to pull the lever reduce the overall loss of life but become the agents of that smaller number of lives lost. Many variations have been posed since the original (XXX), manipulating, for example, age, sex, relationship to and distance from those that die, and each provides an interesting tool with which to understand trade-offs which involve life. In this tradition, and pursuant to the present discussion, I pose a variant thought experiment: “the distracted trolley”. “Given the number of serious reported trolley accidents, an operator now sits in each trolley, ready to pull a lever and divert the vehicle to a side track. Monitoring the side track and the main track is difficult, and so a technology has been developed with detects people on the side tracks. When placed in the trolley, this technology naturally requires the operator’s eyes and mind to be regularly removed from the main track. Is the technology a distraction? Should it nonetheless be allowed in the trolley?” In order to provide a solid argument one way or the other, there is a need for information. Factors include the predicted number of helpless people tied to the main track, number of unaware workers on side tracks. Consider the operator’s ability, the technology’s efficacy, and the efficacy of the human-machine system they comprise, both in terms of the main track and side tracks. It is the collection and interpretation of such data that have driven the previously detailed growth in the area of driver distraction.

The distracted trolley is real, and real world “operators” and “track-dwellers” alike have elevated the term ‘driver distraction’ into law, and from there into common English discourse, where it became Webster's' 'Word of the Year' (Agnes, 2004). Primacy, and a resultant view of just what constitutes distraction, may be clear in the case of social test-messaging but become less so in the case of text-messaging professionals. Consider the data delivered to an ambulance driver, which may simultaneously increase potential resultant harm on-road and reduce potential resultant harm at a nearby medical emergency. Imagine the tracks, drivers surrounding the ambulance on one, a man with a coronary infarction on the other. Should the technology be allowed in the ambulance/trolley? Arguments can be made both ways. Arguments *should* be made both ways. Design, training, and legislation resulting from the argument will result in life being lost on both sides of the balance. It is through thought experiments like "the distracted trolley" that an an actuarial view of primacy may be formed, and exported.

The widespread impact of the findings generated by driving distraction inquiry has not been duplicated in non-driving domains. In order to better understand the success of distraction studies in the context of driving, and great challenges of this construct in other contexts, it is important to gain a broader view of what distraction is. As such, consider the following survey of the psychological constructs and theories underlying the construct. This is presented with an eye both to which existing bodies of knowledge are already used in distraction studies, and which have potential to be used. Throughout, the question of why this useful and impactful construct, distraction, remains widely unexploited outside a single context, driving, will remain central.

A Generalized View of Multitasking-based Attention and Distraction

William James said of attention "it is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought. Focalisation (sic), concentration of consciousness are of its essence. It implies withdrawal from some things in order to deal effectively with others, and is a condition which has a real opposite in the confused, dazed, scatterbrained state…”. The psychological study of distraction is necessarily closely tied to that of attention, and both predates and extends beyond driving research. In sexual research, distraction is studied relative to its ability to delay or prevent orgasm (as in Geer & Fuhr, 1976). In this same vein, a line of research into the ability of distraction to reduce or prevent pain also exists (as in McCaul & Malott, 1984). Auditory distraction in the workplace has been described (see Beaman, 2005 for a review), and details loss of efficiency and increases in training time. Clinical deficit study of distraction centers on disorders of attention and executive function, including attention deficit disorder (ADD) (as detailed in a seminal piece Driven to Distraction by Hallowell & Ratey, 2011, which shares its name with a seminal driving paper). In applied psychology the word distraction is applied to similar dysfunction, albeit with a different root cause. Distraction here is undesirable departures from perception and action directed toward desirable tasks. The shift, therefore, is one toward perception and action relevant to less desirable tasks, and a neglect of the primary task.

Neglect of a task does not always mean it stops; few of life’s activities occur in the vacuum of competition referred to as ‘single-tasking’. Individuals proverbially ‘walk and chew bubblegum’ through life, multitasking in myriad task and subtask combinations. Such rapid, autonomous skill execution is the hallmark of experienced multiple-task performance (Fitts & Posner, 1967). Learning to drive is an excellent example of such skill acquisition (Anderson, 1983), as the task driving is itself a constellation of a series of subtasks. When first learning, rapt attention must be given to each, and novice drivers must be taught sequences of subtask operations by rote. In short order, however, drivers find these previously cumbersome combinations so easy as to engage in a complex secondary tasks, like messaging. Their confidence notwithstanding, as shown by Strayer and Johnston in the context of driving (2001), resources available for such concurrent performance of tasks are limited. Drivers, and indeed all who seek to conserve limited cognitive resources, often perform not to the best of their ability, but to the minimum requirements of the present situation. This act is known as satisficing (Simon, 1969), and further frees resources for peripheral tasks. When attention is divided between tasks, resource supply and demand is the arbiter of performance. Within said divide, stimuli must actually be available for perception and action; situations which remove or occlude stimuli are not distraction (Hancock et al., 2008). However, divided attention can result in failure to perceive available stimuli, as in change blindness, in which attending closely to one set of stimuli renders another set of clearly apparent stimuli cognitively ‘invisible’ (Simons & Ambinder, 2005). The above ideas of divided attention and multitasking are closely related, with the primary difference between the two being the locus of inference. When an individual engages in more than one task, the result is multitasking, whereas when an individual has their attention divided across multiple stimuli, the result is divided attention. For clarity, the term multitasking has been used preferentially in the present work.

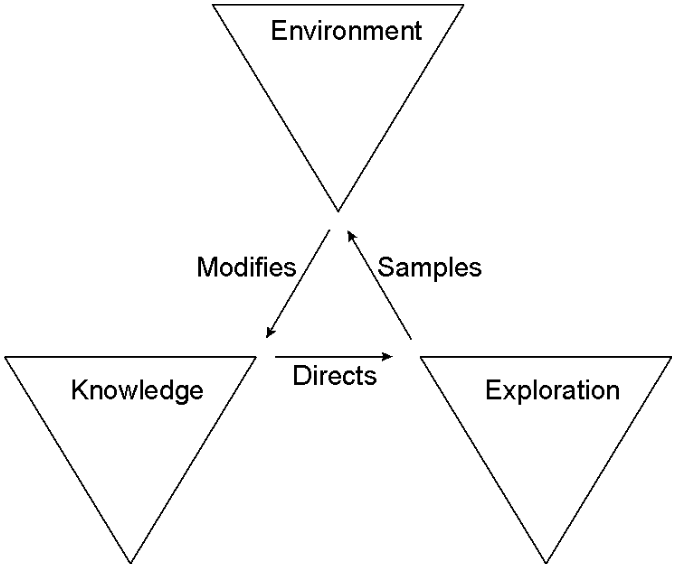


Figure 1: The perception-action cycle
The perception-action cycle (from Neisser, 1976) describes a loop of sampling the environment through sensory organs, using that information to modify knowledge of the world, and responding by directing or withholding action. Resultant changes to the environment require re-sampling, which in turn starts the cycle again.

Multitasking-based distraction should first be considered within the framework of the perception-action cycle (Figure 1). An organism samples the environment through sensory organs, uses that information to modify knowledge of the world, responds by directing or withholding action, and samples the impact on the environment which in turn starts the cycle again (Neisser, 1976; Smith & Hancock, 1995). While the organism may make interactive changes, it is also able to react to changes stemming from outside agents or forces; it is changes within the environment which are the impetus for re-engaging in the cycle (Gibson, E., 1969; Gibson, J., 1979). Such capabilities are underwritten by anatomical organization; humans share with other organisms a functional dichotomy of structure in which sensory (perception) nerves are largely attached to the anterior nerve axis while motor (action) nerves are largely attached to the posterior (Betz, 1874). As such, it is unsurprising that the perception-action cycle underlies diverse psychological constructs, such as attention, memory, situational awareness, and distraction. Multitasking-based distraction, in this light, might be framed as concurrent performance of tasks requiring multiple concurrent perception-action cycles, each of which becomes more likely to experience delay and/or failure. With such complexity available, literally at our fingertips, it is surprising that the cycles within cycles do not more often end in confusion and failure. Certainly, observers must have protections against overindulging in the sheer volume of perception and action available in the environment around them.

Modeling Theft of Attention

Observers are able, and sometimes unable, to avoid being overwhelmed by the variety and complexity their world. Consider perception of, and action in response to, a complex visual scene: objects in the real world reflect light to the eyes of an observer, where it is encoded preliminarily at the retinas. The resultant information is transmitted to the brain, processed, and then informs decision-making to guide action. Human observers lack infinite processing capacity, and so must have the ability to limit processing demand. This inferred existence of a protective mechanism is the basis of the filter model of attention (Broadbent, 1958), which proposes that the brain filters sensory information at an early stage of processing so as to focus only on necessary stimuli. While this model successfully generalizes to many situations, it cannot explain the ability of humans to recognize and attend to irrelevant information streams when they become suddenly relevant, for example attending to one’s own name in a previously ‘filtered’ conversation (Moray, 1959), or presumably identifying a string of sudden brake lights in a roadway lane adjoining the one of travel. If relevant, timely semantic cues can be extracted from supposedly ‘filtered’ stimuli, it seems likely that any filtering must occur far enough into processing to extract this meaning (Deutsch & Deutsch, 1963).

This debate, surrounding early selection and late selection filter models, has only begun to cool relatively recently, in part due to the introduction of new hybrid models that account for aspects of each. Load theory (Lavie, 1995; 2005; 2010), suggested that protective filter mechanisms only come into effect when processing capacity is in danger of being reached or exceeded. Load theory relies on the concept of executive control to explain how observers choose which channels are attended to in the absence of spare capacity. Such a theory for control of attention in overload situations is close, in terms of definition, to that previously offered for distraction. In evaluating the applicability of load theory to distraction, it is worth considering the most common methodology used in this paradigm. Speeded discrimination tasks using simple (i.e. letter, basic shape) stimuli are manipulated by the addition of irrelevant, non-critical signals introduced in the periphery. The influence of a ‘distractor’ is then measured in terms of response time decrement (Lavie, 1995). Elements of this design match well with driving distraction work, in which a discrimination task (identifying brake lights) is responsible for alerting the driver to engage in a speed sensitive response task, and a distractor (secondary task) is introduced in the periphery. Load theory would suggest that at higher levels of overall workload, distractors exert less influence. Indeed, when participants are required to make more complex discriminations (features only vs a multiple conjunction search), distractive decrement are reduced. This effect holds even when distractors are complex, as in the case of the introduction of cartoon characters by Beck and Lavie (2005), or when distractions are internal, as in the ‘mind wandering’ reports studied by Forster and Lavie (2011). In each case higher load tasks remain less affected by distractive stimuli.

These findings are interesting in the context of driving, a task with long periods of relatively low demand punctuated with brief periods of very high demand (see Miyake, Hancock, & Manning, 1992). This so called “hours of boredom, moments of terror” task landscape is shared with many other domains (Hancock, 1997), and in conjunction with the findings from load theory suggest the following questions. Is the low workload majority of the time spent at such ‘boredom and terror’ tasks also the time when individuals are most susceptible to distractions? How quickly, during a transition from boredom to terror, do high-load protections from distractive influences become available? Once the transition has occurred, how long does the state last? Such a transition seems strikingly close to the idea of hysteresis, in which load from a task continues to affect performance on a second task even after the first ends. Further, when the line between any momentary distractor and the nominally vital task is not always clear, might such fixation be not protective, but a problem? Consider the phenomenon of cognitive tunneling, in which fixation on one task prevents engagement with another necessary task (Dirkin, 1983; Dirkin & Hancock, 1985; Thomas & Wickens, 2001), a cases where a desirable transition may be slow or simply fail to occur. Load theory, then, suggests one explanation for driving distraction. When a secondary in-vehicle task becomes primary, even briefly, the driver becomes less sensitive to stimuli in the roadway environment. As long as load remains so elevated, the transition back to the driving task may be inhibited.

Load theory, as a form of basic experimental work, is nominally unconcerned with addressing such applied realities, but does suggest applicability to known applied phenomenon. When multiple tasks are undertaken simultaneously, success of the overall effort is defined by the ability to smoothly maintain and transition between all subtasks. This may be analogous to the ability to abandon interaction with a distractive influence and move back to a critical task.

However, it also suggests that at high load levels individuals may be less likely to engage with distractive influences. This may underlie the phenomenon seen in driving of slowing when under high workload from dual-task device use (as in Törnros & Bolling, 2006; Sawyer & Hancock, 2013; Sawyer et al., 2014a). Drivers may actually be reducing load until they are able to engage in the secondary task effectively, essentially satisficing (Simon, 1969) until capable of multitasking. More generally, the above findings in sum are yet another affirmation that effort level and collateral resource use are vital considerations in multitasking-based distraction.

Effort and Risk of Failure

The effort involved in a task can be said to be directly related to the toll it takes on resources in order to meet demands, and can be referred to as workload. Workload is primarily collected through subjective self-report measures. Among the most popular of these is the NASA task load index (NASA-TLX, Hart & Staveland, 1988), which provides reliable assessment across six dimensions of workload. Effort, performance, frustration, mental demand, physical demand and temporal demand may be assessed individually, or averaged to provide a composite workload score. The NASA-TLX also contains a series of forced-choice questions identifying the importance of the contribution of each component. In practice such questions are often omitted due to experiment time constraints. It is notable that the NASA-TLX was validated in a variety of multitasking situations. These include a) similar concurrent tasks, b) "Fittsberg" tasks, in which the output of a memory search task (Sternberg, 1969) provides the input for a sorting task (Fitts & Peterson, 1964), and c) full multitask, multimodal supervisory control (Hart & Staveland, 1988). The NASA-TLX remains the most used self-report workload technique, and is generally well-regarded in the human factors community and beyond. Nonetheless, it suffers from potential shortfalls that can be levied against any subjective scale; namely whether respondents are willing or indeed able to accurately report their own internal states and experiences (Natsoulas, 1967), and whether retrospective reports are more susceptible to memory concerns. This line of inquiry rapidly reverts to a philosophical debate on functionalism (see Block, 1980 for contrasts with physicalism), but also highlights a number of related practical concerns. Tasks competing for similar resources require more effort to perform in tandem, and so adding a subjective assessment technique in real time invites failure of the task or, speculatively, in the meta-cognition needed for accurate self-assessment.

In addressing this issue, various attempts at passive measurement of workload through physiological measures have been attempted. In general, these measures attempt to quantify arousal directly, as in electroencephalographic (EEG) monitoring (Hancock & Szalma, 2003), or by proxy through measuring resource use, as in trans-cranial doppler (TCD) (Shaw et al., 2009) or heart rate variability (Veltman & Gaillard, 1996). Although many such techniques show promising patterns, these tools are relatively new. Different techniques tend to show different results, and all have varying degrees of agreement with established measures like the TLX (Hancock & Szalma, 2003). What is more, not all task combinations produce the same level of workload. Notably, spanning information across several sensory modalities has long been shown to moderate overall workload, as compared to moving the same information across a single modality (Wickens, 2002).

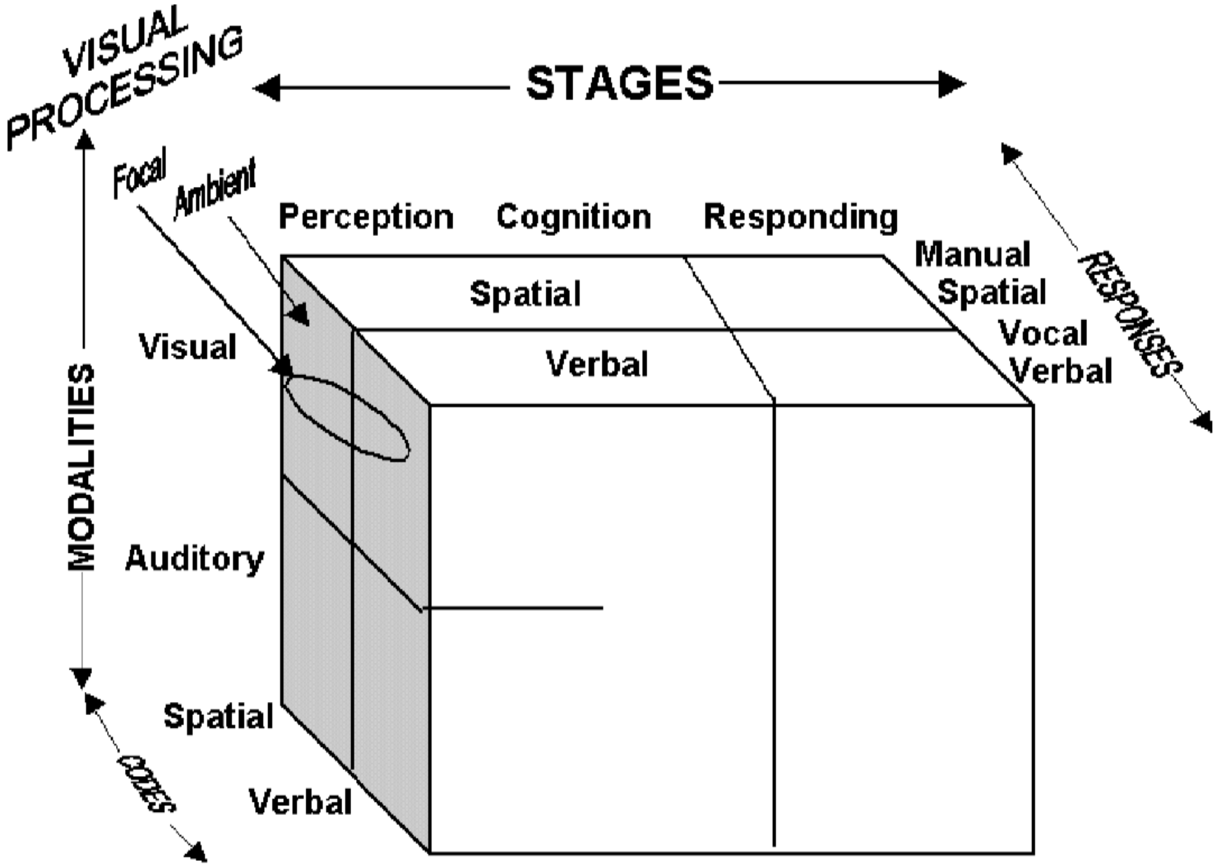


Figure 2: Wickens' multiple resource theory (MRT)
Wickens' (2002) 'box model', multiple resource theory (MRT) presented visually, represents the potential outlets for cognitive resources, and identifies overlaps where resource use may be exacerbated. For example the model indicates that driving and text messaging, both primarily visual and spatial tasks, would require more resources when performed in tandem than driving and listening to audio books, an auditory and verbal task. It is notable that in 'stages' perception and cognition occupy the same space.

To understand the concurrent performance of tasks and subtasks across modalities as well as the perception-action cycle, one popular model is multiple resource theory (MRT, see Wickens, 2002). Wickens' box model (Figure 2) builds off evidence that cognitive work relies upon differential resources which are limited in nature, and may be fluidly redirected to multiple tasks. In multitasking, such direction of resources is not a zero-sum (as in Von Neumann, 1953).

Tasks which draw on similar resources are anticipated to produce greater interference, leading to reduced performance and enhanced likelihood of task failure. Notably, MRT as a model combines the perception-action cycle stages of perception and cognition together. This competition can include gross physical or “structural” limitations. Foveal fixations, for example, are constrained by the physical speed with which the six extraocular muscles can throw and catch their organ. The competition can also include cognitive processing, likely constrained by the rate at which the body can provide blood, oxygen, and thus energy, to the brain. Under such constraints a compound task like texting-while-driving might be modeled as two concurrent visio-manual tasks with physical and manual response elements, which compete for a limited pool of resources. As resources are exhausted, failures of perception or response result.

A debate exists as to whether a) perception-based visio-manual structural interference or b) working memory and processing based cognitive interference is the larger contributor to driving distraction and associated workload (Strayer & Johnston, 2001). A growing body of empirical work presently shows that, in driving, cognitive factors are involved in the lion’s share of distraction-related elevations in workload and detriment (see He, McCarley, & Kramer, 2013; Sawyer et al., 2014a). However, this support of the cognitive interference hypothesis is not unanimous, and it is indeed difficult to argue that failing to look at the road is not detrimental to safe passage (McCallum, Campbell, Richman, Brown, & Wiese, 2004; Owens, McLaughlin, & Sudweeks, 2011). Structural interference undoubtedly does contribute to overall workload elevation, but it should also be noted that it is more easily observed. Psychology exists now a scant half-century from a time when consideration of ‘internal processes’ was a taboo subject precisely because evidence of their existence cannot be directly observed (Skinner, 1977). Early inquiries literally ‘showed’ structural interference and it seems likely that such readily observable phenomenon fueled an availability heuristic-based (Tversky & Kahneman, 1973) erroneous perception that it is the major contributor. Such a perception is not limited to science. Hands-free headsets or dashboard systems do little or nothing to reduce the detriment of talking and driving (Strayer & Johnston, 2001; Backer-Grøndahl & Sagberg, 2011), and yet they are allowed on military bases, and on the roads of the majority of states (Ibrahim, Anderson, Burris, & Wagenaar, 2011). Such policy is the very manifestation of availability heuristic-based erroneous acceptance of the structural interference hypothesis. Still, the limited driving distraction policy that does exist is a reflection of the consensus between these two theoretical views. Both the cognitive and structural interference perspectives support the underlying concept that multitasking effort is resource driven, limited, and can lead to failure.

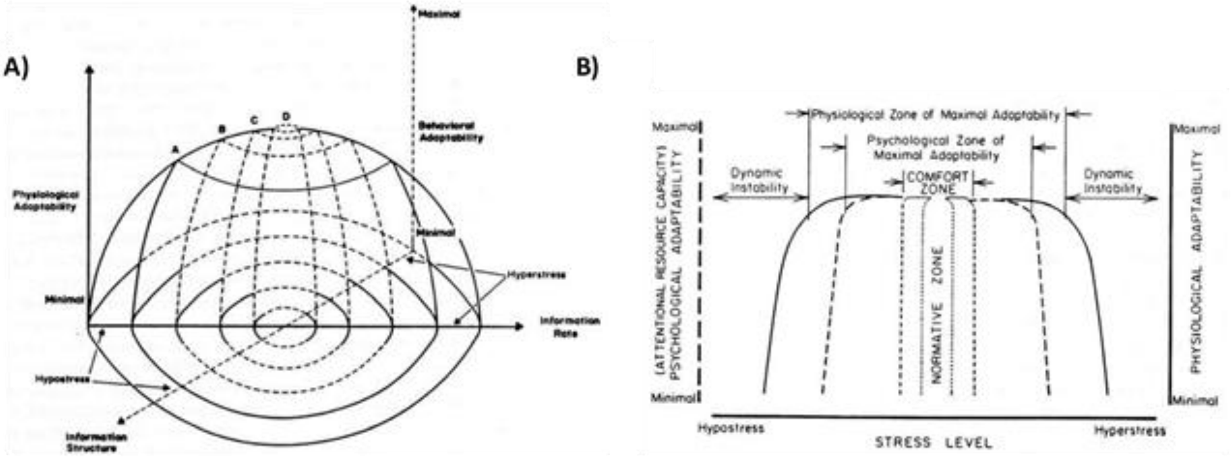


Figure 3: The Hancock-Warm model
The Hancock-Warm model (1989), shown in A) three dimensions with the constructs of Information Rate and Information Structure as dual base axes. In B) two dimensions the region of comfort, which is the core of the left-hand model, is revealed to give way to adaptation, dynamic instability, and failure through either hypostress or hyperstress.

Each of the above ideas addresses factors that might influence failure, and so the study of failure is worth addressing here. The present preeminent model is that developed by Hancock and Warm (1989), who examined failures in the context of sustained attention, but produced a model that is broadly applicable to many tasks where attention mediates performance (Figure 3). As individuals navigate the environment, dynamic change provides challenges which threaten the goal at hand. The model is rooted in the concept of homeostasis, a trait passed down from the thermal stress models preceding it (Hancock, 1986), and describes a progression from stability to either overload or underload states. Failure occurs upon exhaustion of resources required to counter such adverse environmental pressures, and is not instantaneous. Instead it is described (Hancock & Warm, 1989) as “dynamic instability”, a progressive inability to respond adaptively which results in performance dropping ever more rapidly until a state of complete functional failure is reached. This nomothetic description of failure does not preclude individual differences, and indeed such idiographic variations are expected (for a discussion of nomothetic and idiographic patterns, see Cone, 1986). Still, in situations of high load when resources are at a premium, available response strategies to avoid failure dwindle, then disappear.

Time, and Stopping in It

In situations where task failure has become a possibility, what strategies can minimize the chance of such an unfavorable ending? Clues can be found by comparing the ideas of the Hancock and Warm model (1989) with the findings of load theory (Lavie, 2010), in which under high levels of load distractors exert reduced influence. In both sets of findings, as load on the cognitive system increases so variation of response decreases. Ideally, the response chosen will successfully lead to a reduction in load, and a return to stability. Indeed, in situations where a ‘Moment of Terror’ (Hancock, 1997) is successfully navigated, the response chosen can be said to have been sufficiently appropriate, and the transition to engaging in it to have occurred in a sufficiently timely manner. The goal of systems deployed into environments where such high load crisis may occur should be to enable such timely and correct responses. In the system of law, driving-while-multitasking bans (Ibrahim et al., 2011) attempt to facilitate timely and correct response by prohibiting known dangerous concurrent tasks (i.e. messaging and driving). In the context of civilian motor-vehicle operation, this kind of prohibitory message is appropriate, if potentially not sufficient. Lawbreaking aside, the earlier discussion of primacy and “the distracted trolley” highlights contexts in which failing to multitask is not the rational or ethical choice. Emergency vehicle operators and those engaged in military operations rely on concurrent tasks for the mission as a whole to succeed. Drunk drivers may be court ordered to use distractive devices (Sawyer & Hancock, 2014). Where prohibition is not appropriate, design interventions must instead minimize risk, again by enabling timely and correct responses. Resultant interfaces would, ideally, preserve workload in the course of multitasking and afford smooth transitions to appropriate “final responses”.

How does one build interfaces which do not simply reduce the number of concurrent tasks, but provide superior information delivery to reduce the duration, frequency and resource cost of tasks engagement? One answer can be found in dividing ‘single tasks’ such as messaging into representative constellations of a series of subtasks, each corresponding to an interface element. Each is a potential affordance, and each individual contribution builds the aggregate distraction potential. As such, in tasks that can be engineered there is a need to understand the available components. Indeed, in a macro sense, the field of driving distraction is engaged in such an endeavor even now, with studies isolating the potential gains of using voice recognition (He et al., 2013; Strayer et al., 2014a), HUDs (Sawyer et al., 2014a; Tippey, Sivaraj, Ardoin, Roady, & Ferris, 2014), and other new interface one comparison at a time. The arduous nature of such undertakings aside, there is one cause for concern with the analysis of in-vehicle interface as described: time.

When measuring discrete roadway events (as in Sawyer & Hancock, 2013; 2014; Drews et al., 2009; and many of the works described in Caird et al., 2014), participant free will creates difficulties with precise temporal placement of a treatment relative to a measured behavior. How can a researcher ever be certain the use of a potentially distracting roadway device will coincide with the measurement of a dependent variable when the participant can do whatever they wish? A number of strategies can be used to compensate. For example, experimenters can flood the duration of the experiment with the manipulation (Figure 4A), under the correct assumption that when a participant never stops performing the task, the issue of temporal placement becomes less of an issue, or at least a less obvious issue. In continuous manipulations, for example conversational vocal cell phone use (as in Strayer, Drews & Johnston, 2003), this strategy is arguably quite a naturalistic one, although not representative of many shorter vocal communications. In a discrete manipulations, such as delivery of messages (as in Sawyer & Hancock, 2013; Sawyer et al., 2014a; for a review of such studies see Caird et al., 2014) flooding is less optimal. For example, Drews and colleagues (2009), presented 42 brake events at "freeway speeds". Sawyer and Hancock (2013) sent text messages continuously, a new one delivered as soon as the previous message had been responded to, while a single brake event was measured. It is difficult to argue that either situation proves to be a common handling of tasks within time on the road. Further, the amount that participants indulge in a manipulation may be affected by the participants' abilities, as in faster text-messaging typists, or by the affordances of a given experiment. Some studies code matches between messaging activity and response maneuvers, and then analyze (Drews et al., 2009), giving the relative levels of activity in which participants chose to indulge. Still, results of this sort still arguably lack strong experimental control. Replication, then, becomes problematic.

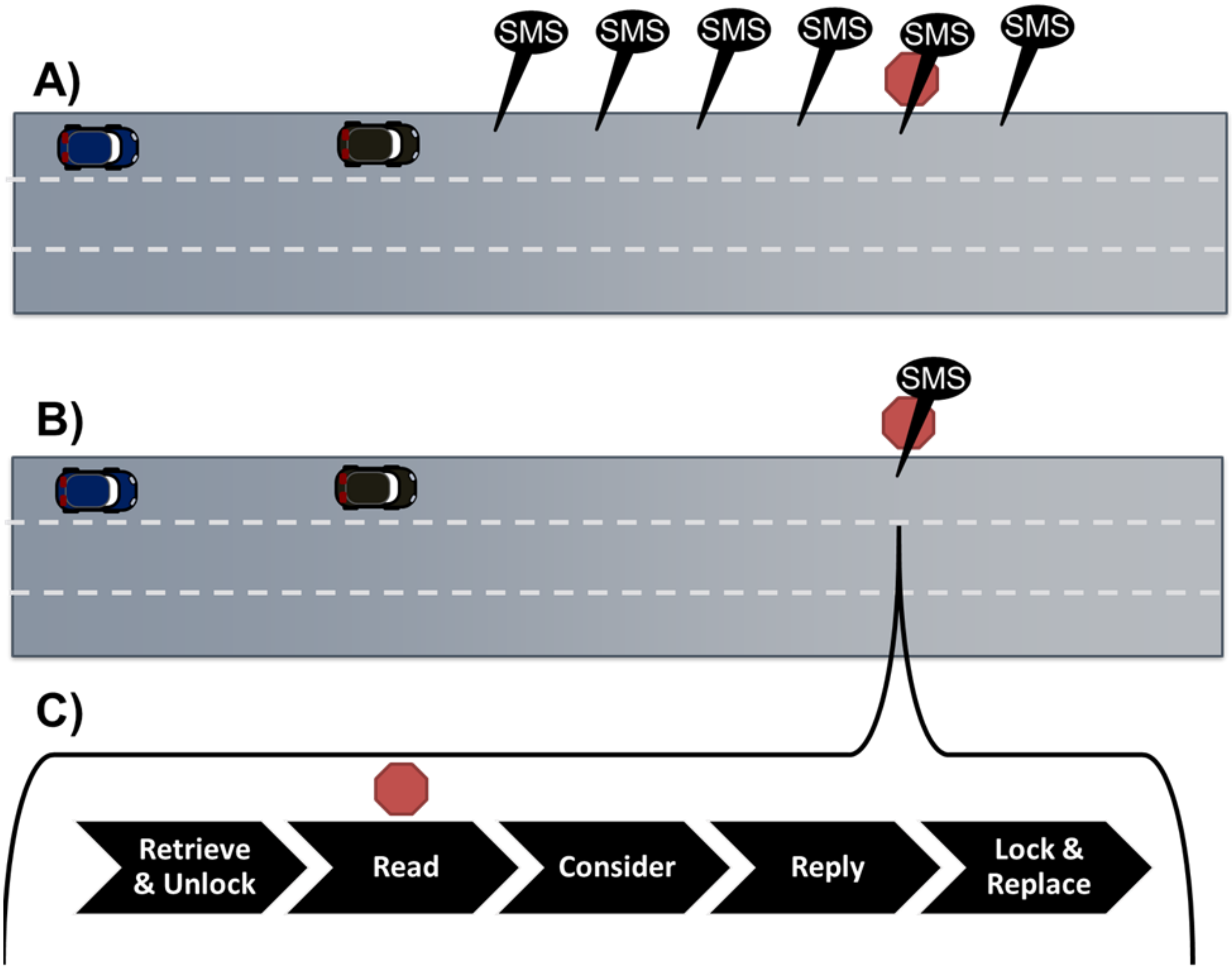


Figure 4: A driving study utilizing epoch-based methods
A) A pace car experiment, in which the participant’s car (blue, left) follows a computer controlled lead car (black, right) which brakes at predetermined points seemingly random to the participant. At the same time the participant is sent messages (here, SMS text messages). In such a paradigm, the free will of the participant makes it impossible to ensure consistent temporal placement of the message relative to the brake event (in red, Figure A shows such an alignments, but note that it is in no way guaranteed). In order to get some of the treatment effect (messaging) to overlap the event producing the dependent variable (brake event) the environment may be flooded with one or the other. This can lead to suboptimal experimental procedure that deviates considerably from real-world situations. B) Device status reporting (DSR) allows the precise targeting of a brake event contingent upon user interaction with a device. In the described project (Sawyer et al., 2014a) a user unlocked the phone and 1800ms later the leading pace car braked sharply. C) In the present example a brake event is targeted to time when the participant is

reading, but the system could likewise be used to trigger actions in the simulation relative to any user interaction event.

In discussing such tactics, consider the idea of hysteresis (as in Morgan & Hancock, 2011), in which load from an event continues to affect performance for a span of time. Flooding tactics rely upon hysteresis effects to ‘smooth out’ load, which is likely effective but can ignore or even mask the concentration of the manipulation within the time of the experiment. By way of metaphor, consider the validity of a drug study in which all participants are administered a five drug cocktail. Further assume that dosage is not controlled; availability of the drugs and unmonitored propensity of participants to indulge decides dosage levels. The consideration that each drug has a duration of effect is not an argument for experimental control. While naturalistic studies must contend with such issues, in an experimental setting such lack of control is suboptimal, greatly limiting the useful information such a study might provide. In driving distraction studies, a similar lack of precision creates issues with interpretation. Larger trends, such as the detrimental effects of multitasking, can be compared between studies (as in Caird et al., 2014), but the finer details that might allow the construction of better interface are less replicable. Hysteresis should be studied in its own right, not used to mask the effects of poor independent variable control.

A different approach was used by Sawyer and colleagues (2014a) when comparing Google Glass and an Android Smartphone. This project ‘built’ the serendipity to precisely overlap an on-road event with an in-vehicle task, a kind of coincidence that had in past work been created through flooding (as in Drews et al, 2009, but see Hancock & De Ridder, 2003 for a previous example). The technique was dubbed device status reporting (DSR, see Figure 4 as well as Sawyer, Calvo, Finomore, & Hancock, 2015a). In it, the transmission of timestamps for any given action on a device transmitted throughout the system allowed stimuli, such as a brake event from a pace car, to be accurately targeted within a specific component of the interface (Figure 4C). As a message was received the driver would 1) use a head-nod gesture to unlock the device, 2) read the message on a screen presented above forward vision, 3) consider a reply, 4) issue a voice command to reply and dictate the reply, 5) wait for transcription, check the message for accuracy, and wait for the device to lock. In the experiment, a brake event was targeted to component 2, reading. Analysis technique was borrowed from event related potential (ERP) research (see Luck, 2014; Sawyer, 2014). Data was analyzed by defining epochs, time windows for each participant starting at time-locked responses measured by DSR or the time synchronized simulator's data collection. By looking at levels of dependent variables within these time-locked windows, the contribution of the created high load dual-task event to driving and recovery could be evaluated.

To drivers so caught reading a message during a roadway brake event, Glass did provide some benefits over the use of a smartphone, although mainly in terms of recovery, as analyzed by this epoch method. In terms of discrete response time, participants showed no significant differences (Sawyer et al., 2014a). Despite a novel 'heads-up' display, it appears that Google Glass does not provide reading drivers protections against the slowed response of multitasking. In the present example, it is easy to see how another interface affordance, for example the head-nod gesture, could be compared to the smartphone unlock-gesture equivalent. The use of DSR triggering of events to target subtasks reveals new possibilities in terms of managing and recording time within driving distraction experiments. This strategy, however, does not directly address how to handle time in research design and analysis.

In many contexts, there is a need to build interfaces that allow operators to successfully navigate high-load multitask situations. Strayer and Johnston's (2001) experiments contain seeds of the successes and challenges of fourteen years, to date, of such inquiry. Their methodology provides a strong tool to look at the broad patterns of workload that stem from engaging in complex tasks alongside one vital and itself complex, driving. However, as the field turns toward more nuanced questions, more precise tools are needed. Without such, experimental tasks conducted in the laboratory may have little advantage, in terms of experimental control, over naturalistic work (such as Fitch et al., 2013). It is important to further consider that such lack of control may be why the construct of distraction, popular in the context of driving, has made so little headway in other domains.

What is needed, then, is a framework which can account for time, and the relative concentration of experimental variables within it. This framework must necessarily account for space as well: in many tasks naturalistic movement through time entails movement through space. In experimental stimuli, the 'terrain' of such space/time may necessarily be quite different from that encountered in naturalistic settings. It is impossible to predict what creative solutions will be found by researchers in the abstraction of applied tasks. Present driving distraction methodology allows great flexibility in addressing such experimental issues, and therefore any candidate framework should as well. It further seems unwise to build some new, untested standard for driving distraction work, which would likely never be used outside this work (see Figure 5). The goal, therefore, will be to build on a known and accepted technique. Ideally, advantages should be paired with backward compatibility to methodology used in (and methodologists used to) existing distraction experimentation.

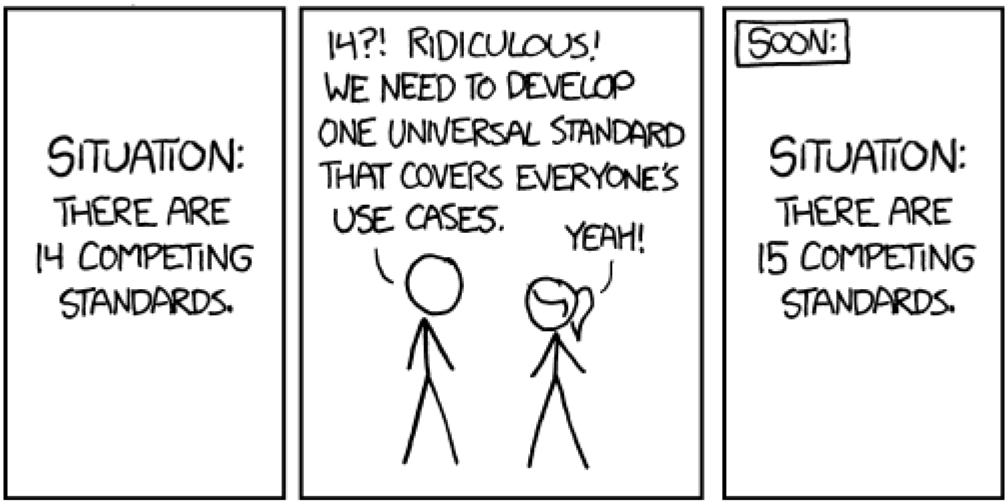


Figure 5: The problem with new standards

The goal of the present work is not to build a new standard for driving distraction, but rather to appropriate one already in use and apply it to driving distraction. The resultant cross-pollination would ideally result in a stronger methodological framework for driving distraction work, while providing distraction as a construct for use in more diverse contexts (XKCD, 2011).

CHAPTER TWO: TOWARD A FRAMEWORK FOR DISTRACTION STUDY

The present empirical question is whether a superior alternative to existing methods of evaluating multitasking-borne performance decrements can be identified. This is the task attacked here. Any alternative must offer a) a framework which can accurately specify space/time and the relative concentrations of experimental variables within and b) integration with existing methodologies, to include equivalent flexibility in supporting self-determined driver behavior. Ideally, a solution will incorporate an existing methodological standard, and will support a wide range of experimental designs. As such, the effort will first look to widely used, versatile experimental standards outside driver distraction research.

Vigilant Attention

In light of the above requirements, one particularly promising area is vigilance, also referred to as vigilant attention or sustained attention (Warm & Jerison, 1984; Hancock, 2013, for an excellent review see Warm, Finomore, Vidulick, & Funke, 2015). This line of research has been heavily influenced by the methods and principles developed by Norman Mackworth, an experimental psychologist investigating failures in radar operator attention for The Royal Air Force (Mackworth, 1948; 1950). His “Clock Task” looked to conceptually replicate the operational requirements of such wartime observers. In the clock task, participants viewed a blank timepiece with only a single hand, which advanced in one second increments. At apparently random times the hand would jump two increments, and it was this double jump which constituted a critical signal to which participants were asked to respond. Mackworth’s study revealed vigilance as a particular form of observational task differentiated by a strong performance decrement over time. Mackworth found that the rate of events (event rate or ER) retarded detection while speed of response also decreased (Mackworth, 1969; Guralnick, 1973; Parasuraman, 1979). As the probability of any given event containing a critical signal, or signal probability (SP), declined, the result was also retarded detection rates and decreased speed of detection (Mackworth, 1970; Warm & Jerison, 1984). In the worst-case scenario of a low SP, high ER task, operator speed and accuracy can quickly dip to dangerous levels of performance, instability, and failure (as in Hancock & Warm, 1989). This danger, and various strategies to mitigate it, have been studied in diverse contexts including aviation, medical monitoring, cyber-defense, and driving (Warm et al., 2015; Sawyer, Finomore, Funke & Warm, 2014b; and see Matthews & Desmond, 2002). Across such mitigation work, cuing (Hitchcock et al., 2003) and knowledge of results (KR, Hitchcock, Dember, Warm, Moroney, & See, 1999) feature prominently among strategies that can improve outcomes. At present, this broad and well-developed body of inquiry has produced methodologies which have benefited from an early and prolonged attention to the roles of time and probability.

Although the central thrust of the present effort is the analysis of the vigilance framework as a candidate for exploring multitasking-based decrements to performance, it is worth briefly acknowledging that vigilance decrements have been sought, and found, in driving tasks. Long haul drivers experience a linear elevation of response time with increased time on task (Schmidt et al., 2007). There is evidence that such decrements, within the driving task, are minor and uncommon (Parasuraman & Nestor, 1991). This, along with the difficulty of differentiating vigilance decrement from fatigue may account for the scarcity of research directly addressing vigilance in driving tasks (Matthews & Desmond, 2002). Presently, it is the use of driving in the framework of a vigilance task which will be further explored.

In general, vigilance experiments share a clear and useful set of methodological affordances to deal with time. First, they clearly describe time relative to the presentation of various classes of stimuli. In a typical vigilance experiment a series of candidate signals, comprised of both critical signal and background (non-signal) events, are presented. Critical signals are clearly detectable to observers, but are presented at unpredictable times and may be missed during the ongoing task. The ratio of critical events to total (background + critical) events is referred to as SP and presented as a percentage (i.e. 1% would constitute a single signal and 99 background events) which can therefore be calculated as the quotient of number of critical signals over total signals presented. Events are most frequently presented at a consistent speed, which is referred to as the ER and is generally expressed in terms of events per minute (ex. 6 events/minute). The interval length (ex. 10 seconds) is sometimes also produced, and in this form event presentation speed is generally referred to as the inter-stimuli interval (ISI). A second advantage offered by the vigilance experiment methodology is an extensive set of well understood principles and basic procedure that allow for the control of uncertainty. For example, within the ISI of each event one or more stimuli, each a candidate signal, can be displayed. This approach allows a tight control of the complexity of the environment. In arrays of multiple candidate signals, the regular temporal pattern of the ER can lead to strategies (ex. on each event scan left to right) which facilitate detection (as in Warm & Jerison, 1984). Likewise, regular spatial patterns facilitate strategic perception (Adams & Boulter, 1964; Helton, Weil, Middlemiss, & Sawers, 2010). Temporal uncertainty can be increased by varying the ISI. For example, providing multiple temporal initiation points for signals within the window of each event while holding presentation time constant provides additional entropy to the observer while preserving the temporal structure and ER of the experiment (Adams & Boulter, 1964; Moore &

Gross, 1973). Spatial uncertainty can be increased by simply moving stimuli around the overall display environment within each event (Helton et al., 2010). Many more methodological permutations exist for vigilance. The breadth of the current body of work is such that, with knowledge of a particular need, it is quite probable a similar question has already been addressed.

One great challenge to using the vigilance experimental methodology in a dynamic performance situation such as driving distraction is the requirement to preserve the free will of the operator. ER, in vigilance experiments, violates this requirement by holding constant the rate of exploration of the environment. Operators move through their environment at a pace of their own choosing unless a system or other constraint prevents them. Naturalistically, such impedance can occur on the individual level, as with the decisions and judgements of the driver. Personal convictions, phobias and habits are examples of such individual impediments to exploration. It may also occur on the social level, as with rules or conventions. Law and religious prohibition are an example of such social impediments to exploration. Finally, the force of interference can be natural, driven by the causal and entropic developments in the environment surrounding the operator. The incursion of foreign bodies into a task space, in addition to weather and other so-called “acts of god” fall into this category (Hancock, 2013). Non-naturalistic impediments, for example those arising from the experiment itself, call the goal of naturalistic, environmentally valid experimentation into question. The control and rigor of the rigid ER framework on which vigilance experiments are built can stand in opposition to these goals.

How can flexible but rigorous management be realized? In driving distraction work, the driver may have a goal speed, but will likely not adhere to it. The chosen speed of movement through the environment may in fact be information bearing. Adherence to speed, as a dependent variable, can signify workload (Törnros & Bolling, 2006; Sawyer & Hancock, 2013; Sawyer et al., 2014a) just as adherence to lane center does. Placing lateral or longitudinal control on a 'rail', so to speak, that is to allow speed to be set by the experimenter in a kind of mandatory cruise control, works contrary to a naturalistic approach to driving. A very similar issue has been encountered in work seeking to generalize vigilance methodology to complex video game environments (Szalma, Schmidt, Teo, & Hancock, 2014). Despite the fact that video games generally afford the player control of exploration of the environment, in this experiment speed and direction were controlled by the experimenters. This methodological choice was in line with traditional vigilance work, and enhanced experimental control. It also placed the participant in the less naturalistic role of passive observer. In driving distraction work, where the present trend is toward naturalistic observation, such artificiality would not be tolerated by the community. It seems likely that the idea of directly adapting vigilance experimental methodology to address driving distraction is not possible. As this work seeks to achieve this aim, the question becomes where a more flexible approach might be found. In doing so, can the many advantages of the vigilance experimental approach be made portable?

Self-paced Search

An additional candidate framework may be found in a task used by Wolfe, Horowitz, and Kenner (2005) to explore the effects of SP on visual search detection rates. In what is referred to in this work as a self-paced search task (SPS task), 1) participants engage in a series of trials containing multiple critical and background stimuli. Crucially, unlike a vigilance task, 2) the rate of exposure to trials is left up to the observer. The task environment can be observed at the participant’s leisure before providing a response. The absence or presence of a critical signal is reported; 3) response is binary. 4) Exercising either option advances the observer to the next ‘event’, with its associated stimuli. As in vigilance tasks, spatial uncertainty can be used in SPS tasks to increase difficulty (Adams & Boulter, 1964). This method has similarities to other forced-choice approaches used in visual search (see Eriksen & Eriksen, 1974; Eriksen, 1995). When performed with a manipulation of SP, the SPS task produced has produced in accord with previous vigilance work; lower SP levels led to reduced performance (Warm & Jerison, 1984).

The ER in an SPS task deserves a moment of special consideration: as participants respond their response times are also their speed of passage through the task, and the environment formed by its stimuli. ER, therefore, can be calculated as the sum of all reaction times within a period of interest over the number of events within that period of interest. Even though ER is here relegated to the role of dependent variable, as in vigilance faster ER is associated with lower accuracy (Fleck & Mitroff, 2007). It is also logically associated with faster response times, overall. Parenthetically, no significant effect of time of task, the central feature of vigilance experiments, is generally found in these SPS studies (Wolfe et al., 2005; Fleck & Mitroff, 2007). Thousands of stimuli are presented over long periods, and the principal difference methodologically is the change from ER as an independent variable to a dependent variable. It seems likely the “passive observer” status of those in vigilance tasks may therefore hold some responsibility for eliciting the vigilance decrement (and see Hancock, 2013). The change of ER from independent variable to dependent variable in SPS tasks leads to another interesting possibility: enforcing a minimum or maximum ER. The former is a strategy that has been used to determine if the loss of response in low SP conditions was merely a symptom of moving too quickly through the stimuli (Wolfe et al., 2007). The task used involved playing a series of beeps when the participant moved above the ‘speed limit’. This had the effect of slowing response times, but not of eliminating the prevalence effect, leading authors to conclude that no mere speed-accuracy shift was taking place. The other option offered, enforcing a maximum ER, is one that is more interesting in light of driving distraction research. In such an experiment the participant would move at their own pace up to the maximum ER for the presented stimuli, at which point the event would be terminated, any critical signals coded as a miss, and the participant moved forward to the next event. In other words, a passive observer would experience some or all events within an SPS experiment much as a classical vigilance task, but could in always choose to take a more active role and move through the stimuli at a faster pace. Such a hybrid SPS/vigilance task approach to ER has practical parity in the task of driving, where a driver moves forward at the speed of their choosing, but obstacles (critical signals) impose an imperative to respond before the ‘end’ of the event. This hybrid is, however, speculative.

The SPS task, as described above, is a framework potentially capable of being used in a driving, and therefore other complex contexts. This endorsement must come with following caveats. First, in some applied tasks movement through the environment is at a pace set by the system, by availability of information, or some other external force. In these cases, use of an SPS task would itself be artificial. Therefore, it should be noted here that SPS tasks are likely better candidates for evaluation of multitask detriment in contexts in which the observer is an active, aggressive seeker of information (as in Sawyer, 1985). In passive contexts, the traditional vigilance framework offers more benefits. In indeterminate situations, the untested hybrid SPS/vigilance framework above presents a possible solution. The second concern relates to false alarm rates, which in cases of multitasking are concealed by the binary response of SPS. When three varieties of critical signal are being searched for simultaneously, but only the presence or absence of such are reported, how may a researcher compute an accurate false alarm rate by critical signal type? At issue is the question of whether differences between experimental conditions is are a result of a criterion shift (for example, an inflation of β), or a sensitivity shift (for example, a decrement to d’). Such inflation and decrements can have counteractive effects, each canceling the other’s influence. Without accurate false alarm rates, analysis techniques that might illuminate these differences, such as signal detection analysis, calculations of predictive power, and construction of ROC curves, are not an analysis option. The loss of these tools is not ideal, but even without them the SPS tasks provide a superior toolset than is available in conventional driving distraction inquiry.

SPS tasks satisfy the requirements outlined above. The SPS framework can accurately specify space/time and the relative concentrations experimental variables within each. It also supports the free will of the operator to explore the environment at a pace of their own choosing. SPS tasks show potentials for utility in tasks both basic and applied. In visual search experiments where stimuli is organized in static slides on a computer screen, movement through the space of the experiment, the events, can be said to occur at the speed of participant response, or event rate as a dependent variable. On a roadway, or in a driving simulation, movement through the space of a roadway environment unfolds more quickly the faster the participant moves through it. Therefore, consider the division of the driving environment into sections of equal distances, invisible to the participant but known to the researcher, each considered one event. These ‘event-sections’ could be driven by the participant at a speed (event rate) of their own choosing. Longitudinal movement, in such a task, connotes movement through events. Note that this arrangement allows for many of the methodological strategies used in both vigilance and SPS tasks. Presentation of stimuli can be assigned a SP in terms of how many event-sections it appears in, relative to the total number of event-sections. Spatial uncertainty can be built within the environment, while temporal uncertainty can be built by changing longitudinal location within an event-section. The ‘speed limit’ described above (from Wolfe et al., 2007) would become quite literal. Such tasks could overlay traditional driving distraction investigations without changing their underlying methodology or mechanics. Consider a ‘temporal overlay’, where in a manner similar to the use of epochs (Sawyer et al., 2014a) or windowing (Morgan & Hancock, 2011), a researcher would map interactions within a framework of event sections, revealing the concentrations of independent variables in terms of SP. Notably, data recorded in simulation or in GPS outfitted vehicles might be retroactively analyzed in this fashion, allowing for reinterpretation and extension of existing driving datasets. Therefore, SPS tasks satisfy the final requirement of backwards compatibility to previous methodologies. In the present work, having identified a framework for use, we will now look its application in driving distraction inquiry and beyond.

Projected Results and Theoretical Implications

Having settled upon SPS tasks as a framework for distraction inquiry, it is instructive to look to the body of work already conducted using these techniques, which suggest a significant additional utility. Wolfe and associates aptly titled “Rare items are often missed in visual searches” (2005), chronicles their findings regarding a manipulation of SP in visual search, and relates the pattern to applied situations including baggage screening and radiological screening. As in vigilance work, low SP critical signals are responded to less often (see Warm & Jerison, 1984). The pattern is not linear; in searches comprising millions of trials (Mitroff & Biggs, 2014), probability of a given critical signal being presented (SP) plotted against probability of it being reporting forms a logarithmic relationship (see Figure 6).

Mitroff and Biggs coined the name ‘ultra-rare-item effect’ to describe this logarithmic prevalence effect in accuracy data, such that accuracy rates remain stable in gradual decay before divergence toward ever steeper decrement. These signals, it should be noted, may not actually be perceptually missed. Given the opportunity to correct their searches (Fleck & Mitroff, 2007), observers reconsider and rectify the majority of the missed rare critical signals, and so it may be more accurate to say rare critical signals are rarely responded to (Hancock, 2014).

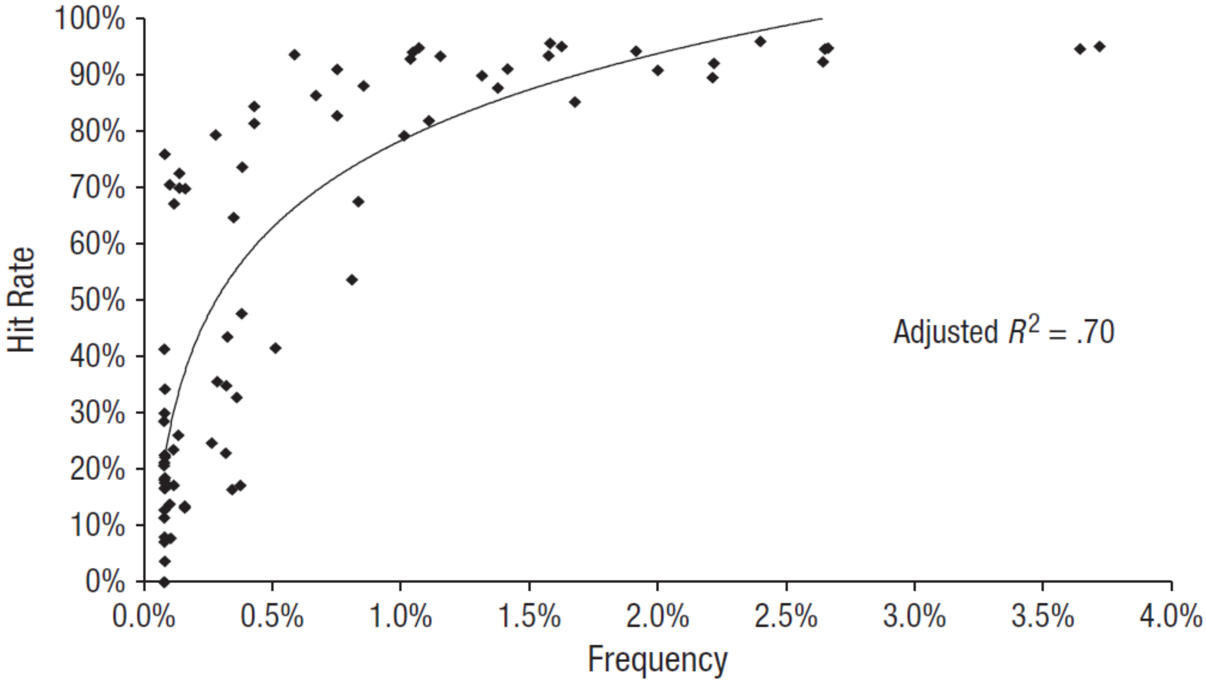


Figure 6: Log pattern of accuracy as a function of signal probability (SP) in Airport Scanner Detection accuracy as a function of SP for 78 classes of critical signals (diamonds) in the smartphone game “Airport Scanner”, in which players identify dangerous objects in simulated x-ray bag footage. Natural logarithm fit of this data is made clear by the trend line. Detection rates remain fairly high and steady at higher SP, but decay quickly at lower signal rates. The above is from Mitroff and Biggs, 2014, although a similar pattern exists in the data from Wolfe et al., 2005. Note that, mathematically speaking, this is a ‘logarithmic growth’ pattern, although the focus of the above and present work will be on the pattern’s divergence toward lower accuracy.

Work highlighting the correctable nature of these failures to respond should be viewed in the context of applied settings. While radiologists might have the time to check their suppositions again, in all too many contexts failure to respond is tragically uncorrectable. The ultra-rare-item effect therefore has significant applied relevance, as real-world critical signals of grave import are often extremely rare (for example see Szalma et al., 2014). Hundreds of hours of uneventful driving pass before a sudden evasive maneuver is necessary. Hours of email work pass before a decision about a malicious file is required. Hours, days, months or even years of uneventful watch in enemy territory pass before a firefight. The same effect can be considered in relative terms. Per brake action, stop-and-go traffic carries less risk of a rear-end accident than unexpected braking from a forward vehicle, but both risks increase should the driver engage in messaging. Likewise, the grim reality of a continuing clash of armies does not carry the same terror and excitement of a sudden encounter, or the even more disastrous ambush, in which one side is taken while their attention is elsewhere. Checking the contents of a spam folder carries less risk than encountering a malicious email in the inbox, even though the concentration of malicious email within is likely higher. In each of these applied contexts, a lesser danger of high probability in the environment is contrasted with a more dangerous one of low probability.

Here, then, is a good juncture to discuss why the logarithmic nature of a pattern is a useful piece of information. The default statistical assumption is that relationships are linear (as expressed in the general linear model), and as such it is easy to make linearity an assumption when conceptualizing a given system. When a linear model is fit to a system better expressed (and better fitting, in terms of R^2) as a logarithmic function, the result is an assumption of a homoscedastic fit of the data, when in fact a systematic form of heteroscedasticity prevails (see Figure 7). In practical terms, two collateral errors are made in interpretation: overestimation and underutilization. The impact of these errors can be significant. For example, the log pattern created by binary choices in search, effectively eliminating half of the problem in each decision, reaches solutions much more quickly than a linear approach (see Wolfe, 2012). Such logarithmic patterns are common in natural systems which attempt to produce optimal result within a set of constraints. This behavior can be seen graphically: in Figure 7 the bottom of the log divergence toward minimum performance may come close to the maximum value for the factor of influence, while the broad, slowly declining top of the curve hugs maximum performance. Such is the case with the visual search pattern shown in Figure 8A. A linear pattern fit here would represent less overall success, while wasting area under the line by extending into the illogical areas such as over 100% hit rate. The benefit of understanding these patterns as logarithmic, therefore, is precisely superior prediction: by fitting a better model to the data, erroneous heteroscedastic interpretations are avoided.

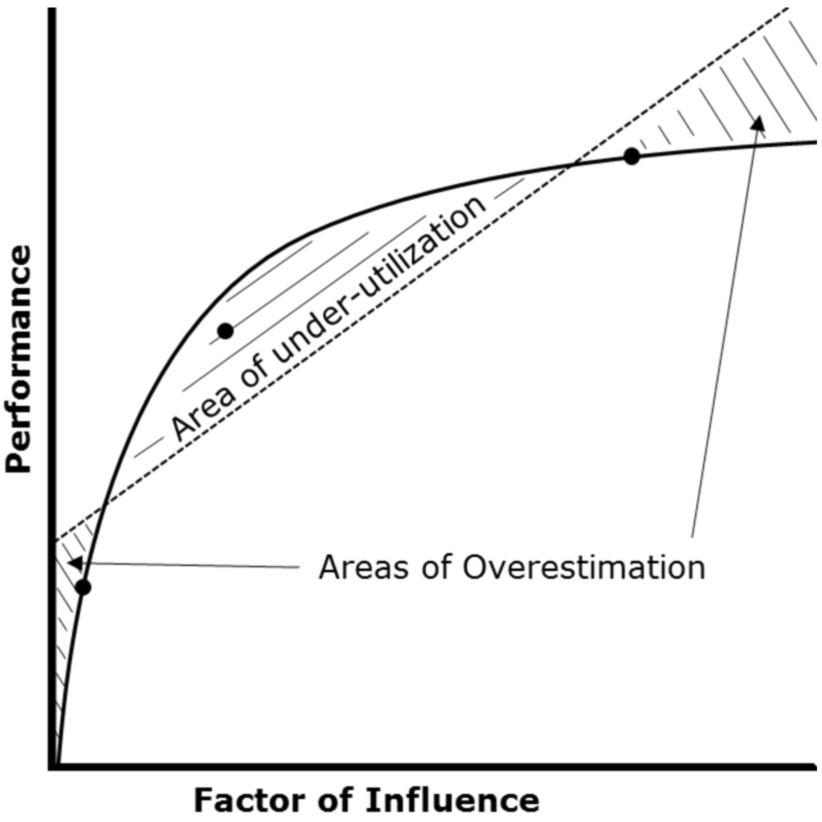


Figure 7: Under-utilization and overestimation in log systems misidentified as linear
When erroneously fitting a linear model to points better modeled by a log function, the heteroscedastic fit produces two collateral errors result which can skew interpretation. First, areas of overestimation, in which the linear model overstates performance, and second, an area of under-utilization, in which the linear model understates performance.

Consider examples where logarithmic interpretation of the data underlying SP-based signal detection decrements (Figure 8) would provide better proscriptive guidance for design of equipment or tasks. For example, take the design of algorithmic email filters which remove malicious emails. Assuming that detection of malicious emails by the operator follows a pattern such as that shown in Figure 8A, the logarithmic model suggests that there may in fact be a point at which removing more signals (i.e. removing more malicious emails) actually leads to a greater probability of the operator failing to detect and so engaging with these cyber-attacks. It could therefore be desirable to allow 2% or so of such emails through (although presumably defanged) so as to allow the human enough ‘signal’ to have a reasonable chance to detect. The linear interpretation of the data, in this case, overestimates operator performance at low SP and erroneously suggests that greater removal of signal must always be a net positive. In another example, assuming that reaction to roadway events by a driver follows a pattern such as that shown in Figure 8, the logarithmic model sheds light on the previously mentioned real-life example of stop and go traffic vs the proverbial child running into the road. It becomes clear why the latter, very rare signal is so much more of a threat. Such understanding is useful for accurate causal induction, and the resultant deduction of appropriate design.

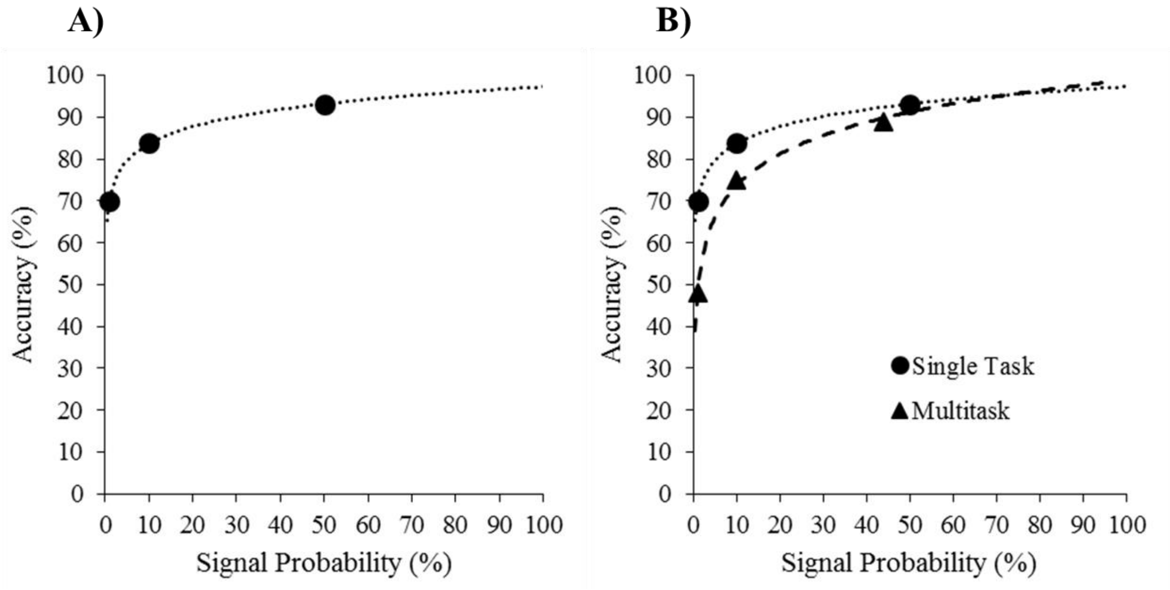


Figure 8: Log patterns found in single and multitask data from Wolfe and colleagues, 2005
Plotting of single-task data from Wolfe and colleagues, 2005, representing aggregate detection data of 24 participants completing over 2000 trials. Log functions can be seen for A) single-task (R^2 = .96) and B) multitask (R^2 = .96) visual search. In single-tasking the rare critical signals are least frequently responded to. In multitasking it is the rare critical signals that experience the greatest decrement relative to their singe-tasking counterparts.

To date, no comparisons have been made between single-task and multitask search in cases of the ultra-rare-item effect. However, a brief thought experiment will reveal that such is happening on the roadways at this very moment. Further, each applied context explored in this work involves a rare target type, and involves multitasking. To have utility, the SPS tasks used experimentally must generalize to multitasking. One SPS study, at least, has asked participants to ‘multitask’, although not in pursuit of any multitasking-related research question. Wolfe and colleagues (2005) were curious if asking baggage handlers to look for common and rare signals together (i.e.: handguns, hopefully rare, and iPods, quite common) might mitigate the effects of missed rare critical signals. They therefore asked participants to search for three categories of critical signal simultaneously (see Figure 8B). They found, not to the surprise of those versed in driving distraction literature, that simultaneous, multitask search did not mitigate the rare critical signal issue, but detection rates were in fact lower for all three categories. In comparing this data (see Table 1) now, a decade later, the comparative patterns of single and multitask trials reveal further utility for evaluating multitasking based distraction, in driving and other contexts.

Table 1 : Accuracy Data Taken from Wolfe et al., 2005

	Signal Probability (%)		
	1	10	50/44
Single Task	70	84	93
Multitask	48	75	89
Decrement	-22	-9	-4

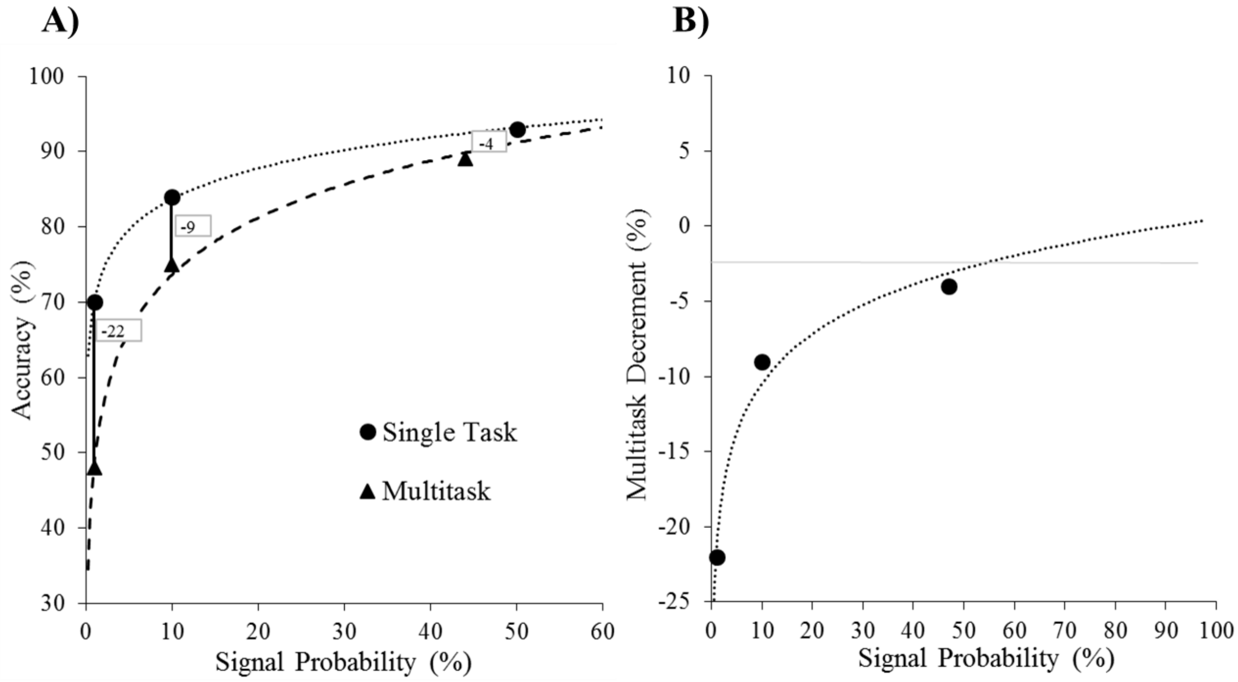


Figure 9: Decrement differences in log patterns from Wolfe and colleagues, 2005
A). Decrements between single and triple-task multitasking are plotted. Note that Wolfe et al. (2005), chose different ‘high’ signal rates for the two experiments plotted (44 vs 50), and so the rightmost decrement can be considered an approximation which likely overestimates the actual decrement. B) The decrements from each level are plotted and fitted to a log function (R^2 = .96). In single-tasking, rare critical signals are rarely responded to, and in multitasking it is rare critical signals which experience the greatest decrement relative to their singe-tasking counterparts. This rare-multitasking decrement inflicts egregious penalties to detection.

A decrement may be calculated in moving from single to multitasking, analogous to the type of decrement that might be calculated in a driving distraction study. In this sample, low, medium and high SP critical signals show high, medium and low decrements to performance in multitasking (see Figure 9). As such it seems likely that, in terms of detection, rare critical signals are the most impacted by multitasking. That is to say, 1) in single-tasking rare critical signals are least frequently responded to, and 2) in multitasking it is the rare critical signals that experience the greatest decrement relative to their singe-tasking counterparts. It would seem that, in multitasking, low SP tasks carry a compound risk of failure. This rare-multitasking decrement, in the present data, is egregious when considered in the context of practically any applied task.

The pattern of these rare-multitasking decrements themselves constitutes a log function (see Figure 9B), which mathematically stems from two natural logarithmic patterns offset by intercept. In practical terms this rare-multitasking effect has substantial applied import, and is not presently described elsewhere in the literature. Parenthetically, the pattern bears resemblance to information theory (Shannon, 1948), and to the Hick-Hyman Law (Hick, 1952; Hyman, 1953). It is also presently constructed from data harvested from a previously published paper (Wolfe et al., 2005). If the pattern holds in experimentation, it could be useful in predicting the broader pattern of multitasking-based accuracy decrements, such as driving distraction.

Research Questions

The previous theoretical discussion of distraction covers significant scope. Before stating formalized hypotheses, it is important to stop and summarize identified areas of exploration and concepts to be used in the experimental portion of this work. Three major suppositions can be identified. First, 1) the application of the distraction construct to contexts beyond driving can be justified and supported. Suggested strategy for making such arguments, including the distracted trolley, are outlined. Second, 2) the superior control of SP space/time afforded by SPS task methodology will allow for better understanding of relationship between number of concurrent tasks, target density, and resultant performance. This performance is best measured in terms of a) response time, b) accuracy, and c) subjective workload, as measured by the NASA TLX. Third 3), the better understanding provided by use of SPS methodology to investigate distraction in varying contexts will result in nomothetic principles which span context. Prior research into prevalence effects and vigilance has associated lower SPs with lower accuracy, slower response time (as noted by Fleck & Mitroff, 2007) and higher workload (as in Warm, Parasuraman, & Matthews, 2008). Accuracy across levels of SP, in past work, were also better fit (in terms of R^2) by a logarithmic function than a linear function (Wolfe et al., 2005 as reinterpreted above, Mitroff & Biggs, 2014). Prior research into driving distraction has likewise associated performance of concurrent tasks with lower accuracy, slower response time, and higher workload (Sawyer et al., 2014a).

To explore these research questions, a series of four experiments were conducted. First, I) a conceptual and much expanded replication of the basic psychophysical experiment performed by Wolfe et al., 2005 was used explore the relationship between number of concurrent tasks, target density, and resultant performance. The findings from this experiment were compared to results gathered in three applied contexts: II) driving distraction, III) battlefield threat detection and IV) email-based cyberattack.

Overarching Hypothesis Structure

In order to provide a common reference for all experiments and common notation for all figures, six generalized hypotheses are forwarded (and see Figure 10):

It was hypothesized that 1) higher SP conditions would lead to higher accuracy, 2) faster (lower) response times, and 3) lower subjective workload. 4) Performing more tasks concurrently was hypothesized to lead to lower accuracy, 5) slower (higher) response times, and 6) higher subjective workload.

The individual experiments may also have additional hypotheses. For example, the logarithmic nature of accuracy data at varying SP levels is tested only in experiments which have three levels of SP presented at a single level of multitasking. Also, the structure of some experiments means data is not collected to test all overarching hypotheses, and so experiment may have non-contiguous hypothesis numbers. This choice was made in order that hypotheses could be easily compared across experiments. Thus, referring to “hypothesis one” will always evoke the prediction that higher SP conditions should lead to higher accuracy.

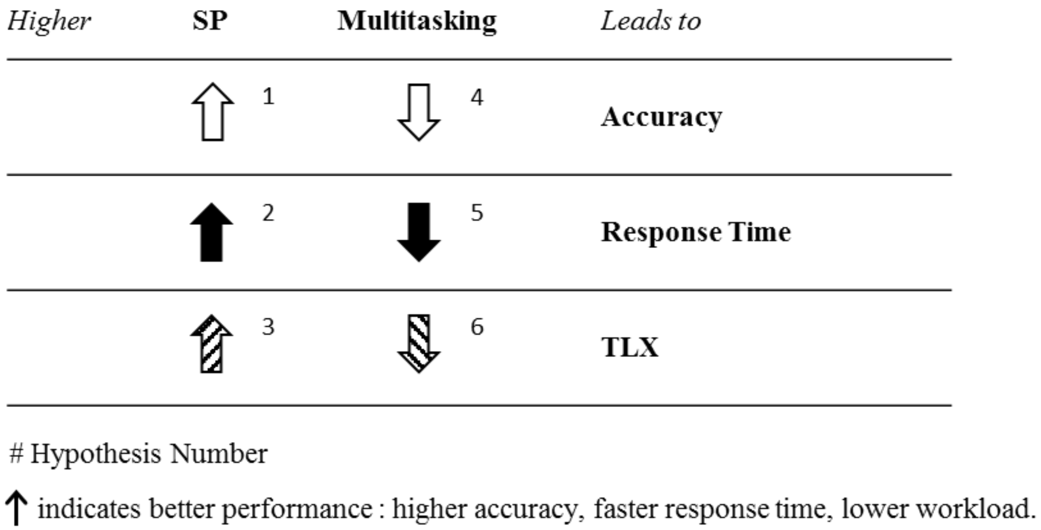


Figure 10: Overarching hypotheses for experiments I, II, III & IV
These hypotheses are common to all four experiments, and choices have been made to allow for easy comparisons. Hypothesis numbering is consistent throughout, and so experiments which do not produce data to test all hypotheses will have non-contiguous numbering. Hypothesis figures here and below will also refer to “quality of performance”, with up-arrows signifying higher accuracy, faster response time, and lower workload. The first line may be read: “In the hypothesis one higher SP was predicted to lead to better accuracy performance and in hypothesis four higher multitasking was predicted to lead to lower accuracy performance.”

CHAPTER THREE: EXPERIMENTATION & METHODS

Experiment I: Basic Psychophysical SPS Task

This conceptual replication and extension of Wolfe and colleges (2005) had the following goals: first to replicate the data seen in Table 1, and second to extend this data to include dual-tasking. As this experiment is based on a non-applied task, no argument regarding inherit ontological primacy can or will be made. Instead, an attempt was made to produce and use stimuli within the task which would relate to an applied setting: driving.

It was desirable to avoid saturating more than 50% of candidate events with critical signals (to avoid moving from a go task to a no-go task, see Helton et al., 2010), and so SP values of 1%, 10% and 35% were used, which results in a triple-task condition familywise SP of 46%. Due to the instrument's placement at the end of a trial, TLX values could only be calculated within a task. As such, TLX scores were necessarily separately analyzed by multitasking level, collapsed across SP conditions. Finally, the design used (see Figure 11) was not perfectly symmetrical with regard to the distribution of the 1% SP condition within levels of multitasking. Specifically, single-task conditions and the triple-task condition each collect data for the 1% SP only once, while in the dual-task condition it is collected twice. These levels should not differ significantly, as previous research (Wolfe et al., 2005) has suggested that concurrent performance of a high SP task with a low one apparently does not affect performance on the low SP task. As such, the plan is to collapse across the dual-task 1% categories. The result is a 3 SP (1%, 10%, 35%) x 3 of multitasking level (single, dual, triple) design, within subjects, with collection of the dependent variables of accuracy, response time and subjective workload (see Figure 11).

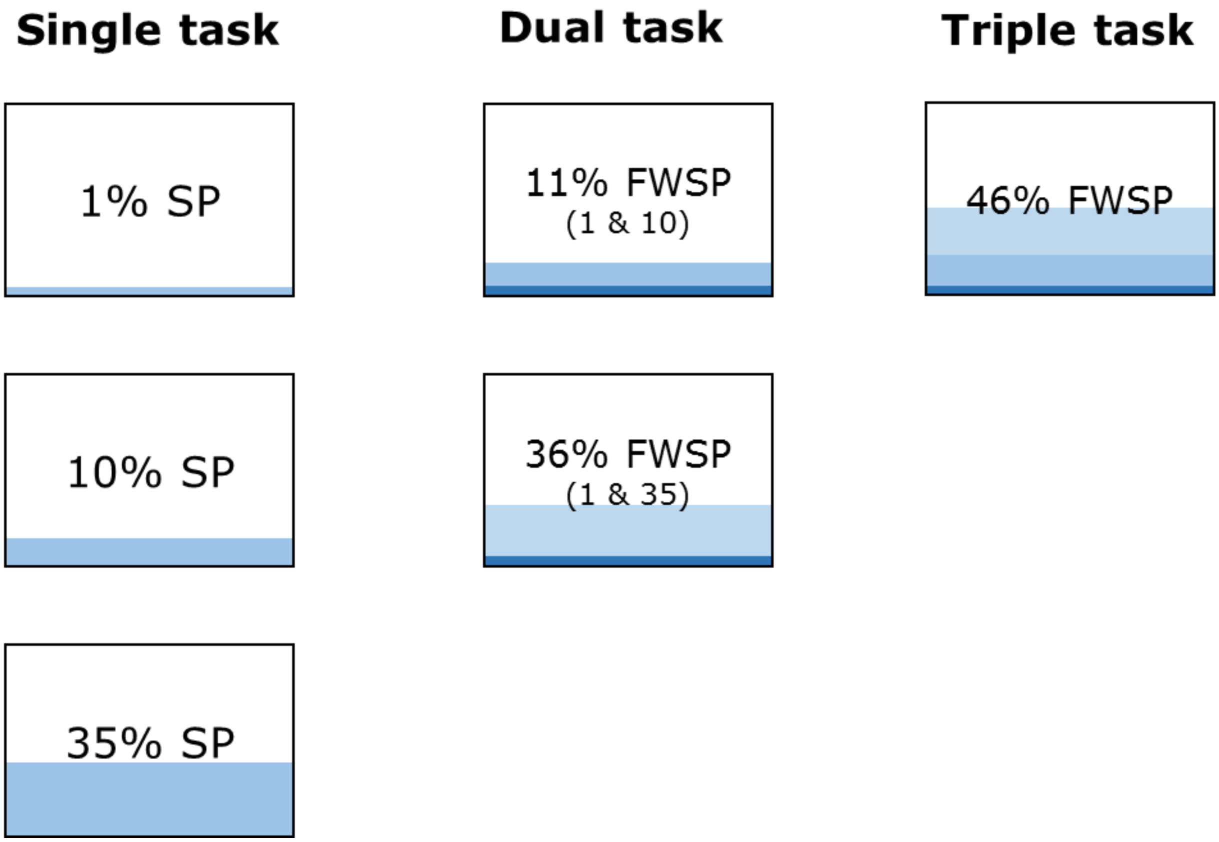


Figure 11: Blocks presented within-subjects in experiment I
The set includes each SP in single-tasking as well as all novel combinations of SP that include the 1% reference category. White represents non-signal background stimuli. In order to create a 3 SP (1%, 10%, 35%) x 3 of multitasking level (single, dual, triple) design, the two dual-task 1% conditions were intended to be averaged. Note: three multitasking levels is not a coincidental number; it is the minimum to plot a logarithm.

Participants

A sample of 30 participants was recruited from the undergraduate population of the University of Central Florida, and provided class credit in return for one hour of time. Each held driving licensure, had 20/20 or corrected vision, and self-reported having no neurological impairments. Based upon analysis of effect sizes from previous research (Wolfe et al., 2005), this sample size was more than sufficient to power the present design.

Apparatus

Informed consent and collection of demographic data was achieved with an internet-connected laptop using Qualtrics (2013). A researcher script was kept in Google Docs, allowing researchers to see which trial they were on at a glance and also to report any difficulties. Training and experimental stimuli, as well as data collection, were achieved through use of PsychoPy, an open source experiment generation suite (Peirce, 2009). The custom PsychoPy script used in the experiment was run on eight Windows 7 desktops (Figure 12), each with Core2Duo processors, 4GB of RAM, a 128GB SSD, and a 15" external Dell LCD monitor displaying the desktop environment at 1024x768 resolution. Participants responded via standard Dell ANSI QWERTY keyboards, accepting input only from the up and down arrow keys; key presses outside of these keys were ignored. Sound was presented with stereo, over-the-ear closed-back sound isolating earphones.

Figure 12: Terminals used to run experiment I
The bottom, horizontal monitor was used in experiment III, and in this experiment was powered off. Al monitors were of similar make, model, and displayed the same resolution.

Stimuli & Task

Stimuli were taken from the context of driving distraction studies (such as Sawyer et al., 2014a), and consisted of 40 greyscale images from one of 10 categories: navigation, street signs, text messages, billboards, passengers, pedestrians, unlit warning lights, cars, lit warning lights, and braking cars (see Figure 13B). Any category could serve as critical signal or non-signal stimuli, although in the present experiment only braking cars, text messages and pedestrians were used as critical signals. In a given block one to three categories would be set as critical signals, with each displayed at a different SP. The following signal probabilities were used: 1%, also referred to as "rare critical signals", 10%, also referred to as "uncommon critical signals", and 35%, also referred to as "common critical signals".

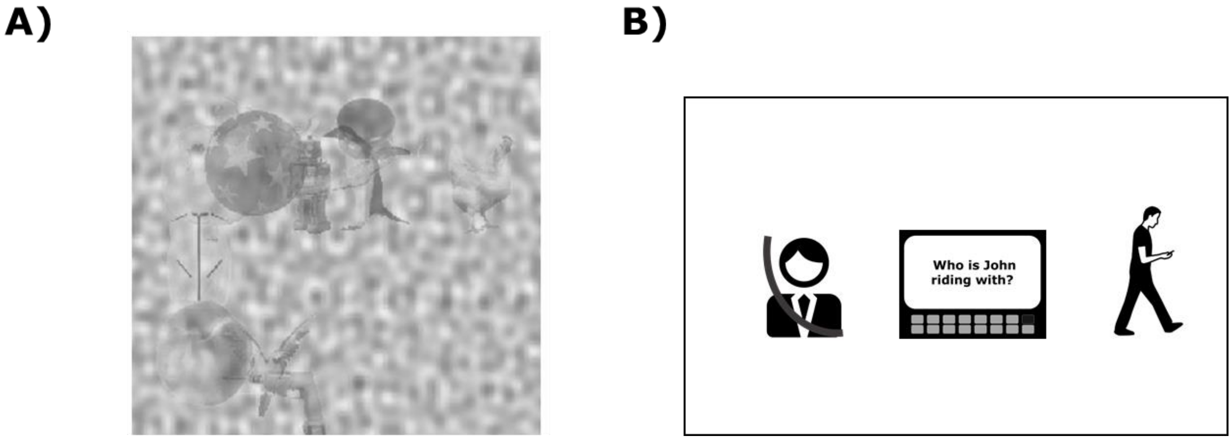


Figure 13: Comparison of past stimuli with that from experiment 1
Stimuli as presented in a single event within A) Wolfe and colleagues (2005) study and B) the present study. In the former A) the drill just left of bottom-center was the critical signal. In the latter, B) the pedestrian was the critical signal. The mask applied to A was forgone in the present experiment.

Before each block participants viewed a screen informing them which categories would be critical signals, and how frequently those categories would appear. Candidate stimuli were displayed on the screen side-by-side in groups of three (Figure 13B). Participants were instructed to respond using the down and up arrow keys: down to ‘brake’ for critical signals, up to move forward. Pressing either key advanced the screen to a new event. Auditory feedback was provided. The sound of wind passing was played for correct key presses, including correct detections and correct rejections. A buzzer was played for incorrect key presses, including misses and false alarms. There was no overlap of critical signals; a maximum of one critical signal was available in each event.

Training occurred at the beginning of the experiment, and consisted of slides bearing instruction on response buttons, feedback, and how to treat critical signals and background stimuli. In this training, the experiment was framed as a ‘game’ with ‘levels’. Short training blocks were conducted, with the following critical signal combinations at an aggregate 50% probability: braking cars alone, text messages alone, and pedestrians alone, then braking cars and text messages were presented together, and finally, braking cars, text messages and pedestrians all presented at the same time. A criterion of 80% accuracy on each of the three single-task portions of the training block was used to determine whether the participant would be retained to complete the full study.

Experimental blocks presented are shown in Figure 11. In blocks containing a 1% SP critical signal, 500 events were presented. In all other blocks 200 events were presented. In the single-task conditions, the following blocks were presented: braking cars at 1%, 10% and 35%, text messages at 10%, and pedestrians at 35%. In all multi-task blocks 1% critical signals were braking cars, 10% critical signals were text messages, and 35% critical signals were pedestrians.

At the conclusion of each block, a NASA TLX asking about "your experience in the last level" was completed by the participant. Timestamps for all key-presses were recorded, allowing for the coding of accuracy and response times.

Procedure

Participants were run in groups of up to 8. After group informed-consent, participants were provided time to ask questions. Each was asked to power off and surrender all electronics capable of producing any alert, including watches, which were held until the conclusion of the experiment. Each participant sat at a separate computer station. Training was followed by experimentation, as previously described. At the end of experimentation, participants were debriefed, and released.

Experiment-Specific Hypotheses

The following specific hypotheses were forwarded for experiment I (see Figure 14):

1) Higher SP conditions would lead to higher accuracy, 2) faster (lower) response times. 4) Performing more tasks concurrently would lead to lower accuracy, 5) slower (higher) response times, and 6) higher subjective workload. 7) Within a level of multitasking, the pattern of SP related accuracy changes would be better fit (in terms of R^2) by a logarithmic function than a linear function.

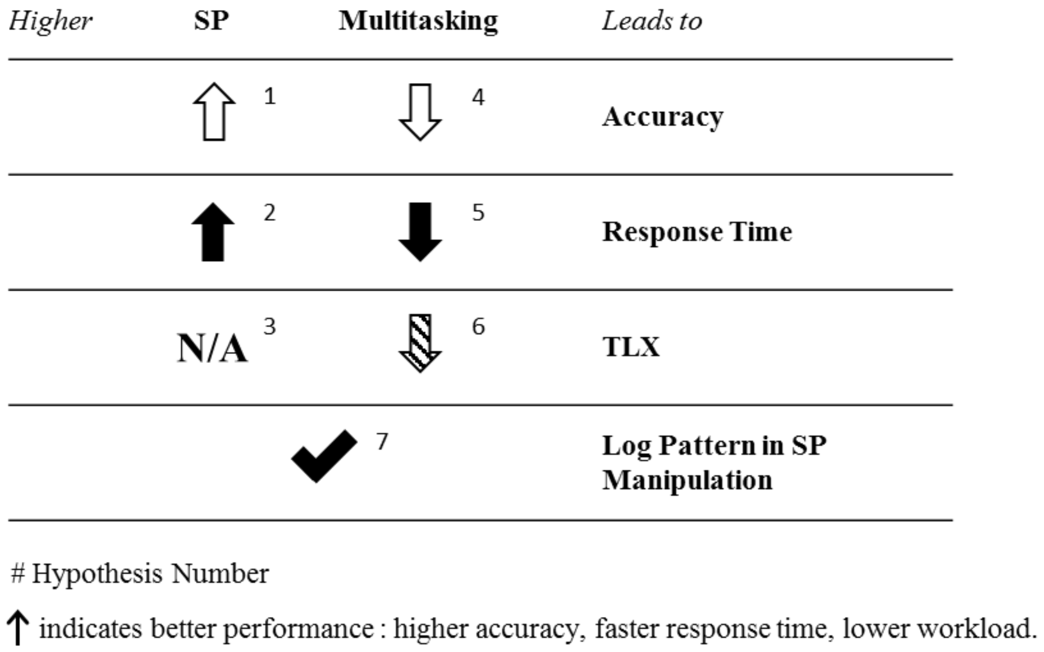


Figure 14: A visual representation of the hypotheses for experiment I
The first line may be read “In terms of accuracy, “Higher signal probability (SP) conditions were hypothesized to lead to better performance, higher multitasking was hypothesized to lead to poorer performance.” Please note, as no SP independent TLX data was expected, no SP specific TLX hypothesis (3) was forwarded.

Experiment II: SPS Task - Multitasking in Driving Simulation

The second experiment was conducted in the context of driving, a context named for the well accepted ontological primacy of the driving task. Methodologically, it was modeled on driving simulation pace-car experiments (esp. Sawyer et al., 2014a) and adapted to fit an SPS framework. Participants drove and received text messages, and were tested in reaction time by a braking lead vehicle. A ‘temporal overlay’ of event sections allowed the control and measurement of SP for both of these events. As in experiment I, the dependent variables of interest were accuracy (in terms of collision avoidance) and response time of the 1% signal, a stopping car. Two levels of dual-task performance were tested, one pairing 1% brake and 1%

text SP, one pairing1% brake and 10% text SP. This led to three conditions to be evaluated between subjects (see Figure 15). False alarm rate was not calculated, as use of the brake defensibly occurs in driving speed maintenance. TLX scores were recorded following each participant's drive. Speed over time was likewise collected; in the present experiment the sum of response times was not equivalent to ER, although they would covary to some extent.

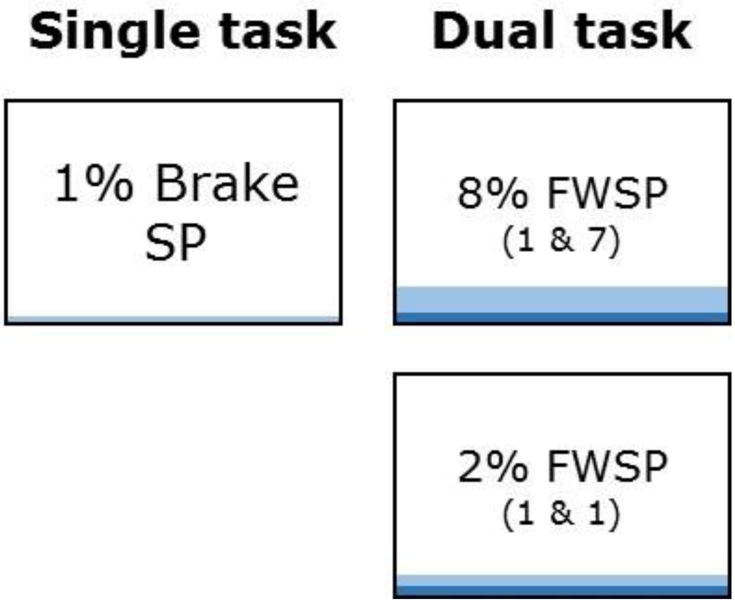


Figure 15: Task combinations in experiment II
All conditions are to be evaluated between-subjects. The set includes single-task, baseline 'brake only' condition and two levels of dual-task performance, one pairing 1% brake and 1% text SP, one pairing1% brake and 7% text SP. White represents background (non-critical) event-sections.

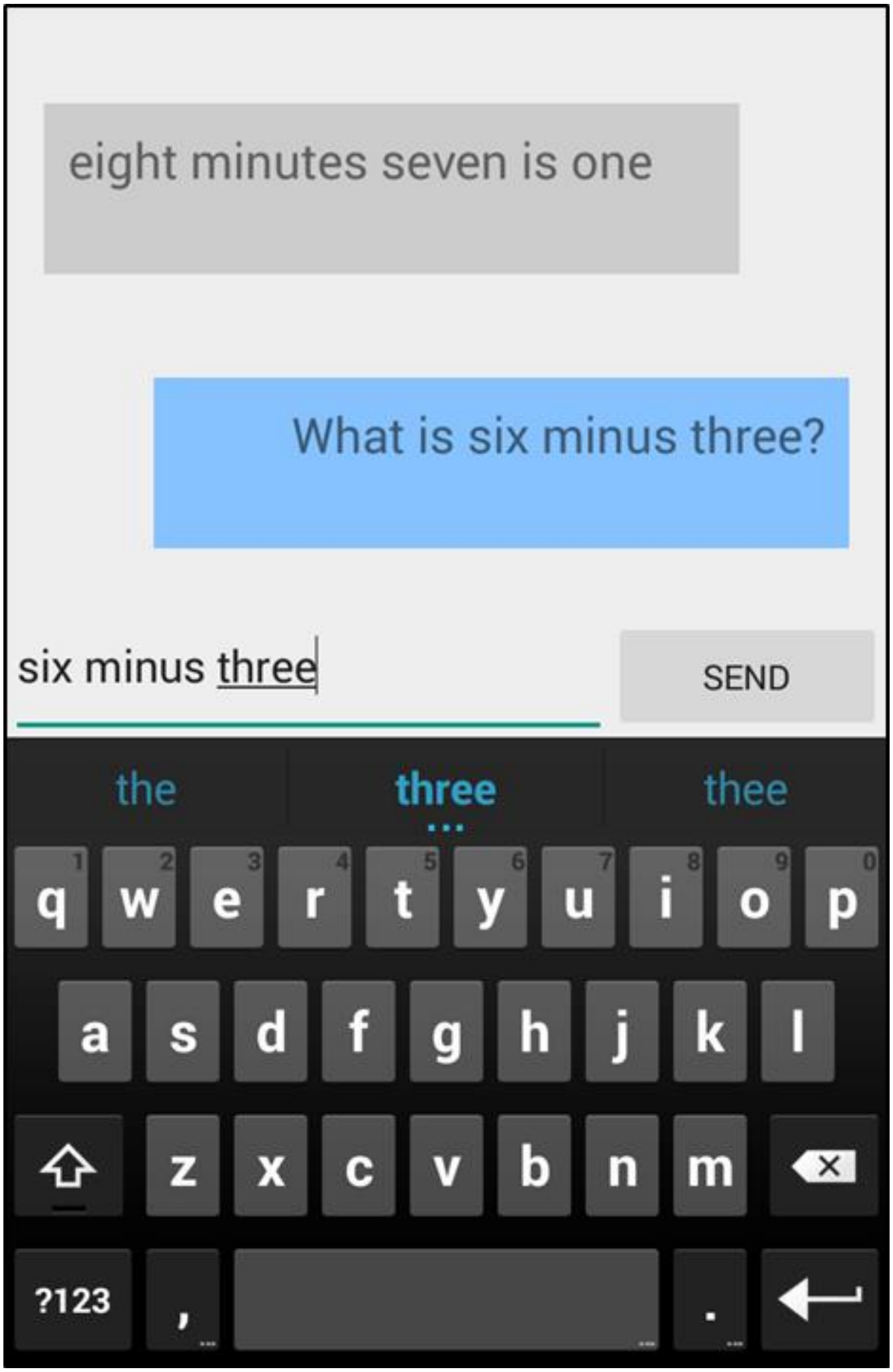
Participants

A sample of 36 participants was recruited from the undergraduate population of the University of Central Florida, and provided payment of 10 dollars in return for thirty minutes of time. All were required to hold driving licensure, to have 20/20 or corrected vision, and to self-report having no neurological impairments. This sample size was based upon analysis of effect sizes from previous research (Sawyer & Hancock, 2013).

Apparatus

Informed consent and collection of demographic data was achieved with an internet-connected laptop using Qualtrics (2013). A researcher script and schedules coordinating researchers was kept in Google Docs, allowing researchers to see which trial they were on at a glance, and also to report any difficulties.

Stimuli, in the form of a virtual driving environment, was generated by a fixed platform PatrolSim driving simulator with a 270 degree rear-projected field of view. The stock PatrolSim software was modified to allow delivery of text messages according to the x, y position of the participant's vehicle within the simulated environment (see Sawyer et al., 2014a; Sawyer et al., 2015a). Such messages could be delivered to and received on a smartphone running Android 4.4 and a specially designed app, visually similar to the stock android messaging app (Figure 16). All participant interaction with the smartphone and simulator was recorded by a purpose-built application at a rate of 60 Hz (for more information see Sawyer & Hancock, 2012; Sawyer et al., 2015a). The messaging app was modeled after the stock SMS application included with Android 4.4, which itself has similar appearance and functionality to the messaging application included with Apple iOS. Message delivery to the smartphone could be triggered by the participant's physical location in the simulation.



eight minutes seven is one

What is six minus three?

six minus three

Figure 16: Smartphone app used in the driving study
A screenshot of the native Android app displaying the text messaging task from the present study. This tool allowed not only for the delivery of text messages, but for millisecond precision records of all user interactions (as in Sawyer et al., 2015a).

Stimuli & Task

The virtual environment used for this experiment consisted of a 5.4 mile stretch of three-lane highway posted at 65 MPH, at which speed it could be driven in 5 minutes. The environment was sub-divided into 100 event-sections (see Figure 17), each comprising 3 seconds of driving at 65 MPH or 285.5 feet. As such, a desired SP could be assigned to a task based upon the number of event-sections in which it appeared. Critical signals and background stimuli, both in the simulation environment and presented through the phone, were delivered within event-sections with some temporal uncertainty: they could be presented at ‘0’ feet, or 100 feet, or 200 feet. This uncertainty allowed the presentation of critical signals and background stimuli within event- sections to be less predictable to the participant. There was no overlap of critical signals, either brake or messaging; a maximum of one was available in each event-section.

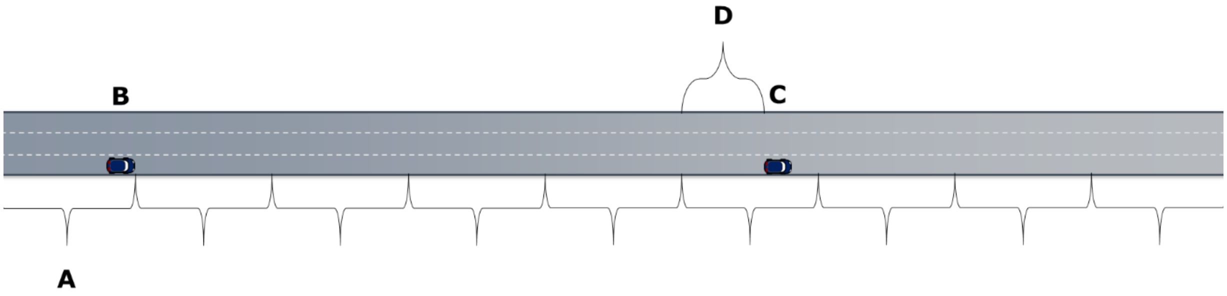


Figure 17: Use of event-sections to control signal probability (SP) in a driving study
Shown here are nine of a sequence of 100 event-sections (A), each comprising 3 seconds of driving at 65 MPH (285.5 feet). The participants, although trained to drive the prominently displayed speed limit of 65 MPH, could drive their vehicle (B) through the environment at the speed of their choosing. As such, when at the beginning of critical event-section C a lead vehicle was presented stopped in the road 2 seconds of driving distant at 65 MPH (190.36 feet) (D) the following was true: 1) each participant had the same amount of space, but dependent of speed of travel, different amounts of time to react, 2) perception, action, and a result of either hitting or missing the car was carried out within the single event section, and 3) the SP (SP) of the interaction could be quantified at 1% of the 100 available sections in which the stopping car stimuli might have been presented.

The task was modeled after a driving distraction pace-car study. Background stimuli were only presented in the simulation- all text messages were signals. Such background stimuli were presented in 19% of event-sections, in the form of billboards, pedestrians, and signage. A single critical event-section (SP 1%) triggered the appearance of a stopped car, brake lights lit, in the lane at a distance of 190.36 feet, or 2 seconds at 65 MPH. To minimize the non-environmentally correct and potentially saliency enhancing experience of having a vehicle appear in front of the driver, night-time conditions including rain and dense fog were included in the simulation, such that forward visibility was limited to around 200 feet. Response times and accuracy (in the form of collision avoidance) were recorded. A crash sound was played when participants failed to avoid collision, as they collided with the pace car. Textual critical signals, “messages”, were in the form of simple mathematical questions (i.e. “What is seven minus four?”). Participants drove under conditions of receiving no signal text messages, rare text messages (SP 1%, a single event-section), or uncommon text messages (SP 7%, 7 event-sections).

Procedure

Participants were run individually. After informed-consent, including time to ask questions, each was asked to power off and surrender all electronics capable of producing any alert, including watches, which were held until the conclusion of the experiment. Each participant then engaged in 1) messaging training, 2) driving training, 3) a single experimental drive, and 4) exit demographics. In text message training a) the participant practiced using the phone to receive text messages and send replies. Ten training math problems to be responded to were sent. In driving training b) the participant was instructed in how to start and shift the simulator to drive, then drove for five minutes through an environment that included the necessity to steer, brake, and maintain a speed of 65 MPH. Next, c) the participant engaged in an experimental drive under one of the three conditions above (Figure 15). Finally, the participant completed d) a NASA TLX and demographic questionnaire, which included debriefing.

Experiment Specific Hypotheses

The following specific hypotheses were forwarded for experiment II (see Figure 18):

First it was hypothesized that 1) higher SP would lead to higher accuracy, 2) faster (lower) response times, and 3) lower workload. Dual-tasking would lead to 4) lower accuracy, 5) slower (higher) response times, and 6) higher workload. It was also hypothesized that 7) these data would produce patterns consistent with the basic pattern seen in experiment I.

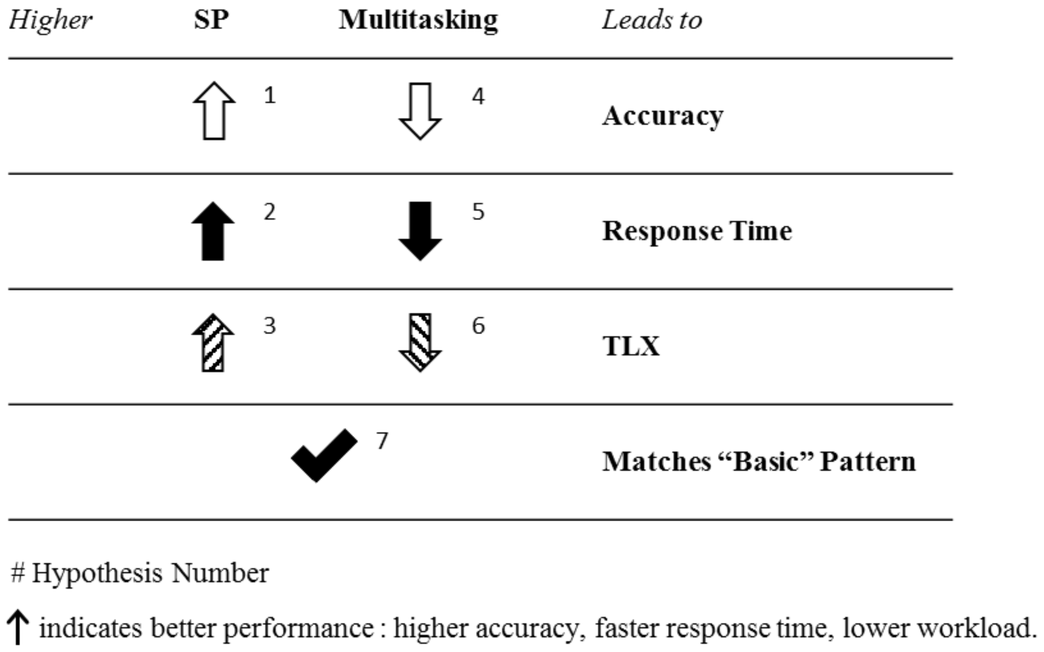


Figure 18: Hypotheses for experiment II
The first line may be read "Higher SP conditions were hypothesized to lead to better performance in accuracy, higher multitasking was hypothesized to lead to poorer performance in accuracy."

Experiment III: SPS Task - Battlefield Threat Detection

In dismounted military contexts, mission success may rely on timely action not only in the immediate environment, but in monitoring remote data not in semantic context to the operator's present location. Such multiple-display multitasking degrades task performance in other domains, (Fischer & Haines, 1980; Wickens & Long, 1994, Sawyer et al., 2014a), elevating workload, and reducing the stable load level that can be sustained over time. While consideration has been given to what sizes of display best facilitate performance of cognitive tasks (Hancock, Sawyer, & Stafford, 2015), very little research exists evaluating such battlefield displays in terms of distraction potential (Yeh, Wickens & Seagull, 1999). Unintended, iatrogenic consequences (as in Hancock, 2013) to behavior, task performance, and collateral risk are likely effects of such displays. Soldiers with chest-mounted smartphones or helmet mounted displays arguably shoulder the risk of distraction in pursuit of their duties.

The present experiment was conducted within a dismounted battlefield context, and involved detecting and reporting threats, both in an immediate world environment and in remote data delivered to a digital display. Designers of wearable information-delivery systems often think of visual interface elements in terms of layers (as in Roberts et al., 2012; see Figure 19), and in such a paradigm, the real world can be considered the final layer. It should also be considered the most important, as attention to critical signals in the real world battlefield environment protects against a variety of lethal environmental hazards, including improvised explosive devices (IEDs) and enemy combatants. In the present experiment, available signals emanated from either a 'world layer', meant to represent the participant's immediate environment, or the 'digital display layer', meant to represent a chest-mounted display. The digital display layer was divided into two streams of remote data: the first was a video downlink (VDL) displaying images, the second was a text message feed delivering enemy movement information. Participants were instructed to treat detection of critical signals from the real world layer as the primary task, and accuracy, response time in this layer were the dependent variables of interest, for performance here was representative of human performance in classifying dangerous threats.

Figure 19: Digital information sources overlay the real world
A graphical interface for a digital display in the near domain in attentional competition with a real world scene in the far domain. The dangers present in the real world far domain, including enemy combatants, must be weighed against the value of information in the near domain.

Three sources of critical signals were available; 1) the primary real-world layer, and within the digital layer both 2) a video downlink (VDL), and 3) a text message feed. Signal probabilities were set as close as possible to levels identified by subject matter experts as applicable to real dismounted battlefield situations, although it should be noted that even a 1% SP was deemed high compared to hostile environments where targets may appear with days or weeks between.

Table 2 : Experiment III Signal Probabilities by Condition and Source

Source

	Real	*Digital*		
	World	*VDL*	*Text*	***FWSP***
World Baseline **(1T01%)**	1%	0%	0%	**1%**
Low Digital All **(3T03%)**	1%	1%	1%	**3%**
High Digital All **(3T21%)**	1%	10%	10%	**21%**

Measures of subjective workload were collected following each trial. Accuracy and response times for the world layer, therefore, were analyzed as a 3 SP load (world baseline, low digital all, and high digital all) within-subjects repeated measures design. In this design (see Table 2) differences between the world baseline and low digital all conditions were intended to be indicative of ramping concurrent tasks from one to three, while changes between the low and high digital all conditions were intended to be indicative of ramping SP from 3% to 21%.

Participants

Twenty participants were recruited from the undergraduate population of the University of Central Florida, and provided class credit in return for one hour of time. Each held driving licensure, had 20/20 or corrected vision, and self-reported having no neurological impairments. This sample size was based upon analysis of effect sizes from previous research (Wolfe et al., 2005).

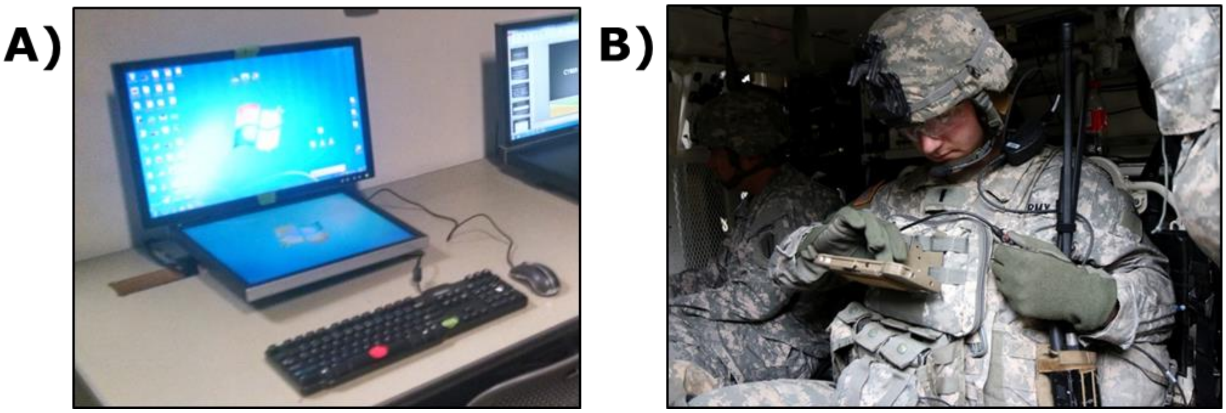


Figure 20: Terminals used to run experiment III compared with actual battlefield interfaces A) In actual experimentation the lower screen was pulled to the edge of the workstation, the keyboard moved to the side, and a gamepad held in the hands (and see Figure 21B). B) The configuration was designed to approximate the visual separation of chest-mounted displays presently used in dismounted combat situations.

Apparatus

Informed consent and collection of demographic data was achieved with an internet-connected laptop using Qualtrics (2013). A researcher script was kept in Google Docs, allowing researchers to see which trial they were on at a glance and also to report any difficulties. Training and experimental stimuli, as well as data collection, were achieved through use of PsychoPy, an open source experiment generation suite (Peirce, 2009). The custom PsychoPy script used in the experiment was run on eight Windows 7 desktops (Figure 20), each with Core2Duo processors, 4GB of RAM, a 128GB SSD, and two 15” external Dell LCD monitors displaying the desktop environment at 1024x768 resolution.

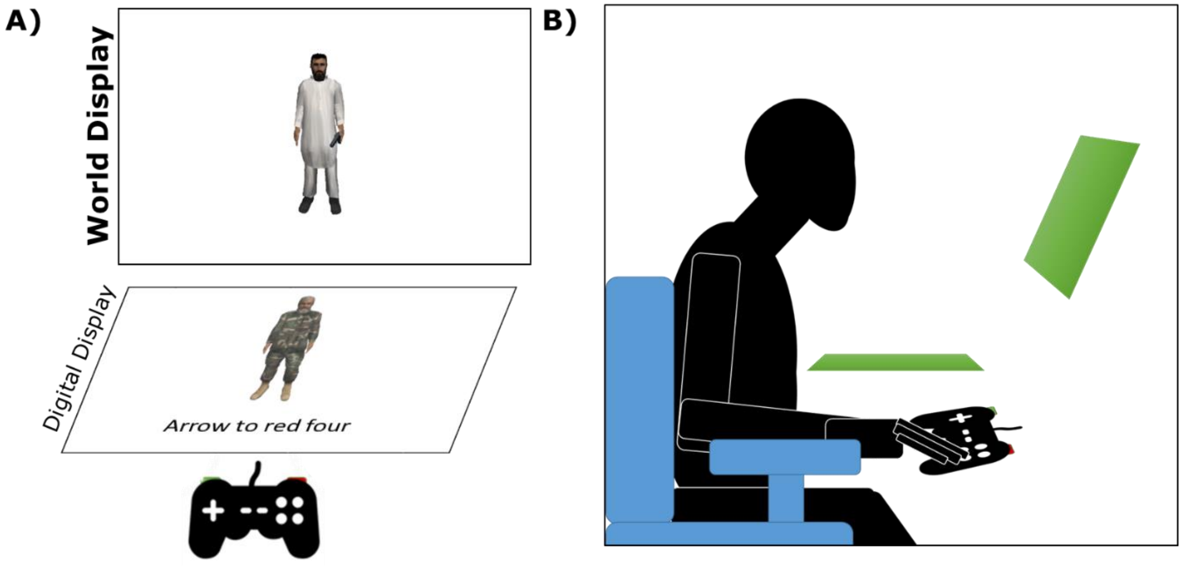


Figure 21: Detailed terminal configuration for experiment III
A) The top display held the world layer and the primary task: identifying and reporting individuals holding guns. Below, the data layer held a video downlink (VDL) in which the task was the same at the primary task, as well as a text message feed in which the task was to identify the code-word “Baron”. There was no overlap of signals. In this example, the world layer holds a critical signal in the form of an individual holding a gun. B) Participants sat with the ‘world’ layer positioned ahead, and the data layer pulled to their chest, while a gamepad was held on the lap. Participants responded via the ‘triggers’, buttons on the upper corners of the gamepad. Key presses outside of the triggers were ignored. Sound was presented with stereo, over-the-ear closed-back sound isolating earphones.

Stimuli & Task

Graphical and textual stimuli were used, the former obtained from an Air Force virtual training environment simulating Kandahar, Afghanistan. A set of fifty static targets, ten holding guns, was created by taking stills of 3D models (see Figure 21A). Critical signals in graphical stimuli were defined as any image in which the individual held a gun. Textual stimuli consisted of fifty messages using call signs to report enemy movement, of which ten used the code-word

‘Baron’, the critical signal. Graphical stimuli were presented on the forward world layer display, as well as on the lower horizontal ‘remote data’ display which contained both VDL and textual message stimuli.

Training occurred at the beginning of the experiment, and consisted of slides bearing instruction on response buttons, feedback, and when to report critical events. Training blocks consisted of 50 trials, with all critical signals (world layer gunmen, VDL gunmen, textual messages) displayed, each at a signal probability of 10%, for a FWSP of 30%. The final training block was used as a criterion for further participation in the experiment: participants unable to achieve 80% accuracy were released.

Experimental stimuli was presented in counterbalanced blocks. Before each block participants viewed instructions about critical signals and the primacy of the world layer. Blocks consisted of five hundred events. Three stimuli, two graphical, one textual, were evaluated in each event. Participants were instructed to respond using trigger buttons on a gamepad, left to report a threat, right to move forward, and either key advanced the screen to a new event. Auditory feedback was provided: a chime was played for correct key presses, including correct detections and correct rejections, and a buzzer for incorrect key presses, including misses and false alarms. There was no overlap of critical signals; a maximum of one critical signal was available in each event. At the conclusion of each block, a NASA TLX asking about “your experience in the last level” was completed by the participant. Between blocks participants were encouraged to take breaks, stretch, and move around. Timestamps for all key-presses were recorded, allowing for the coding of accuracy and response times. While global false alarms could be coded, there was no way to know which critical signal they might be associated with.

Procedure

Participants were run in groups of up to four. Each was asked to power off and surrender all electronics capable of producing any alert, including watches, which were held until the conclusion of the experiment. After group informed-consent, participants were provided time to ask questions. Each was assigned to a separate computer station. Training and experimentation, as previously described, was followed by debriefing, return of electronics, and release.

Experiment Specific Hypotheses

The following specific hypotheses were forwarded for experiment III (see Figure 22):

It was hypothesized that 1) higher SP conditions would lead to higher accuracy, 2) faster (lower) response times, and 3) lower subjective workload. 4) Performing more tasks concurrently would lead to lower accuracy, 5) slower (higher) response times, and 6) higher subjective workload. It was also hypothesized that 7) these data would produce patterns consistent with those seen in experiment I.

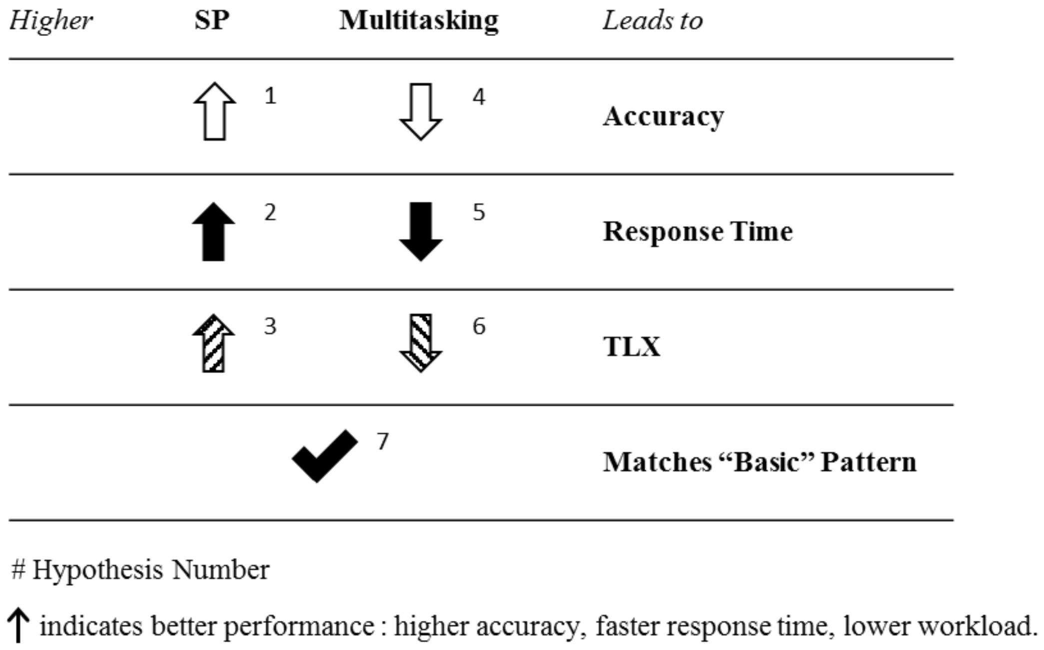


Figure 22: A visual representation of the hypotheses for experiment III
The first line may be read "In terms of accuracy, Higher SP conditions were hypothesized to lead to better performance, higher multitasking was hypothesized to lead to poorer performance."

Experiment IV: SPS Task - Signal Probability in Cyberattack

The forth experiment was conducted within the context of cyber-defense. The former Chief Scientist of the US Air Force (AF) described cyberspace as a domain through which all essential AF operations are performed (Maybury, 2012). In safeguarding this vital space, the focus of research and intervention has to date most often been on automated algorithmic systems and human teams of cyber-defenders (as in Finomore, Shaw, Warm, Matthews, & Boles, 2013; Sawyer et al., 2014b). Such human-machine teaming is responsible for reducing the number and impact of attacks and identifying those missed, but crucially can never intercept all malicious messages. Human decisions, rendered by end users, decide whether such attacks are successful. Email remains a primary form of communication for business and government, despite a growing field of competitors. As such, it represents the critical signal of choice for attackers with the intent of delivering malicious code or harvesting valuable information, respectively referred to as ‘malware’ and ‘phishing’ attacks. Indeed, it is difficult to think of a military or commercial operation which does not in some fashion rely on email. The question is, when faced with an email-delivered cyber-attack, would the operator fail to detect and engage, or ‘reject and report’ the threat?

The question of primacy is more difficult to argue here than in any other context explored in this work. Certainly, if you asked most people engaged in checking their email what they were doing, the answer would be “checking my email”. However, it is exactly this kind of response that pop-up screens used by corporate, military and government mail systems try to overcome, reminding users that they are first and foremost the guardians of cybersecurity of the organization (as in Sawyer et al, 2015). The essential goal of such messages, of online security training, and of public advertising like The Department of Homeland Security’s “Stop. Think. Connect.” campaign (Paulsen, McDuffie, Newhouse, & Toth, 2012). Still, in a “distracted trolley” thought experiment, it is particularly hard to come to the view that email is somehow a distraction from email-based cyber-attacks. This situation is further complicated by the inherently highly multi-task nature of the email task, in which evaluating the legitimacy of a communication is embedded among so many other highly variable subtasks. Indeed, it seems most arguable that email *is* the primary task, and that a subtask of that task is rejecting and reporting email-based cyber-attacks. In light of this conclusion, and of the complex multi-task nature of the task, no attempt was made to manipulate the number of concurrent tasks. Instead, a purely SP based manipulation was conceived, to better understand if the prevalence effect might be applicable even in an operationally complex and diverse task like checking and responding to email.

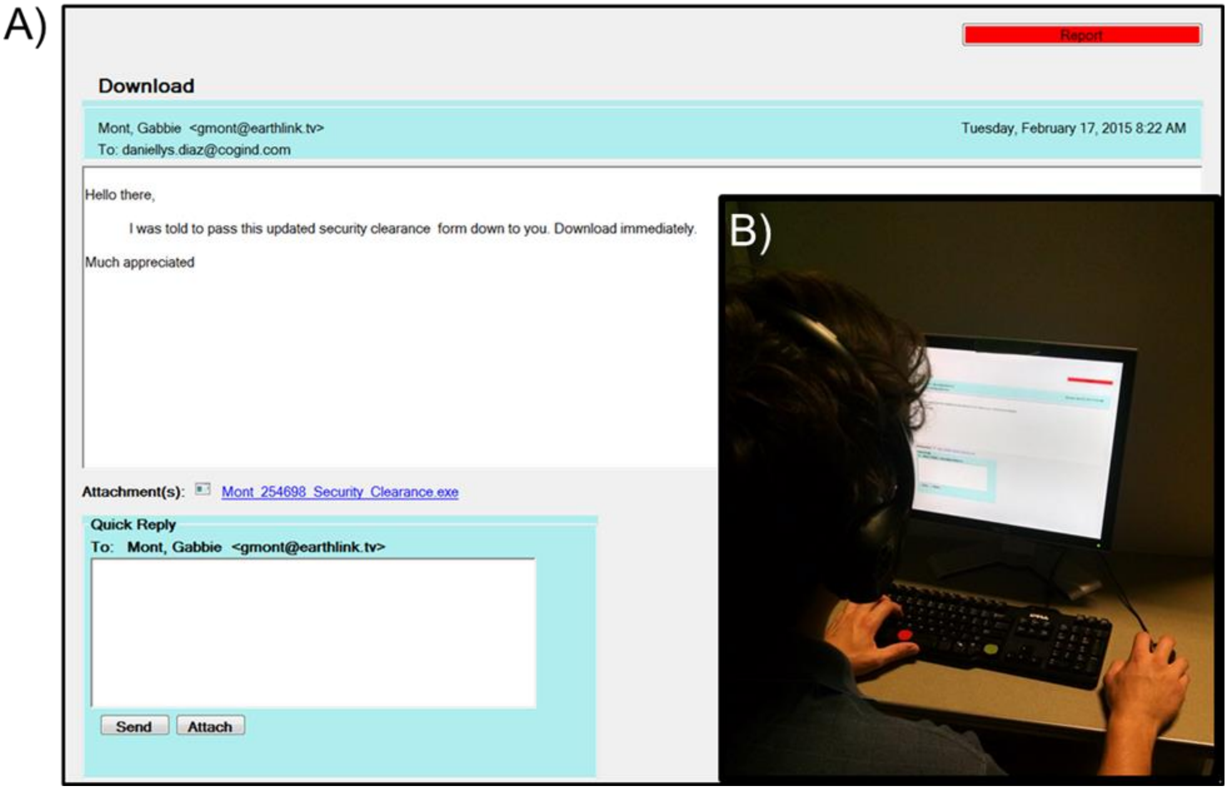
In the present experiment, participants role-playing an administrative position with the fictitious ‘Cog Industries’ received emails either containing or requesting sensitive PDF attachments. Attack emails, containing either malicious code or improper requests for information were also delivered. The question, as posed previously, was whether reducing attack email SP might lead to a counter-productive situation in which more attacks would be downloaded or uploaded as a result of low user respond-and-report rates than were prevented by removal. As a result, attack emails at a SP of 1%, 5% & 20% were included among balanced upload and download background events. This 3 SP manipulation was performed between subjects. The measure of interest in this case was accuracy, in terms of attack email detect-and-report rate and response time. This design differs from that of experiment II and III; no manipulation between single and multitasking exists, and no specific distractive influence was included. Instead, it was intended to see if, in an inherently multi-task domain, mere manipulation of SP could drive a strong difference between groups.

Participants

A sample of 33 participants was recruited from the undergraduate population of the University of Central Florida, and provided class credit in return for ninety minutes of time. All were required to have 20/20 or corrected vision, and to self-report having no neurological impairments. This sample size was based upon analysis of effect sizes from previous research (Sawyer et al., 2015b).

Apparatus

Informed consent and collection of demographic data was achieved with an internet-connected laptop using Qualtrics (2013). A researcher script and schedules coordinating researchers was kept in Google Docs, allowing researchers to see which condition they were in at a glance and also to report any difficulties.



Download

Mont, Gabbie <gmont@earthlink.tv>
To: daniellys.diaz@cogind.com

Tuesday, February 17, 2015 8:22 AM

Hello there,

I was told to pass this updated security clearance form down to you. Download immediately.

Much appreciated

Attachment(s): Mont 254698 Security Clearance.exe

Quick Reply
To: Mont, Gabbie <gmont@earthlink.tv>

Figure 23: An attack email, as presented in the email testbed
A) The interface of the email testbed (ET) was designed to mimic common online webmail interfaces, and to be minimal. In this email the inbound address, ending in ‘.tv’, the attachment, ending in .exe, and the body of the email, lacking a signature, all point to the suspicious nature of the email. B) (Inset) Participants at individual terminals interacted with the ET. Should a participant click on the executable, a miss would be recorded. Should the participant click on the red “Report” button, a hit would be recorded.

Stimulus, in the form of simulated emails, was delivered through the email testbed (ET, Figure 23), which was designed to mimic online webmail. After opening an email in the inbox, participants were able to download attachments, or reply and upload their own attachments. They could also report suspicious emails with a dedicated button. Upon taking any of these three actions, they would be returned to the inbox.

Stimuli & Task

Participants received a series of emails, one at a time. All legitimate emails came from addresses ending in ‘cogind.com’, and asked participants to download a PDF file, or upload an existing file. Attack emails came from ‘.tv’ addresses, and could be a request to download a non-pdf ‘.exe’ file, or an improper (i.e. from outside the company) request for an upload of a file.

In all, 300 emails were delivered. Depending on condition, attack emails accounted for 1%, 5% or 20% of emails (3, 15 or 60 emails, respectively).

Procedure

Participants were run individually. After group informed-consent, including time to ask questions, each was asked to power off and surrender all electronics capable of producing any alert, including watches, which were held until the conclusion of the experiment. The email testbed (ET) provided a webmail environment with which to test participants (Sawyer et al., 2015b). Each participant then engaged in 1) interface training, 2) cyber-defense training, 3) the experimental ET session, 4) finally in exit demographics. In interface training, participants viewed a PowerPoint presentation stepping them through the interface of the ET. In cyber-defense training, participants viewed slides explaining the sensitive nature of the data they would be handling, and giving simple strategies for avoiding attack emails. Finally, the participant completed a NASA TLX and demographic questionnaire, which included debriefing.

Experiment Specific Hypotheses

The following specific hypotheses were forwarded for experiment IV (see Figure 24):

1) Higher SP conditions would lead to higher accuracy, 2) faster (lower) response times, and 3) lower subjective workload. 7) The pattern of SP related accuracy changes was expected be better fit (in terms of R^2) by a logarithmic function than a linear function.

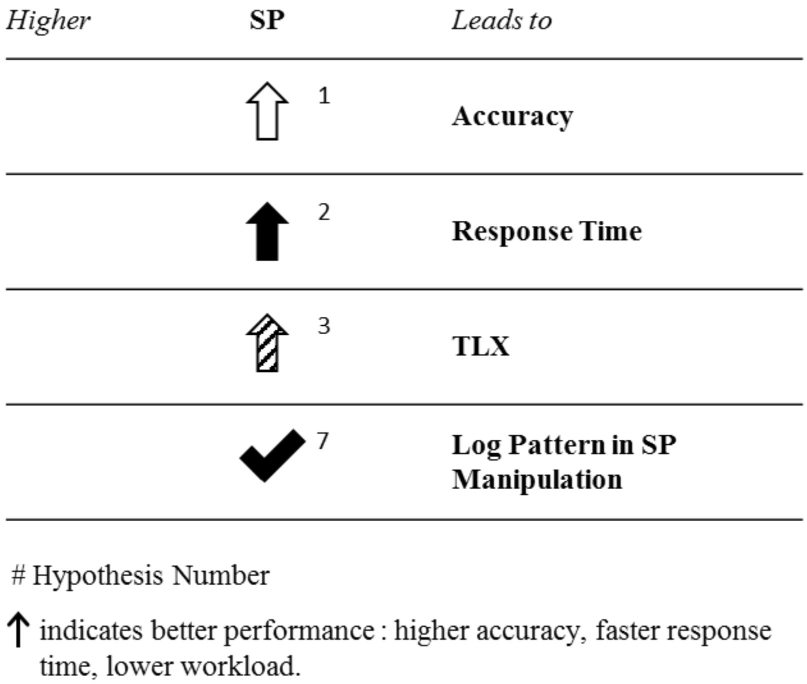
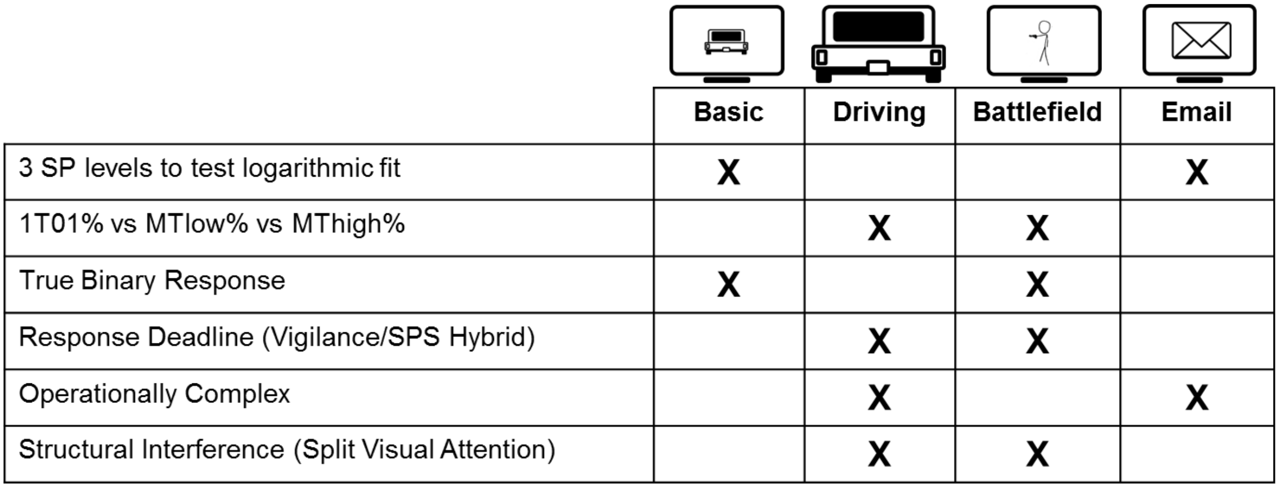


Figure 24: A visual representation of the hypotheses for experiment IV
The first line may be read “In terms of accuracy, Higher SP conditions were hypothesized to lead to better performance, higher multitasking was hypothesized to lead to poorer performance.” Please note, as there was no manipulation of multitasking level, no multitasking hypotheses are forwarded (4,5,6). Best performance was expected to be seen in the highest SP condition, with detection becoming worst in the lowest SP condition, and to be best expressed as a logarithmic function (7).

Comparing and Contrasting Experiments I-IV

The goal of varying levels of multitasking and SP in a psychophysical task and three applied tasks was to understand how concentration of targets in space and time affected accuracy and response across levels of multitasking. The applied nature of these tasks prevented perfect parity of experimental features, but an attempt, illustrated in Figure 25, was made to balance such characteristics. A list of these intentional points of parity follows. In experiments I (psychophysical) and IV (email) performance at all SP levels was analyzed, in order to better understand the effects of varying SP. In experiments II (driving), & III (battlefield) the analytical focus was upon the targets presented at 1% SP, which were theorized to show greatest volatility in the face of manipulations of SP and task number. In these two experiments, three groups tested 1) the change between single-task detection of this 1% SP group and a group with minimally inflated SP but a greater number of tasks, then 2) the change between that group and a group with greatly inflated SP. Experiment I (psychophysical) was an SPS task with true binary response, a characteristic shared with experiment III (battlefield). In contrast, both experiment II (driving) and experiment IV (email) had forms of response (braking and button clicks, respectively) which while essentially binary in nature, are somewhat more ambiguous. While neither experiment I (psychophysical) nor experiment IV (email) had deadlines to response, both experiment II (driving) and experiment III (battlefield) had such restrictions in the time available to respond, with a failure to respond being counted as a ‘miss’. Finally, experiments I (psychophysical) and IV (email) had no structural divide in visual attention, as opposed to experiments II (driving) and III (battlefield), each of which spread attention across multiple displays.



	Basic	Driving	Battlefield	Email
3 SP levels to test logarithmic fit	X			X
1T01% vs MTlow% vs MThigh%		X	X	
True Binary Response	X		X	
Response Deadline (Vigilance/SPS Hybrid)		X	X	
Operationally Complex		X		X
Structural Interference (Split Visual Attention)		X	X	

Figure 25: Comparisons between experimental features
In order to provide environmentally valid applied settings in which to test the ideas forwarded in this work, perfect parity of experiments could not be exercised. A subset of fundamental differences between experiments which were identified as potentially important is given here.

CHAPTER FOUR: RESULTS

Presented here are interactions and main effects from the four experiments: basic psychophysical (I), driving (II), battlefield (III) and cyberattack (IV). Effect sizes here are provided as η^2_p. While not specifically discussed in this section, pairwise comparisons and measures of Cohen’s *d* may be found in Appendix C. Findings for the overarching hypotheses and interpretation can be found in discussion.

Experiment I: Basic Psychophysical SPS Task

Response Dependent Variables

Data from thirty participants ($n = 30$, 18 female, 11 male, 1 non-reporting) were included in the present analysis, after six were removed for non-completion. Please note that some participants failed to qualify to finish the experiment, as described in methods. Three target types were presented at different levels of SP (braking cars at 1%, text messages at 10%, pedestrians at 35%) individually (single-tasking), in pairs (dual-tasking), or in threes (triple-tasking). Accuracy and response times from each of these categories were submitted to a within-participants 3 SP (1%, 10%, 35%) x 3 multitasking (single, dual, triple) MANOVA. These data are visually presented in Table 3.

Each visual category of target was presented associated with a single SP, leading to the risk that greater saliency in a given category of target might masquerade as the effect of the SP manipulation. As a precaution, two trials were run in which the braking car stimuli used in the reference 1% category was presented at the 10% and 35% signal probabilities. A within-subjects MANOVA was then run comparing the braking car stimuli to the two other stimuli sets (text messages, pedestrians). Univariate ANOVA results revealed a significant difference for the pedestrians set only, *F*(1, 29) = 5.535, *p* = .026, η^2_p = .160, such that at an SP of 35% performance detecting braking cars (*M* = .885, *SD* = .023) exceeded performance detecting

Table 3 : Basic Psychophysical SPS Task Dependent Variables

DV	Task	Code	# Task	FWSP (%)	*Mean*	*SE*
Accuracy (%)	Single	1T01%	1	1	54.0	4.7
		1T10%	1	10	64.2	2.3
		1T35%	1	35	83.9	1.7
	Dual	2T01%	2	1	45.3	2.7
		2T10%	2	10	61.9	2.6
		2T35%	2	35	84.2	1.6
	Triple	3T01%	3	1	68.0	4.5
		3T10%	3	10	70.6	2.1
		3T35%	3	35	83.4	2.2
Response Time (ms)	Single	1T01%	1	1	851	30
		1T10%	1	10	820	11
		1T35%	1	35	675	26
	Dual	2T01%	2	1	758	16
		2T10%	2	10	871	22
		2T35%	2	35	666	9
	Triple	3T01%	3	1	748	14
		3T10%	3	10	883	12
		3T35%	3	35	716	11

Note. Dual-task scores at 1% SP reflect averaging of 1% scores from two conditions (1% & 10%, 1% & 35%). Standard errors calculated as within-participants confidence intervals (Cousineau, 2005).

pedestrians (*M* = .839, *SD* = .031). This leads to the possibility that, in the present experiment, accuracy values for the 35% SP group may be depressed relative to the 1%SP group. This decrement is relative to what would be expected from a perfectly saliency-balanced stimuli set. However, the small effect size and four percentage mean difference reveal this confound as small, especially in light of the much larger magnitude of the pattern to be reported below. To compensate for violations of the sphericity assumption, when appropriate the Box correction (Field, 2009) has been applied to the following results. Where violations of sphericity were indicated by Mauchly's Test, degrees of freedom have been adjusted using the Greenhouse-Geisser (1959) correction.

A significant interaction of multitasking and SP was seen, Wilks' Lambda = .276, $F(8, 22) = 7.207$, $p < .001$, $\eta^2_p = .724$, and so univariate ANOVA results were interpreted. The interaction was significant for both accuracy, $F(2.570, 74.521) = 4.413$, $p = .009$, $\eta^2_p = .132$, and response, $F(2.797, 81.106) = 6.419$, $p = .001$, $\eta^2_p = .181$. These data suggests that accuracy decreases as SP decreases, and that the nature of that decrease is greater at lower signal probabilities. Dual-tasking results in lower accuracy than single-tasking at both the 10% and 1% signal probabilities, but triple-tasking outperforms both. In terms of response time, the pattern is less straightforward. At 1% SP poorest (slowest) performance is seen in single-tasking, while at the 10% and 35% levels of SP, poorest (slowest) performance is seen in triple tasking. A visual representation and expanded discussion of the pattern can be seen in Figure 26 & 27. Main effects of both multitasking and SP were significant, but neither are further interpreted here in light of the significant interaction.

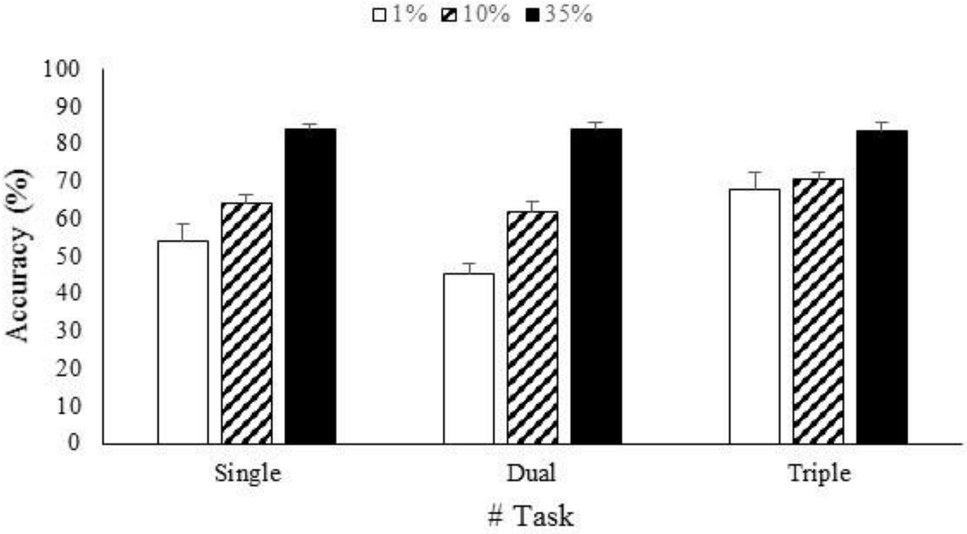


Figure 26: Accuracy results at each level of multitasking in experiment I
Patterns of accuracy at each level of multitasking for a significant multitasking by SP (SP) interaction. Note that dual-tasking exhibits poorest performance, followed by single-tasking, and finally triple-tasking. Error bars represent within-participants confidence intervals (Cousineau, 2005).

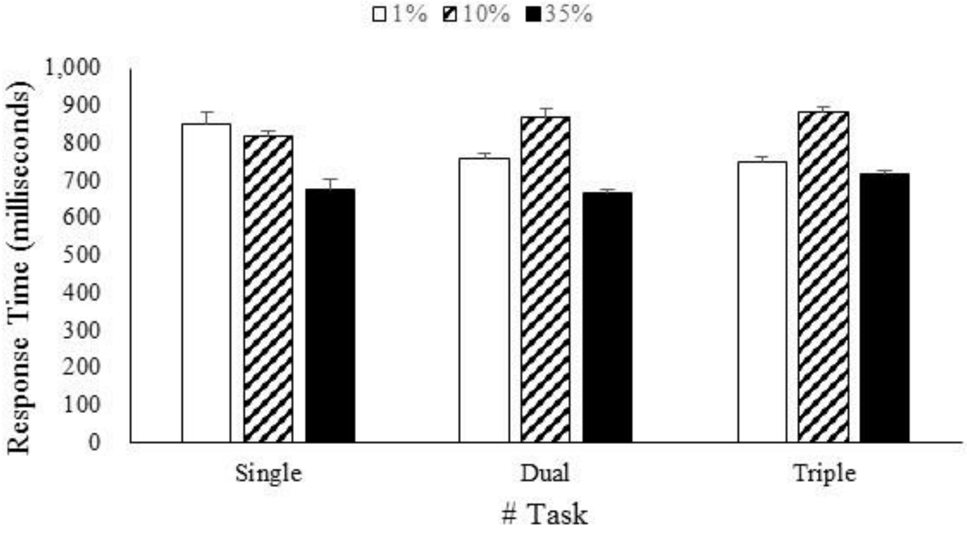


Figure 27: Response time results at each level of multitasking in experiment I
Patterns of response at each level of multitasking for a significant multitasking by signal probability (SP) interaction. It should be noted that faster times in the 1% category can likely be attributed not to better performance, but to accuracy decrement-related reduced opportunity to detect targets that might require longer search time. This phenomenon has been previously described (as in Fleck & Mitroff, 2007). Error bars represent within-participants confidence intervals (Cousineau, 2005).

A final analysis was performed to ascertain the log or linear nature of patterns of workload at each level of SP within a given level of multitasking. Untransformed and log^e transformed data were subjected to linear regression, and the R^2 of each result was calculated. These estimates of fit, presented in Figure 28, make it clear that the patterns seen in the present data are better fit when treated as linear.

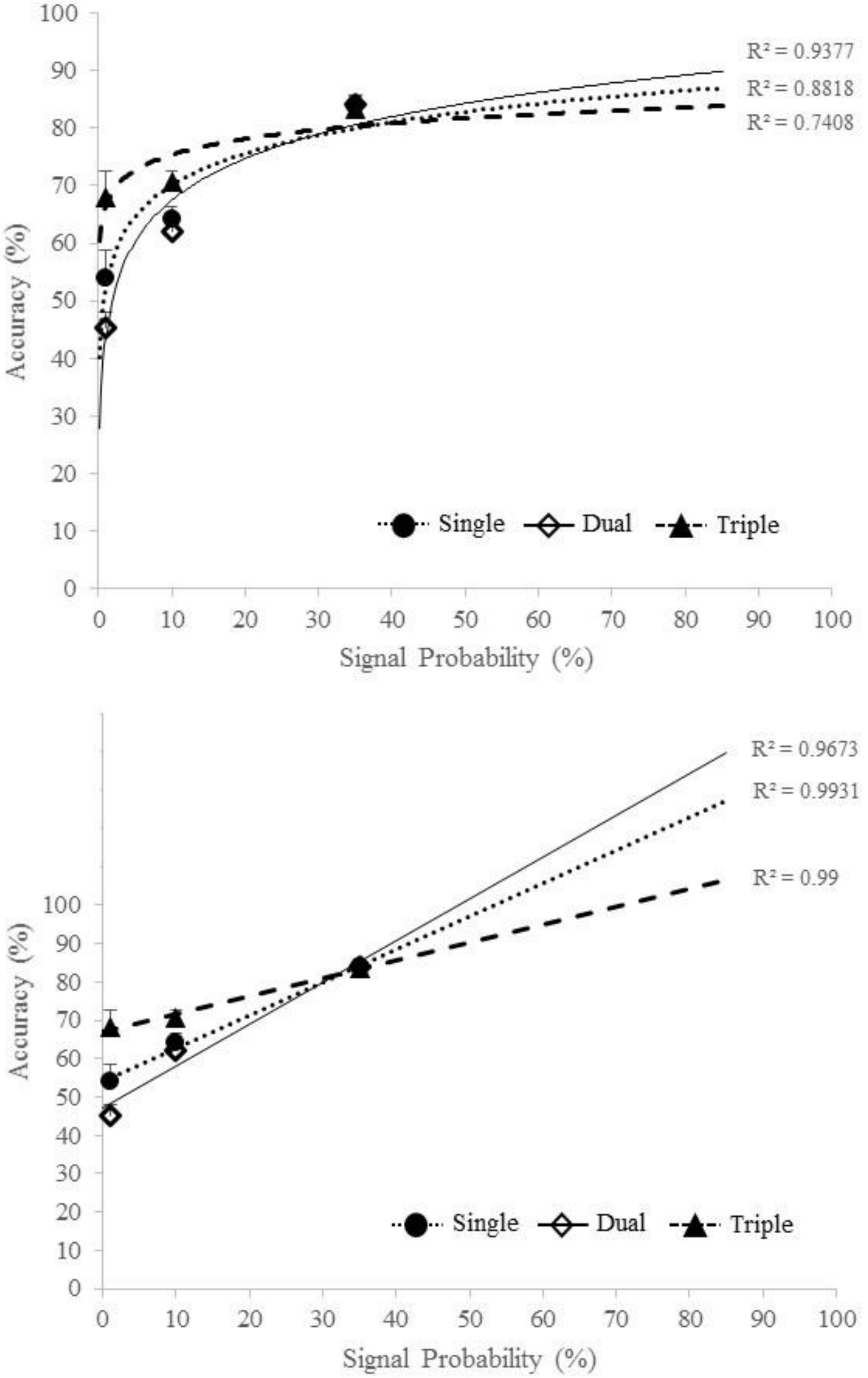


Figure 28: Logarithmic and linear fits of accuracy data in experiment I compared
The *R^2* values for linear fit with and without loge transformation are noted alongside points associated with each task level (single, dual, triple). In all three, linear fit in the non-transformed data is superior, suggesting that the pattern of accuracy decrement is linear in nature.

Subjective Workload Dependent Variables

Data from thirty participants (n = 30, 18 female, 11 male, 1 non-reporting) were included. Six dimensions of the NASA TLX (mental demand, physical demand, temporal demand, performance, effort and frustration) as well as the overall workload score were subjected to a within -participants MANOVA to assess the impact of the three levels of multitasking: single, dual and triple. To provide the data for this analysis TLX scores from all conditions were averaged by multitasking level. In cases where participants failed to answer a single subscale it was replaced with the average for the condition. Resultant descriptive statistics are reported in Table 4. Where violations of sphericity were indicated by Mauchly's Test, degrees of freedom have been adjusted using the Greenhouse-Geisser (1959) correction. No effects of gender or order were seen.

A significant main effect of multitasking was seen: Wilks' Lambda = .339, $F(12, 18)$ = 2.922, p = .020, η^2_p = .661. Univariate ANOVA results were therefore interpreted, revealing a significant effect for the composite TLX, $F(2, 58)$ = 13.677, p < .001, η^2_p = .320, and subscales including mental demand, $F(1.559, 45.205)$ = 23.537, p < .001, η^2_p = .448, physical demand, $F(1.592, 46.182)$ = 7.104, p = .004, η^2_p = .197, performance, $F(2, 58)$ = 10.757, p < .001, η^2_p = .271, effort, $F(2, 58)$ = 13.446, p < .001, η^2_p = .317, and frustration $F(2, 58)$ = 5.959, p = .004, η^2_p = .170. No significant effect was seen for temporal demand, which may reflect the self-paced nature of the task. Overall, these data support the view that subjective workload is elevated primarily by the shift from single to double-tasking, which is in line with accuracy and response results. Visual representations and further discussion can be found in Figure 29.

Table 4 : Basic Psychophysical SPS Task TLX

DV	Task	*Mean*	*SE*
TLX	Single	50.742	0.645
	Dual	58.044	1.138
	Triple	60.211	1.644
TLX Mental	Single	51.587	1.154
	Dual	68.100	2.051
	Triple	69.800	2.303
TLX Physical	Single	25.380	1.448
	Dual	38.167	2.435
	Triple	38.067	3.345
TLX Temporal	Single	54.647	0.999
	Dual	60.100	2.185
	Triple	62.900	3.750
TLX Performance	Single	64.353	0.955
	Dual	52.400	1.784
	Triple	53.867	2.336
TLX Effort	Single	47.320	1.054
	Dual	59.400	1.934
	Triple	64.533	3.010
TLX Frustration	Single	61.167	1.267
	Dual	70.100	2.338
	Triple	72.100	2.640

Note. These data are computed by averaging across all three single task conditions, both dual task conditions, and reporting the triple task condition directly. Standard errors calculated as within-participants confidence intervals (Cousineau, 2005).

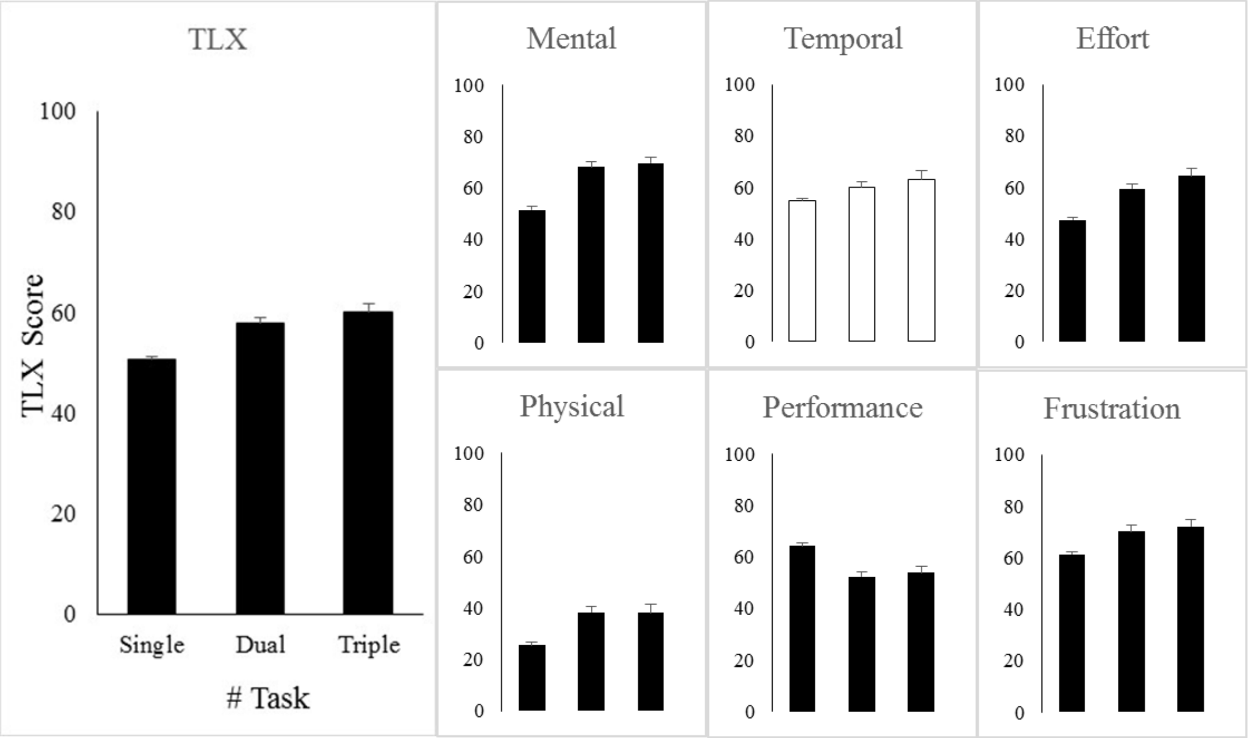


Figure 29: TLX scores at each level of task number in experiment I
Significant subscales are in black, non-significant in white. These data are computed by averaging across all three single-task conditions, both dual-task conditions, and reporting the triple-task condition directly. Note that the largest change is due to the change from single to dual-tasking, not dual to triple-tasking. Error bars represent within-participants confidence intervals (Cousineau, 2005).

Experiment II: SPS Task - Multitasking in Driving Simulation

Response Dependent Variables

Data from thirty-six participants (*n* = 36) were included in the present analysis. Descriptive statistics for all response dependent variables are reported in Table 5. Two response variables related to the driving task (collision, hybrid response time) were analyzed using a between-participants MANOVA to assess the impact of a single manipulation, the presentation of text messages in three levels of familywise SP relative to total event sections driven: driving only (0%), infrequent (1%) and frequent messages (7%). The pace car brake event was consistently presented in a single event section driven, for an SP of 1%.

Table 5 : SPS Task - Multitasking in Driving Simulation Dependent Variables

DV	Code	# Task	FWSP (%)	Result FWSP (%)	*Mean*	*SD*
Collision	1T01%	1	1	1	51.1	11.9
(%)	2T02%	2	2	11	27.4	11.6
	2T08%	2	8	61.3	4.9	12.1
HRT	1T01%	1	1	1	599	113
(ms)	2T02%	2	2	11	1007	110
	2T08%	2	8	61.3	1103	115

Note. FWSP (Signal Probability) reports the number of event sections in which any task was provided. Result FWSP reports the average number of event sections in which a participant was engaged in a task. Because the brake car task was necessarily responded to in a single event section, the entirety of this difference can be attributed to the amount of time it took participants to complete messaging tasks.

Participants were free to indulge in the text messaging task at their own pace, and as a result three concerns were investigated. First, participants might text-message slowly enough to still be engaged in the task when the pace car brake event occurred, but no such overlaps were found. Second, the higher FWSP conditions might cause participants to compensate by driving much more slowly. Mean speed in MPH at each condition was calculated: driving only *M* = 67.30, *SD* = 1.74, infrequent messages *M* = 67.02, *SD* = 4.54, and frequent messages *M* = 67.34, *SD* = 3.86. These manipulation checks led to the conclusion that participants performed in keeping with the intentions of the experiment.

No effects of gender were seen, and as each participant engaged in only one drive there were no effects of order. To compensate for violations of the sphericity assumption, when appropriate the Box correction (Field, 2009) has been applied to the following results. The operating system of each participant's own phone was introduced as a covariate to account for the familiarity that android users had with our android-based texting application. This covariate was significant, Wilks' Lambda = .712, *F*(2, 31) = 6.261, *p* = .005, η^2_p = .288.

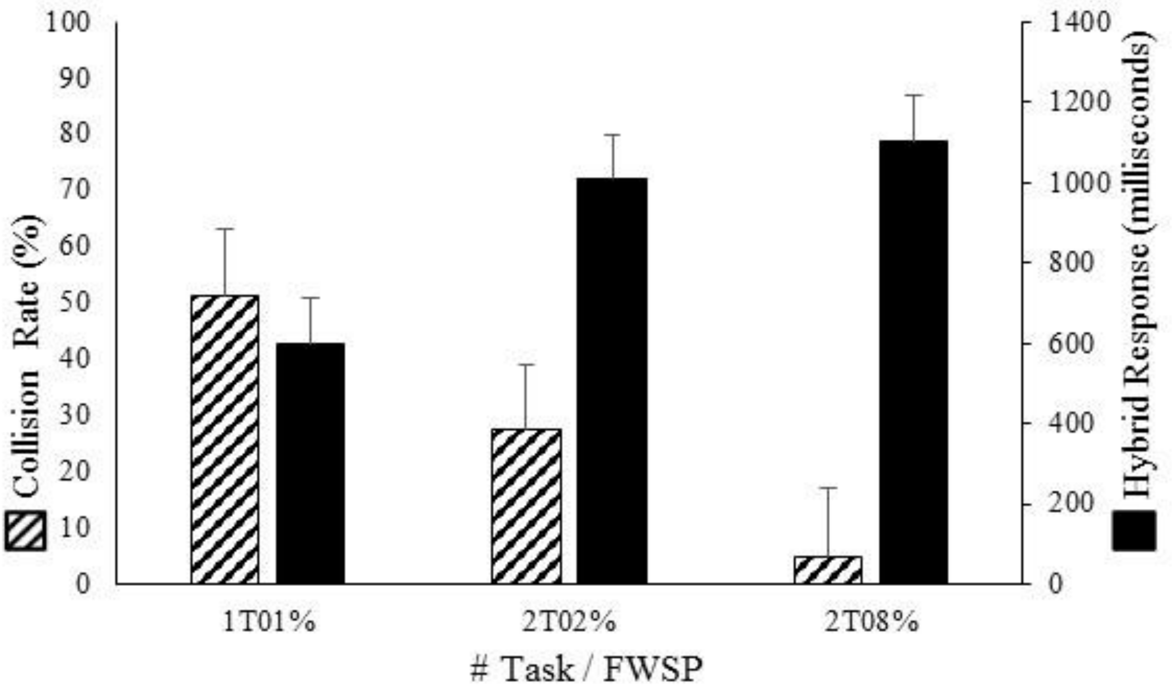


Figure 30: Collisions and response time results across conditions in experiment II
Collision rate is compared with hybrid response time across three conditions. Note that collisions, which can be considered the inverse of accuracy (see text), are reduced over the transition to dual-tasking, and again when the messaging task is expanded sevenfold. This is in spite of poorer response performance, in which the largest change is that shown from single to dual-tasking. In sum, a cost to response time is balanced against an increased chance that the participant will respond at all.

There was a significant main effect of condition, Wilks' Lambda = .618, $F(4, 62) = 4.220$, $p = .004$, $\eta^2_p = .214$. Univariate ANOVA results were therefore interpreted, revealing the effect to be significant as related to both response variables: collision $F(2, 32) = 3.455$, $p = .044$, $\eta^2_p = .178$ and hybrid response time, $F(2, 32) = 5.375$, $p = .010$, $\eta^2_p = .251$. See Figure 30 for a visual representation. These data indicate that, consistent with previous driving distraction research, response time increases with the addition of the messaging task, and when the prevalence of that task is raised response time further increases. However, collision rates concurrently decrease, a pattern consistent with vigilance findings showing that rare signals are less likely to elicit a response.

Subjective Workload Dependent Variables

Data from thirty-four participants ($n = 34$) were included, as two records were removed for excessive incomplete responses. In cases where participants failed to answer a single subscale it was replaced with the average for the condition. Resultant descriptive statistics for all subjective workload variables are reported in Table 6. Six dimensions of the NASA TLX (mental demand, physical demand, temporal demand, performance, effort and frustration) as well as the overall workload score were subjected to a between-participants MANOVA to assess the impact of the presentation of text messages in three levels of SP: driving only (0%), infrequent (1%) and frequent messages(7%). No significant effects were seen.

Table 6 : SPS Task - Multitasking in Driving Simulation TLX

DV	Code	# Task	FWSP (%)	Result FWSP (%)	*Mean*	*SD*
TLX	1T01%	1	1	1	37.858	3.651
	2T02%	2	2	11	40.642	3.572
	2T08%	2	8	61.3	48.573	3.676
TLX Mental	1T01%	1	1	1	43.670	5.004
	2T02%	2	2	11	40.549	4.895
	2T08%	2	8	61.3	63.840	5.037
TLX Physical	1T01%	1	1	1	20.157	6.246
	2T02%	2	2	11	20.176	6.109
	2T08%	2	8	61.3	29.108	6.288
TLX Temporal	1T01%	1	1	1	20.777	6.519
	2T02%	2	2	11	26.659	6.377
	2T08%	2	8	61.3	39.564	6.563
TLX Performance	1T01%	1	1	1	71.615	7.783
	2T02%	2	2	11	64.419	7.613
	2T08%	2	8	61.3	64.875	7.836
TLX Effort	1T01%	1	1	1	39.596	6.931
	2T02%	2	2	11	50.032	6.780
	2T08%	2	8	61.3	57.637	6.978
TLX Frustration	1T01%	1	1	1	31.334	9.468
	2T02%	2	2	11	42.018	9.261
	2T08%	2	8	61.3	36.413	9.531

Note. Non-significant TLX data trends toward increased demand for higher multitasking and higher signal probability.

Experiment III: SPS Task – Battlefield Threat Detection

Response Dependent Variables

Data from twenty participants (*n* = 20) were included in the present analysis. Descriptive statistics for all response dependent variables are reported in Table 7. A within-participants MANOVA assessed the impact of a single manipulation, the presentation of visual targets and text messages on a secondary digital display in three levels of SP: monitoring the world baseline (1T01%), low digital all (3T03%) and high digital all (3T21%). Response measures were captured only within the 'world' display, and included accuracy and response time. No effects of gender or order were seen.

Table 7 : SPS Task - Battlefield Threat Detection Dependent Variables

DV	Code	# Task	FWSP (%)	*Mean*	*SE*
Accuracy (%)	1T01%	1	1	74.0	5.9
	3T03%	3	3	44.0	5.7
	3T21%	3	21	56.0	5.3
Response Time (ms)	1T01%	1	1	672	21
	3T03%	3	3	768	30
	3T21%	3	21	774	41

Note. Standard errors calculated as within-participants confidence intervals (Cousineau, 2005).

There was a significant main effect of condition, Wilks' Lambda = .404, $F(4, 16) = 5.905$, $p = .004$, $\eta^2_p = .596$. Univariate ANOVA results were therefore interpreted, revealing the effect to be significant as related to both response variables: accuracy $F(2, 38) = 5.865$, $p = .006$,

η^2_p = .236 and response time, $F(2, 38) = 3.670$, $p = .035$, η^2_p = .162. These data indicate that response time, consistent with distraction research, increases with the addition of the digital display task, but when the prevalence of that task is raised response time does not significantly increases. Accuracy likewise significantly suffers with the addition of the digital display task, but is unaffected by further elevation of SP. See Figure 31 for a visual representation.

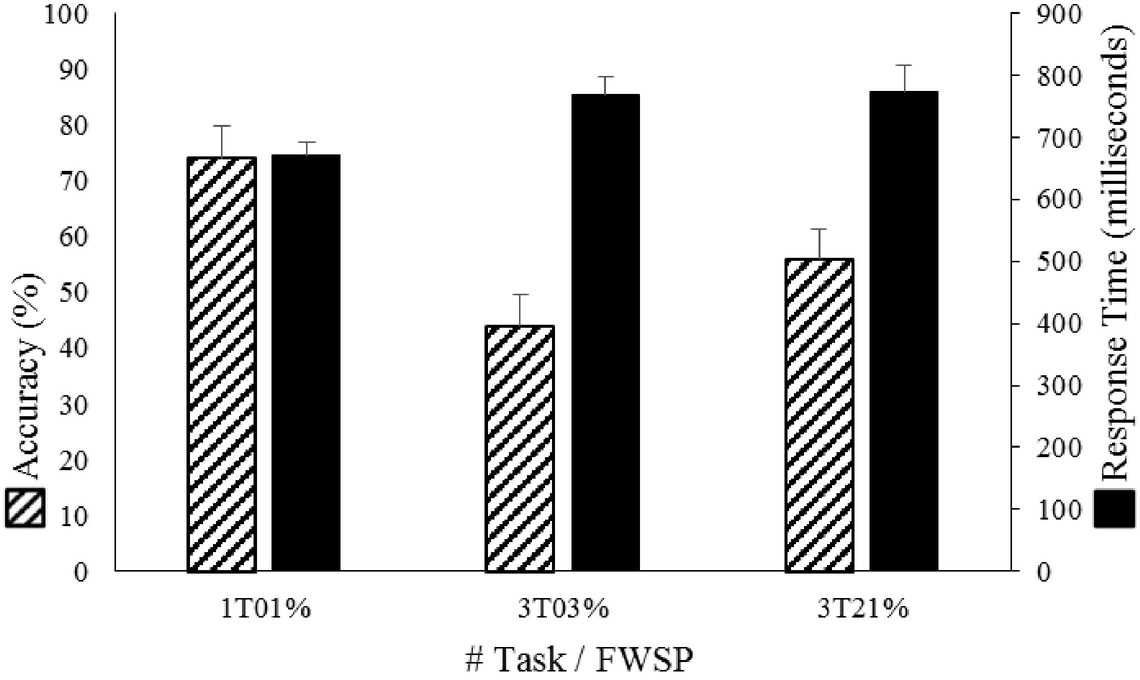


Figure 31: Accuracy and response time across conditions in experiment III
Note that the largest change in accuracy occurs in the transition from single to triple-tasking; the trend toward improved performance seen in increased signal probability (SP) is non-significant. The largest change in response likewise occurs in the transition from single to triple-tasking. Error bars represent within-participants confidence intervals (Cousineau, 2005).

Subjective Workload Dependent Variables

Data from twenty participants ($n = 20$) were included. Descriptive statistics for all subjective workload variables are reported in Table 8. Six dimensions of the NASA TLX (mental demand, physical demand, temporal demand, performance, effort and frustration) as well as the overall workload score were subjected to a within-participants MANOVA to assess the impact of the presentation of visual targets and text messages on a secondary digital display in three levels: monitoring the world baseline (1T01%), low digital all (3T03%) and high digital all (3T21%). No effects of gender or order were seen.

A significant main effect of condition was seen: Wilks' Lambda = .302, $F(12, 66) = 4.503$, $p < .001$, $\eta^2_p = .450$. Univariate ANOVA results were therefore interpreted, revealing a significant effect for the composite TLX, $F(2, 38) = 13.015$, $p < .001$, $\eta^2_p = .407$, and all subscales including mental demand, $F(2, 38) = 16.004$, $p < .001$, $\eta^2_p = .457$, physical demand, $F(2, 38) = 3.308$, $p = .047$, $\eta^2_p = .148$, temporal demand, $F(2, 38) = 10.384$, $p < .001$, $\eta^2_p = .353$, performance, $F(2, 38) = 25.189$, $p < .001$, $\eta^2_p = .570$, effort, $F(2, 38) = 13.514$, $p < .001$, $\eta^2_p = .416$, and frustration $F(2, 38) = 15.350$, $p < .001$, $\eta^2_p = .447$. Visual representations can be found in Figure 32. In aggregate, these data indicate that subjective workload is elevated both by the need to engage in multitasking by monitoring the digital display, and by elevation of SP within the digital display.

Table 8 : SPS Task - Battlefield Threat Detection TLX

DV	Code	# Task	FWSP (%)	*Mean*	*SE*
TLX	1T01%	1	1	41.729	2.142
	3T03%	3	3	51.761	2.041
	3T21%	3	21	57.230	1.137
TLX Mental	1T01%	1	1	37.261	5.135
	3T03%	3	3	63.223	3.861
	3T21%	3	21	73.596	1.998
TLX Physical	1T01%	1	1	15.191	2.759
	3T03%	3	3	17.499	2.030
	3T21%	3	21	24.495	1.992
TLX Temporal	1T01%	1	1	52.420	3.502
	3T03%	3	3	68.431	2.320
	3T21%	3	21	72.596	2.231
TLX Performance	1T01%	1	1	80.459	5.245
	3T03%	3	3	40.196	3.385
	3T21%	3	21	39.204	2.748
TLX Effort	1T01%	1	1	39.570	4.718
	3T03%	3	3	62.955	3.661
	3T21%	3	21	70.937	2.554
TLX Frustration	1T01%	1	1	25.476	4.902
	3T03%	3	3	58.263	4.342
	3T21%	3	21	62.551	3.392

Note. Standard errors calculated as within-participants confidence intervals (Cousineau, 2005).

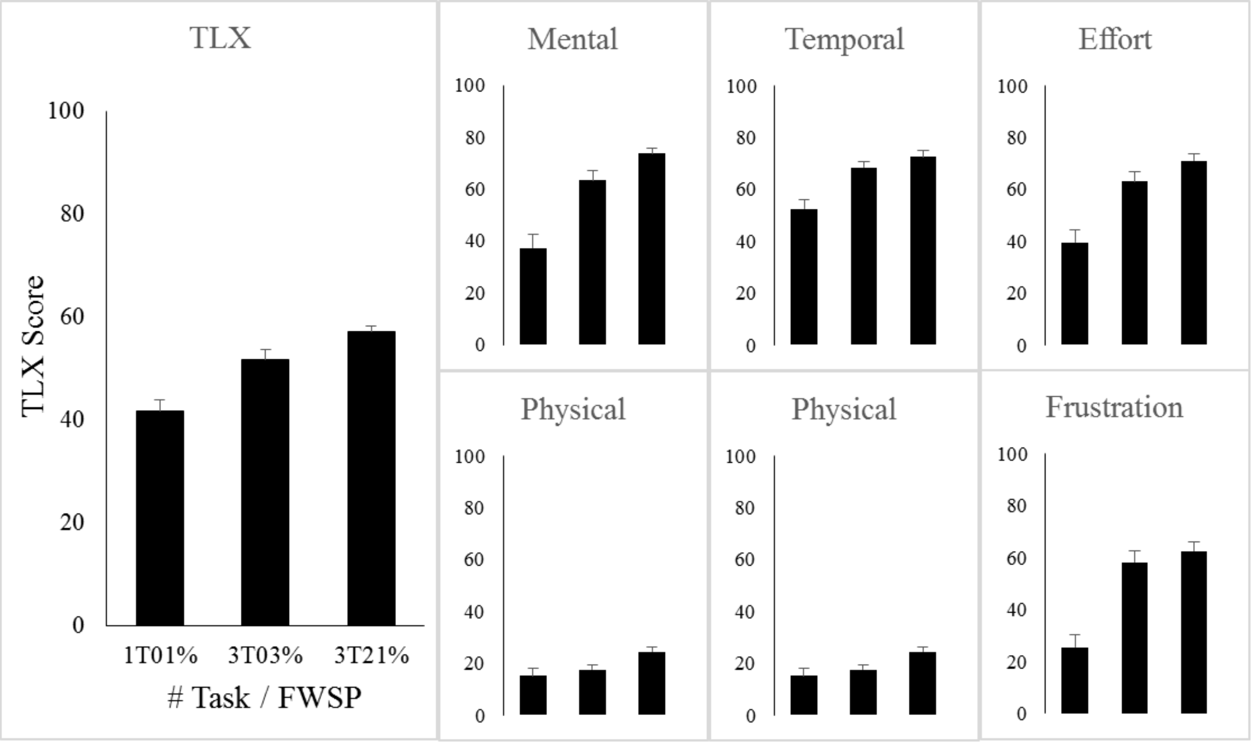


Figure 32: TLX scores across conditions in experiment III
Note that the largest change is due to the change from single to triple-tasking, not the increase from 3% to 21% signal probability (SP) within triple-tasking. Error bars represent within-participants confidence intervals (Cousineau, 2005).

Experiment IV: SPS Task - Signal Probability in Cyberattack

Response Dependent Variables

Data from thirty-three participants (*n* = 33) were included in the present analysis, after two were removed for non-completion and a third was removed for excessive flagging of emails. Descriptive statistics for all response dependent variables are reported in Table 9. Two email attack types: one providing a malicious download and the other requesting an improper upload were presented within subjects at three levels of SP, between subjects. Collected response variables included accuracy and response time to attack emails of both types, but as no significant differences between performance in these attack types was seen they were analyzed together. These data were submitted to a within-between participants MANOVA to assess the impact of these two manipulations in a 2 attack type (upload, download) x 3 SP (1%, 5% and 20%) design. To compensate for violations of the sphericity assumption, when appropriate the Box correction (Field, 2009) has been applied to the following results. No effects of gender or order were seen.

Table 9 : SPS Task - Signal Probability in Cyberattack Dependent Variables

DV	Code	# Task	SP (%)	*Mean*	*SE*
Accuracy (%)	MT01%	M	1	45.5	10.8
	MT05%	M	5	86.6	10.8
	MT20%	M	20	78.0	10.8
Response Time (ms)	MT01%	M	1	7305	601
	MT05%	M	5	5160	631
	MT20%	M	20	4796	601

Note. The email task is multitask in nature, and was not manipulated in the present experiment.

No significant interaction or main effect of attack type were seen, but a significant main effect of SP did emerge, Wilks' Lambda = .572, $F(4, 58) = 4.671$, $p = .002$, $\eta^2_p = .244$. Between-subjects ANOVA results were therefore interpreted, revealing the effect to be significant as related to both response variables: accuracy $F(2, 30) = 4.058$, $p = .028$, $\eta^2_p = .213$ and response time, $F(2, 30) = 4.995$, $p = .013$, $\eta^2_p = .250$. See Figure 33 for a visual representation. These data indicate that when the prevalence of attacks was raised response time decreased while accuracy increased.

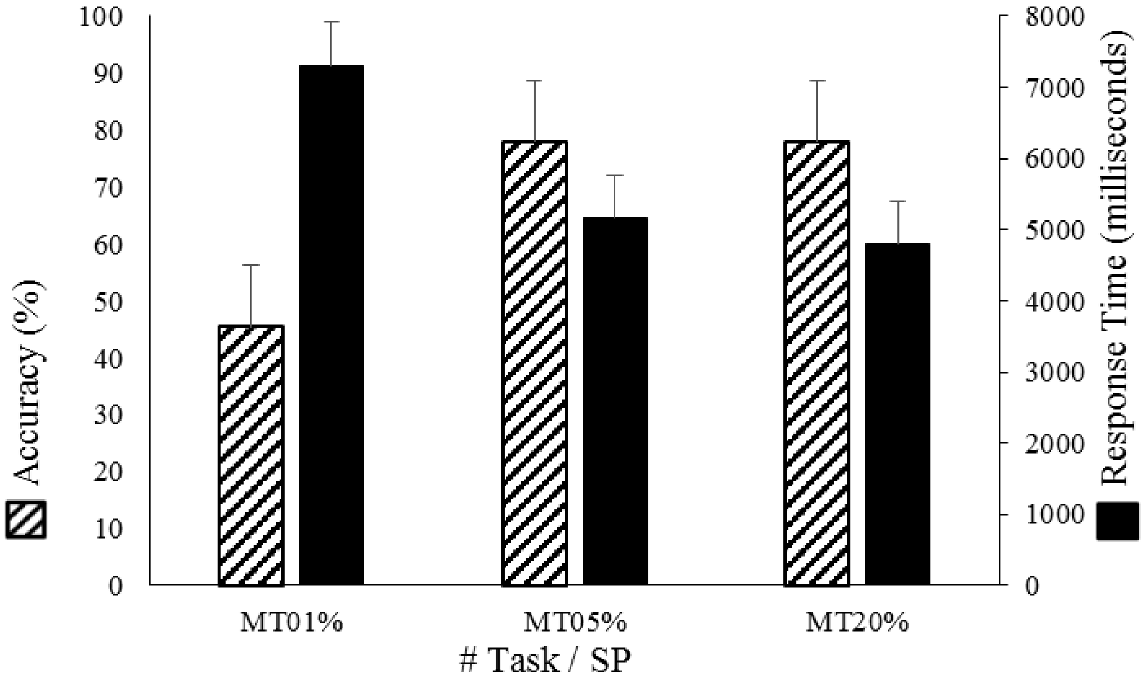


Figure 33: Accuracy and response time across conditions in experiment IV
Note that the large accuracy gain seen between 1% and 5% is significant, while the small drop between 5% and 10% is not. The email task is multitask in nature. The pattern of response, in which that the largest change is due to the change from 1% to 5% SP, should be considered in light of this fact. Many steps and decisions must be made, but at any time the participant can make the decision to end that process and report the email.

Subjective Workload Dependent Variables

Data from thirty-three participants (*n* = 33) were included. Descriptive statistics for all subjective workload variables are reported in Table 10. Six dimensions of the NASA TLX (mental demand, physical demand, temporal demand, performance, effort and frustration) as well as the overall workload score were subjected to a between-participants MANOVA to assess the impact of the presentation of attack emails at three levels of SP: 1%, 5% and 20%. No effects of gender or order were seen. No significant effect of SP was seen: Wilks' Lambda = .662, *F*(12, 50) = .955, *p* = .503, η^2_p = .186.

Table 10 : SPS Task - Signal Probability in Cyberattack TLX

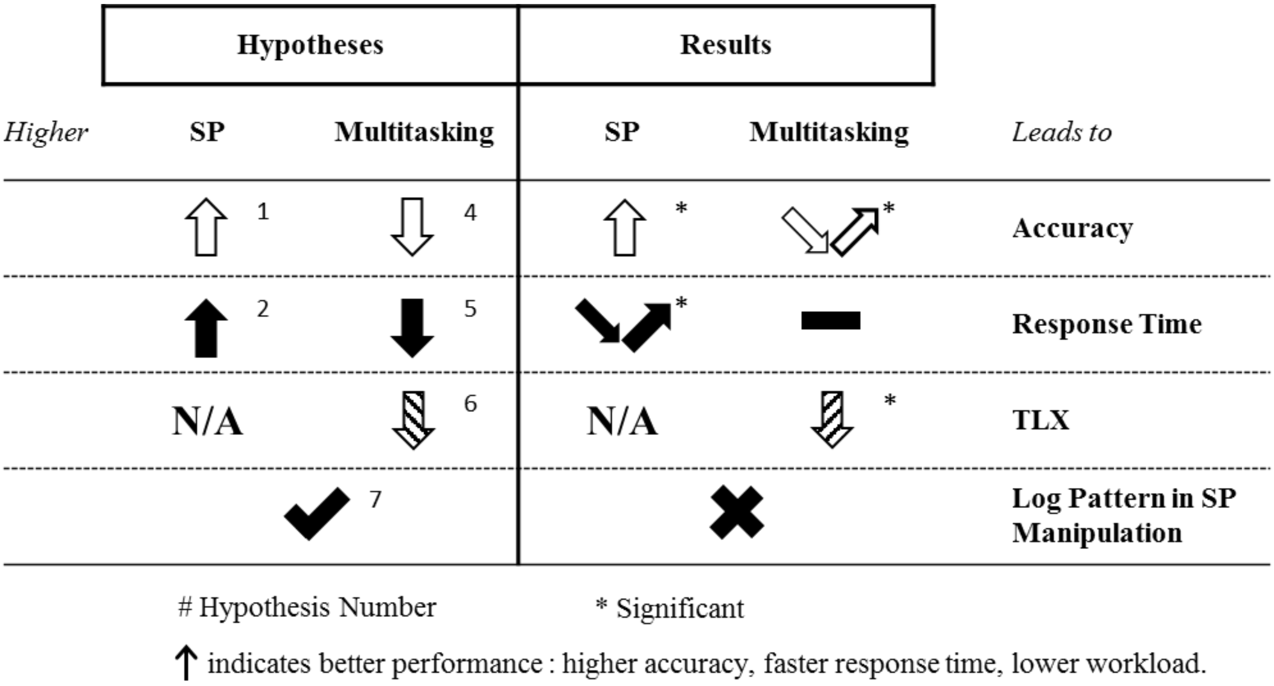
DV	Code	# Task	SP (%)	*Mean*	*SD*
TLX	MT01%	M	1	27.136	4.872
	MT05%	M	5	35.078	4.872
	MT20%	M	20	38.364	4.872
TLX Mental	MT01%	M	1	26.818	6.548
	MT05%	M	5	32.403	6.548
	MT20%	M	20	35.091	6.548
TLX Physical	MT01%	M	1	8.000	5.697
	MT05%	M	5	16.798	5.697
	MT20%	M	20	25.545	5.697
TLX Temporal	MT01%	M	1	30.818	7.963
	MT05%	M	5	42.756	7.963
	MT20%	M	20	32.364	7.963
TLX Performance	MT01%	M	1	45.182	10.840
	MT05%	M	5	39.134	10.840
	MT20%	M	20	37.091	10.840
TLX Effort	MT01%	M	1	25.727	8.010
	MT05%	M	5	36.960	8.010
	MT20%	M	20	47.000	8.010
TLX Frustration	MT01%	M	1	26.273	8.747
	MT05%	M	5	42.418	8.747
	MT20%	M	20	53.091	8.747

Note. These non-significant TLX data trends toward increased demand for higher signal probability conditions, despite the fact that performance was significantly better in lower signal probability conditions.

CHAPTER FIVE: DISCUSSION

In the following pages each experiment will be discussed in terms of hypothesized and reported results, as well as implications. This will be followed by a discussion of which overarching hypotheses were supported, and the implications of these aggregate patterns.

Experiment I: Basic Psychophysical SPS Task



	Hypotheses		**Results**		
Higher	**SP**	**Multitasking**	**SP**	**Multitasking**	*Leads to*
	⇧ 1	⇩ 4	⇧ *	⇘⇗ *	**Accuracy**
	⬆ 2	⬇ 5	↘↗ *	▬	**Response Time**
	N/A	⇩ 6	N/A	⇩ *	**TLX**
	✔ 7		✖		**Log Pattern in SP Manipulation**

\# Hypothesis Number * Significant

↑ indicates better performance : higher accuracy, faster response time, lower workload.

Figure 34: Hypothesized and reported results for experiment I
The first line may be read: “In the hypothesis one higher SP was predicted to lead to better accuracy performance and in hypothesis four higher multitasking was predicted to lead to lower accuracy performance. Results show higher SP led to significantly better accuracy performance and higher multitasking led to a significant mixed pattern of performance.”

In terms of accuracy, these data (Figure 34) make it clear that it increases as SP increases, supporting hypothesis one. Performing more tasks concurrently did not lead to lower-accuracy performance, as suggested in hypothesis four, and greatest performance was instead seen when participants searched for three types of targets concurrently, followed by single targets, and two types of targets. This surprising finding is better illustrated by the interaction shown in Figure 26, in which the impact of lower SP is significantly greater under conditions of single and dual-tasking. This pattern suggests that, when triple-tasking, the decrement associated with low SP search is lessened as compared to single and dual-tasking. Note that this pattern stands in opposition to those reported by Wolfe et al. (2005), who investigated a similar scenario in order to determine if mixing high SP and low SP targets might alleviate the low rate at which rare targets were detected. In the present experiment concurrent search for the three types of targets had little effect upon pedestrians, presented at 35%. Significant gains of increasing magnitude were seen for text messages, presented at 10%, and braking cars, presented at 1%. This suggests, to use Wolfe's suggestion (2005), baggage screeners searching for rare guns might benefit from also searching for more common objects, such as digital music players (iPods). This argument is a part of a striking overall pattern of accuracy: as participants do more, in terms of SP, the impact of the number of tasks they engage in grows less. At the 35% level of SP this results in near parity of accuracy performance (but not of response time performance). As noted before, the trajectories of these levels of multitasking beyond 35% is a worthwhile question, and will be identified as a possible future direction for research.

Higher SP led to better response time, as in hypothesis two, only in single-tasking. In both dual and triple-tasking this pattern fails to hold as faster performance appears to be seen in the braking cars presented at 1% SP. This pattern may in fact be an artifact of speed-accuracy tradeoffs: the score calculated is only for correct detections, and very low rates of accuracy seen at the 1% SP in this experiment may suggest that only participants able to detect and respond quickly are measured. The pattern shows very little variance over levels of multitasking, leading to an inconclusive answer to hypothesis five.

Increases in the number of concurrent tasks led to greater reported workload, supporting hypothesis six. Note that while participants performed better in terms of accuracy under conditions of triple-tasking, this is not reflected in their TLX scores. The change from single to dual-tasking, which had a negative impact upon accuracy performance, was reported as the chief contributor to elevated workload, suggesting some sensitivity on the part of participants to the actual pattern of performance.

Hypothesis seven concerned the log nature of the pattern expected within each level of SP, and was not supported. The best fit, in terms of R^2 and as shown in Figure 28, was for a linear pattern. These data nonetheless show that decrements related to SP are greatest at the lowest signal probabilities. Level of multitasking here determines the slope of the line describing the decrement. It is again an interesting question how the pattern behaves at signal probabilities greater than 35%. One possibility is that the linear trend-lines are correct, and more favorable levels of multitasking, such as triple-tasking, continue to show an accuracy benefit. In light of the evidence discussed earlier, it seems more likely that higher signal probabilities show little difference among multitasking levels. Regardless, as higher SP results in greater performance no participant can in fact exceed perfect accuracy, and this limitation of reality may in fact render the entire pattern closer in function to the log fit predicted.

Experiment II: SPS Task - Multitasking in Driving Simulation

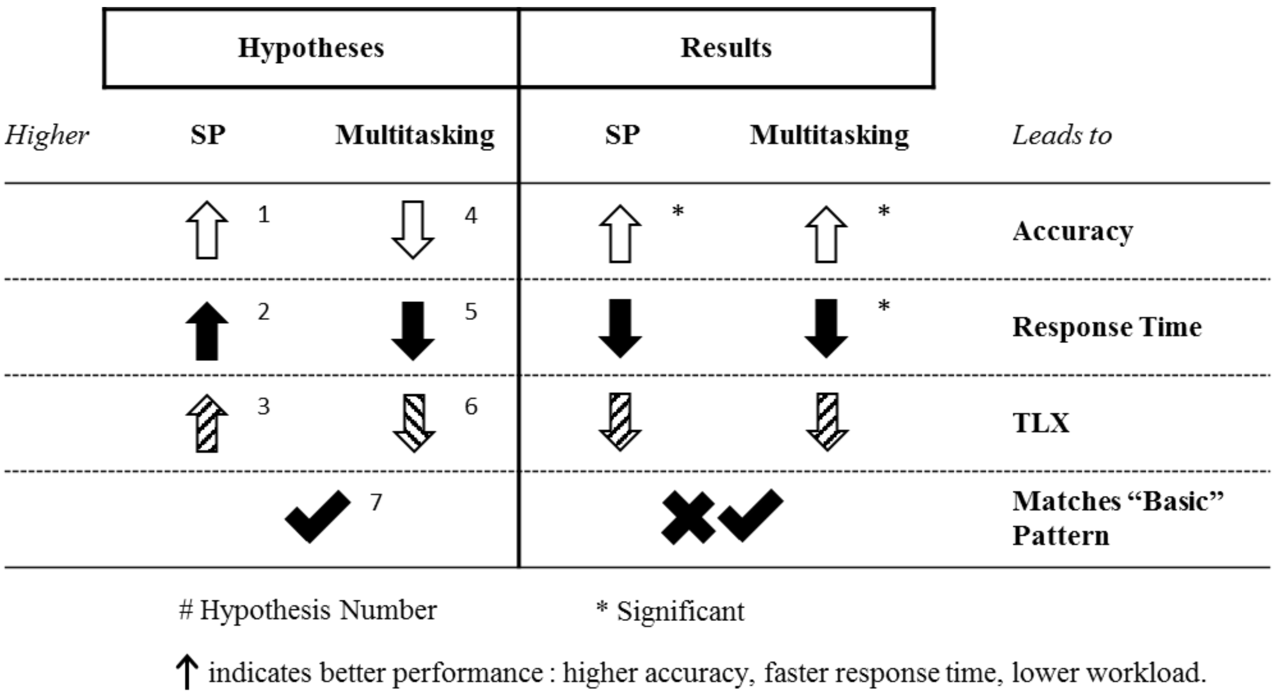


Figure 35: Hypothesized and reported results for experiment II
The first line may be read: "In the hypothesis one higher SP was predicted to lead to better accuracy performance and in hypothesis four higher multitasking was predicted to lead to lower accuracy performance. Results show higher SP led to significantly better accuracy performance and higher multitasking led to a significant better accuracy performance."

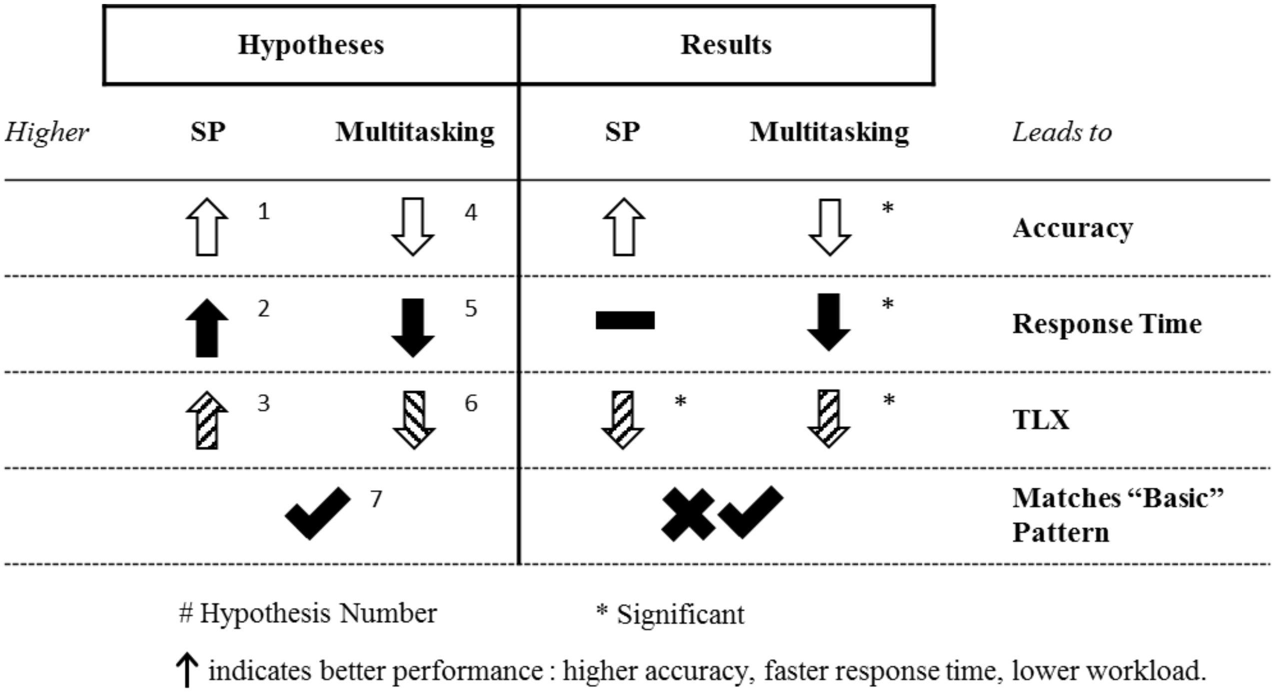
A brief thought experiment will reveal that collision rate can be reframed as the inverse of accuracy (a collision is a miss in and that appropriate avoidance of the braking car stimuli is a correct response). As such, higher SP led to better (higher) accuracy (Figure 35), as previously suggested in hypothesis one. A greater number of concurrent tasks likewise led to better (higher) accuracy, in opposition to the predictions of hypothesis four, but in support of the pattern observed in experiment 1 (psychophysical). It is worth noting, as shown in Table 5, that the calculated FWSP for the 2T02% condition was in fact 11%, and in the 2T08% condition was 61.3%. Because the brake car was necessarily reacted to (or not, in the case of collisions) in the space of a single event section, this means that the texting task manipulation occupied 10%, followed by 60.3% of participants time (and space) on task. All participants finished the texting task before encountering the braking car. The orderly progression of decreased collisions, illustrated in Figure 30, is therefore unrelated to so-called structural distraction factors, and instead a result of hysteresis effects upon the participants cognitive state at the time of brake car presentation. The greater accuracy at higher SP levels of stimuli presentation is in line with much previous vigilance research (Mackworth, 1970; Warm & Jerison, 1984; Sawyer et al., 2014b), as well as visual search investigations of the prevalence effect (Fleck & Mitroff, 2007, Wolfe et al., 2007). That greater multitasking likewise led to better (higher) accuracy is with little precedence in the literature, but can be framed in the Hancock & Warm Model of workload as a case of load lifting a population out of an underload state.

While accuracy increased, response time suffered under higher levels of multitasking as predicted in hypothesis five. Participants in the 1T01% group were significantly quicker to respond than those in the 2T02% group. A non-significant trend toward slower response is seen between the 2T02% and 2T08% groups, indicating that this large ramping of SP in the texting condition had little effect. The finding that multitasking impacts response time is in line with much previous driving distraction work (Strayer & Johnston, 2001; Chisholm, Caird, & Lockhart, 2008; Mouloua, Hancock, Rinalducci, & Brill, 2003; Drews, Yazdani, Godfrey, Cooper, & Strayer, 2009; Sawyer, 2010; Sawyer & Hancock, 2012; Sawyer et al., 2014). The lack of contribution to response decrement from greatly expanded levels of exposure (as measured both by intended and actual SP) to the secondary task is surprising.

In partial support of hypothesis seven, significant results in experiment II match the findings from the experiment I basic psychophysical task both in terms of the contribution of SP to accuracy, and in terms of greater multitasking's contribution to slower response time.

In the context of driving, these findings in aggregate suggest that while engaging in a secondary task while driving does impose costs in terms of response time, it may also increase the likelihood that a driver responds in the first place. Further, the response time decrement may not significantly vary depending upon the SP of the secondary task, while higher SP involvement with the secondary task continues to improve likelihood of response and so reduce collisions. From a practical point of view, this suggests that in situations where secondary tasks are unavoidable efforts to reduce their frequency may not have the intended effect of reducing crashes. Further, in situations where roadway obstacles are unlikely to appear without warning, such as convoy driving, a secondary task may actually be desirable. These ideas should, however, be understood within the study's limitations: no structural distraction was tested, and the impact of situations where the driver looks away from the road at the wrong moment (as in Sawyer, Calvo, Finomore & Hancock, 2015) are not accounted for.

Experiment III: SPS Task – Battlefield Threat Detection



	Hypotheses		Results		
Higher	**SP**	**Multitasking**	**SP**	**Multitasking**	*Leads to*
	⇧ 1	⇩ 4	⇧	⇩ *	**Accuracy**
	⬆ 2	⬇ 5	—	⬇ *	**Response Time**
	⇑ 3	⇓ 6	⇓ *	⇓ *	**TLX**
	✓ 7		✗✓		**Matches "Basic" Pattern**

Hypothesis Number * Significant

↑ indicates better performance : higher accuracy, faster response time, lower workload.

Figure 36: Hypothesized and reported results for experiment III
The first line may be read: "In the hypothesis one higher SP was predicted to lead to better accuracy performance and in hypothesis four higher multitasking was predicted to lead to lower accuracy performance. Results show higher SP led to a non-significant trend toward better accuracy performance and higher multitasking led to a significant decline in accuracy performance."

In terms of accuracy, the change between the 1T01% and 3T03% conditions resulted in decreased performance, in line with hypothesis four (Figure 36). Although a trend toward better performance was seen in the change between the 3T03% and 3T21% conditions, in line with hypothesis one, it was not significant. In terms of response, the change between the 1T01% and 3T03% conditions resulted in significantly slower performance as expected in hypothesis five, but in the change between the 3T03% and 3T21% conditions resulted in a non-significant pattern, and no notable trend. The increase in number of concurrent tasks between the 1T01% and 3T03% conditions led to greater reported workload, in opposition to hypothesis three. The increase in SP between the 3T03% and 3T21% conditions also led to greater reported workload, supporting hypothesis six.

In partial support of hypothesis seven, significant results in experiment III match the findings from the experiment I basic psychophysical task in terms of greater multitasking's contribution to both slower response time and higher reported workload.

In the context of dismounted battlefield operations, these findings in aggregate suggest first that engaging in additional tasks results in decrements in identifying and engaging threats. The decision to impose additional tasks comes at a cost to the primary task of identifying and appropriately engaging threats. This cost can be compared to multitasking-based distraction in driving, and should be taken into account when deciding what equipment and tasks are appropriate for dismounted soldiers. Battlefield distraction shows evidence of endangering warfighters and mission success alike.

Experiment IV: SPS Task - Signal Probability in Cyberattack

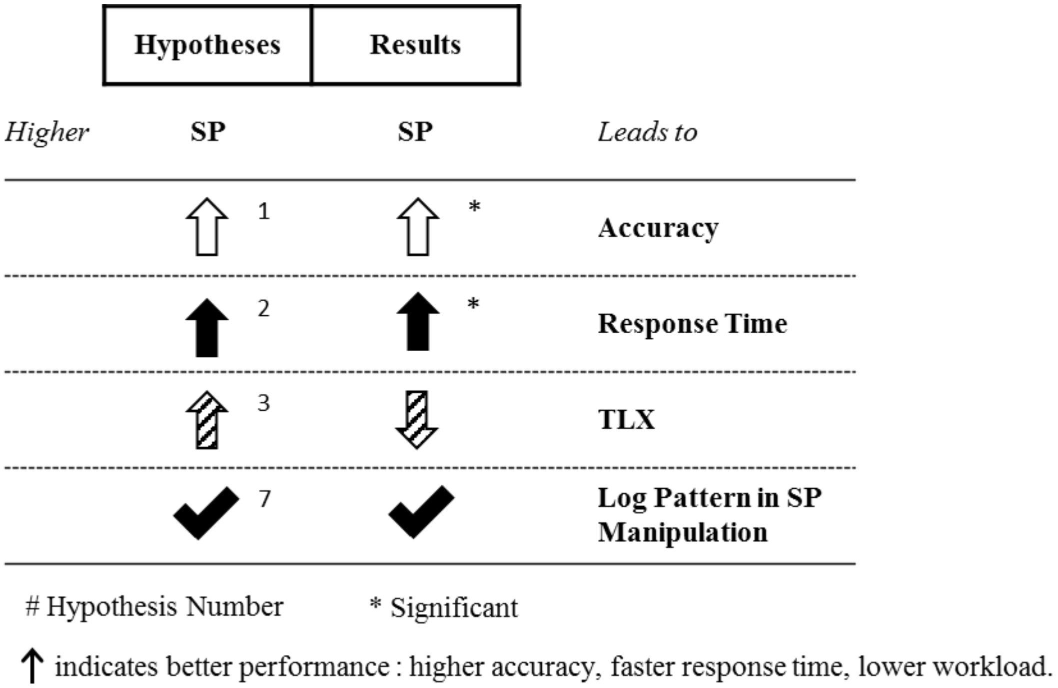


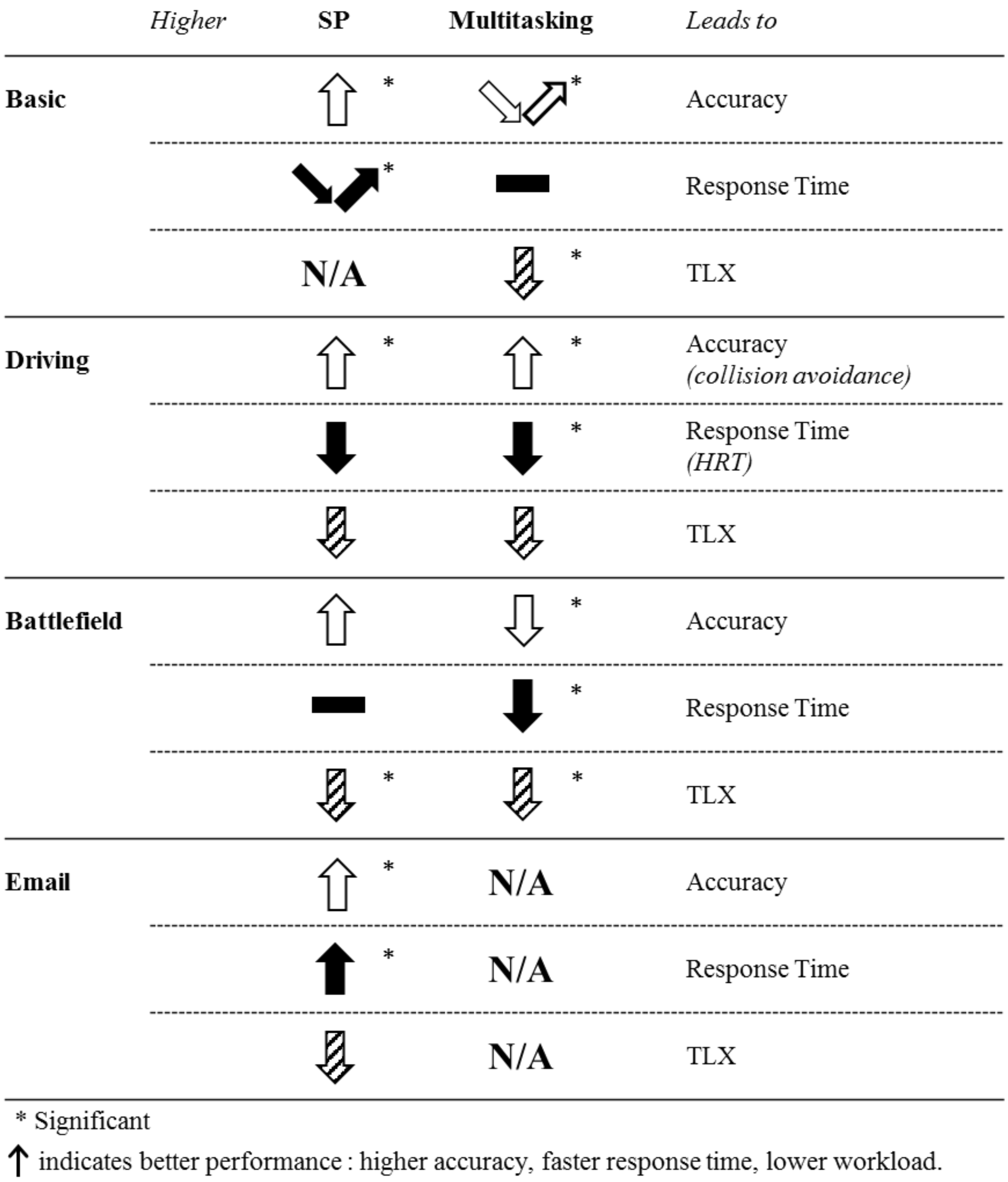
Figure 37: Hypothesized and reported results for experiment IV
The first line may be read: “In the hypothesis one higher SP was predicted to lead to better accuracy performance and in hypothesis four higher multitasking was predicted to lead to lower accuracy performance. Results show higher SP led to significantly better accuracy performance and higher multitasking led to a significant better accuracy performance.”

This experiment did not manipulate levels of multitasking (Figure 37). As predicted in hypothesis one, higher SP resulted in significantly higher accuracy. As predicted in hypothesis two, higher SP resulted in significantly faster response time. Hypothesis three was not supported, and a non-significant trend was in the direction of higher SP resulting in greater workload. Hypothesis seven concerned the log nature of the accuracy pattern expected within levels of SP, and was supported. This log pattern has considerable import in the context of cyberattack and defense, and these findings in aggregate suggest, as suggested above. There may therefore be an optimal level of attack signals in order for the human to serve as an effective detector, and at the lowest end of SP percentage the drop-off in that efficacy may be logarithmic in nature and grievous indeed. This log pattern stands in opposition to the pattern found in experiment I (psychophysical), and adds to the argument that those data may in fact represent a log pattern. One possibility for this discrepancy lies in the training provided for both studies: in experiment I participants who could not perform to criterion on the final phase of training were removed from the study. In the present experiment participants arrived with considerable experience in using email interface, and no such training measure was required. It is possible that in selecting away individuals unable to comprehend training, experiment I was deprived of the bottom of its distribution, producing a linear pattern.

Finally, experiment IV served as a test for the principles for arguing the distractive nature of a task outlined in chapter 1. Although a decrement was seen in some conditions, the thought experiment carried out in methods renders it unlikely these can be considered a form of distraction. One possibility top consider is that is that this is a result of the nature of the manipulation: SP alone. Mere elevation of a specific subtask's frequency does not seem to meet the bar. Distraction is in this work defined a performance decrement in one task are due to attentional allocation to a concurrent task, and it appears that this concurrent task may in fact need to be discrete, as is the case in all other experiments here presented

Overarching Hypotheses and Aggregate Patterns

Bridging a common set of research questions across experiments both psychophysical and applied provided a rich field of results (Figure 38, 39) from which to draw conclusions. In aggregate, the findings of the four experiments support the supposition that better control of time (and therefore space) reveals nomothetic principles. Most pronounced among the aggregate patterns seen (Figure 39) is that related to overarching hypothesis one (1), which predicted that higher SP conditions would lead to higher accuracy. This was supported by significant results in all experiments but III (battlefield), where a non-significant trend is seen. Prevalence effects, therefore, seem quite robust beyond vigilance settings in applied, operationally complex contexts. In such contexts, including driving, common wisdom states that requiring humans to do more leads to greater workload and reduced performance. In contrast, the present pattern resonates more with the view forwarded by the Hancock-Warm Model (1989), in which greater arousal may, in situations of underload, lead to improved performance.



	Higher	**SP**	**Multitasking**	*Leads to*
Basic		⇧ *	⬂⬀ *	Accuracy
		⬊⬈ *	▬	Response Time
		N/A	⇓ *	TLX
Driving		⇧ *	⇧ *	Accuracy *(collision avoidance)*
		⬇	⬇ *	Response Time *(HRT)*
		⇓	⇓	TLX
Battlefield		⇧	⇩ *	Accuracy
		▬	⬇ *	Response Time
		⇓ *	⇓ *	TLX
Email		⇧ *	N/A	Accuracy
		⬆ *	N/A	Response Time
		⇓	N/A	TLX

* Significant

↑ indicates better performance : higher accuracy, faster response time, lower workload.

Figure 38: Dependent variable trends as a function of SP and multitasking by experiments
Color coding here allows comparison across manipulation by dependent variable. For example, by focusing on the white arrows that signify accuracy results in the ‘Higher SP’ column, it can be seen that in all experiments a significant effect (3) or trend (1) toward higher SP eliciting greater accuracy was found. Where two arrows are shown, a mixed effect was found. Bars signify no significant results or discernable trend.

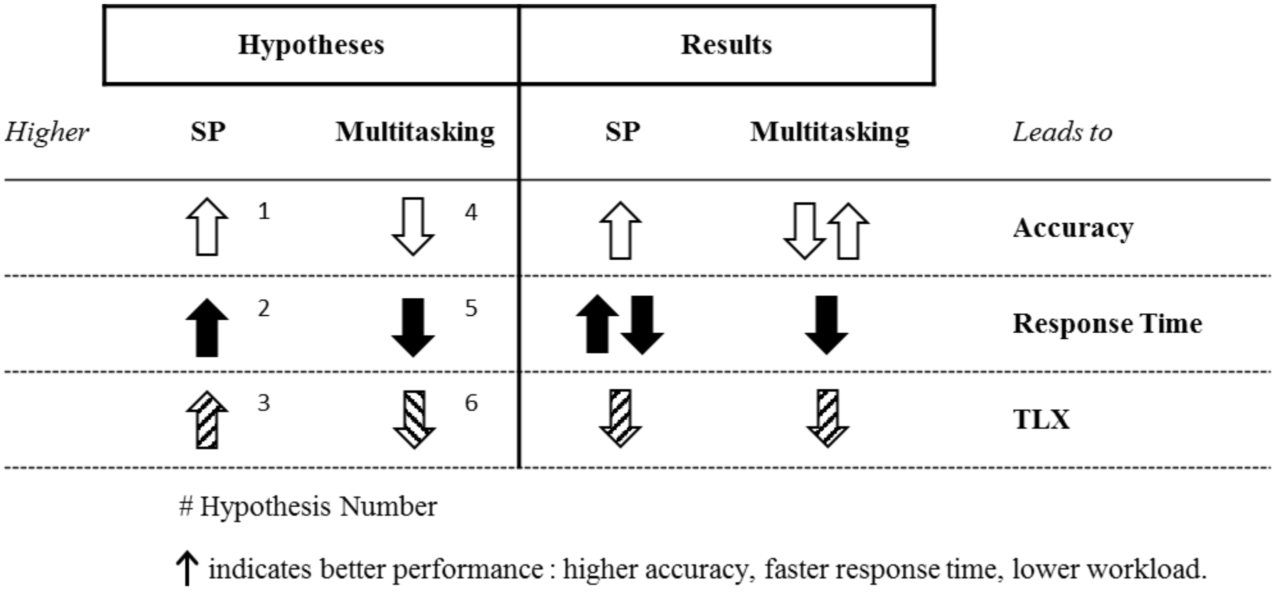


Figure 39: Overarching hypotheses compared to aggregate patterns
Each result on the right was computed by looking at all significant and trend results from table 38. Where two opposing arrows are shown, mixed results were found.

This effect of signal probability upon accuracy (Figure 39) is even more interesting in light of the aggregate pattern found in data relating to overarching hypothesis three (3) and six (6), which had initially predicted that higher SP would lead to better (lower) workload, while performing more tasks concurrently would lead to poorer (higher) workload. In all experiments where such data existed, elevating either SP or number of tasks also led to elevated (poorer) workload scores. This suggests that humans may be poor judges of the accuracy boost that accompanies presentation of stimuli at higher signal probability, perhaps employing a heuristic involving the overall amount of time they find themselves occupied in lieu of some more accurate assessment of performance. This supposition mirrors a broader question, alluded to earlier, of whether humans are indeed able to accurately report their own internal states and experiences (Natsoulas, 1967).

Another aggregate pattern may be found in data relating to overarching hypotheses five (5), which predicted that performing more tasks concurrently would result in slower (higher in terms of ms) response times. In both experiments II and III (driving, targeting) significant results support this hypothesis. This finding, in combination with accuracy and SP findings related to overarching hypothesis one (1), paints an interesting picture of accuracy and response time as they relate to SP and multitasking. Consider the possibility that findings from both driving distraction work and vigilance/prevalence work are correct: accuracy is sensitive to level of target density (SP) just as response time is sensitive to number of tasks performed (multitasking). As levels of one are so often dependent upon levels of the other, and because the speed-accuracy tradeoff might often mask these independent changes, these effects might well have remained entwined in decades of data. The present dataset is not enough to conclusively make such a claim, and additional work will be needed to understand the full pattern.

Of course, we must not neglect the question of distraction itself. Engaging in multiple tasks increases RT, a finding that spans multiple contexts in the present work. Recall the earlier definition of distraction: “When, in the course of multitasking, performance decrements in one task are due to attentional allocation to a concurrent task, the condition can be viewed as a distraction from the first task.” As such, it is possible that this definition should in fact be limited to reflect no all response decrements, but response time alone, as mixed results on accuracy were seen. That said, in a stochastic world in which response time can be ‘tested’ at any level, poorer response time is, in events which move beyond its threshold, indicative of poorer accuracy. In these experiments, in order to separate the effects of signal probability and multitasking, criterion response time was controlled and intentionally long. As such, although we here parse the effects of SP and MT, in real-world situations there is considerably more noise. As such, the present results perhaps best generalize to situations where the criterion response time is more often long, such as long-haul trucking or caravan driving. In light of this, we will retain the original definition, with one caveat: the prevalence effect can and will exert an influence.

What attracts attention? One interpretation of the present study is that past successes in signal detection beget enhanced attention to the signal-bearing channel; information-bearing channels attract future attention. This strategic view of the prevalence effect does rely on the idea that the enhanced accuracy seen in such tasks is the result of enhanced attention, as opposed to simply a reduced criterion (beta). It nonetheless fits the foraging profile of organisms in search of food, mates, and shelter. As information-foraging creatures (predators of information, even), humans’ propensity to return to rich feeding grounds rings true. It also stands to reason that humans would not be adept at surviving the information famine of low SP environments, preferring instead to move to new, more meaning-bearing locale. The present research does not speak to what happens in cases of information-glut, and such might be an interesting future investigation in pursuit of better supporting the present metaphor.

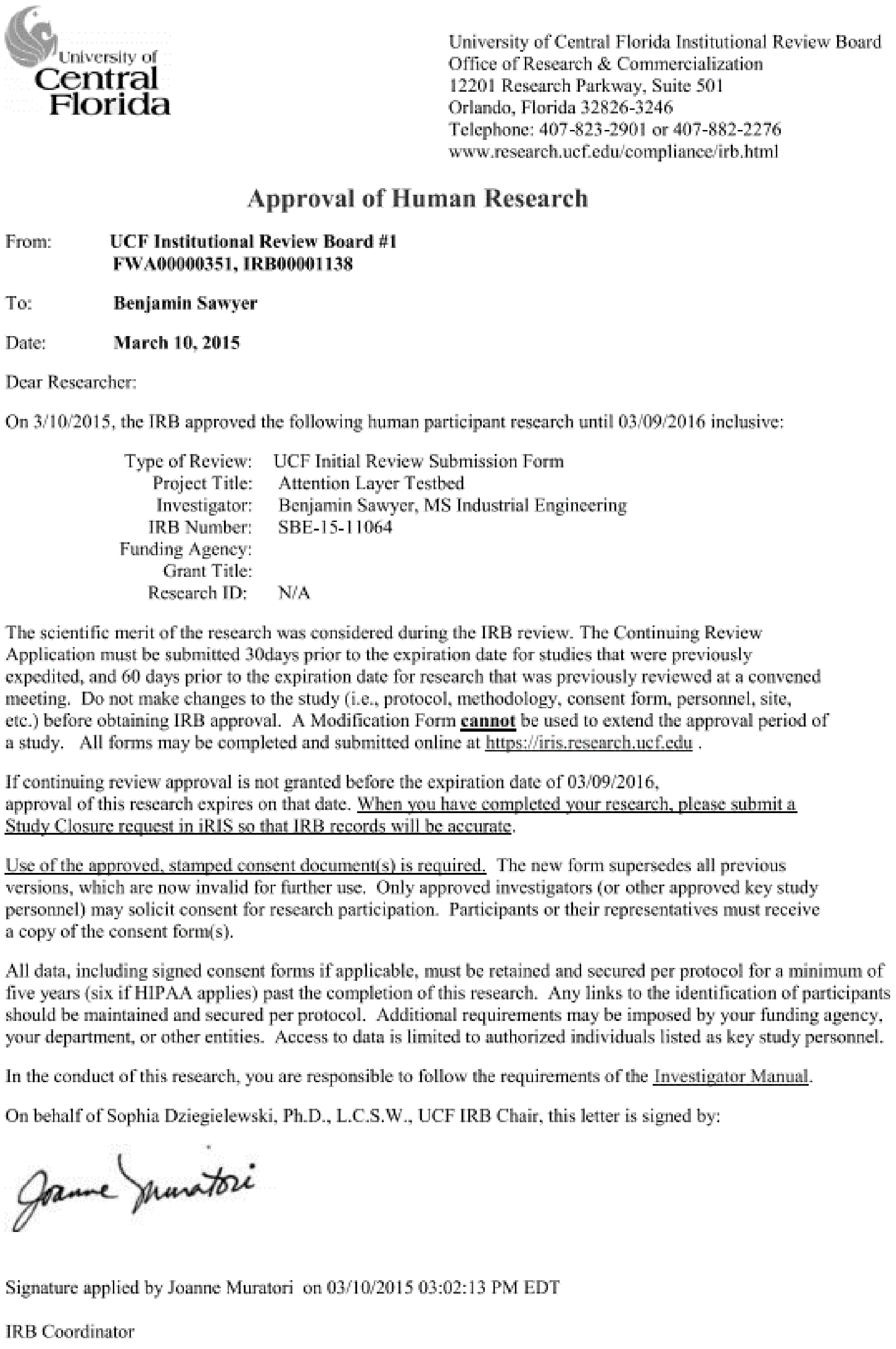
Looking back upon this work, knowing the outcomes, a number of refinements and further avenues of study become clear. Such hindsight, for example, might be focused on repercussions of findings from experiment I (psychophysical), where triple-tasking outperformed single and dual-tasking, and differences between SP levels at the triple-tasking level were the smallest of the three groups. Specifically, the strategy employed in experiment III (driving, battlefield), which tested first the change between single-task detection of a 1% SP group and a group with minimally inflated SP but a greater number of tasks, followed by the change between that group and a group with greatly inflated SP. The latter transition, intended to be made at a level which maximized differences might instead have masked them. Moreover, the findings in experiment I (psychophysical) beg for an expansion of the number of tasks and range of SP tested. A full 4x4 or 5x5 design, either performed over many sessions or between subjects, would provide answers to many of the most perplexing questions to come out of this analysis. In addition to a larger design, more widely used stimuli and normed stimuli (for example, the Snodgrass & Vanderwart object pictorial set) might replace the custom-built set specific to the context of driving used in the present effort. Finally, despite the great effort obviously entailed, it seems the results detailed here suggest that each applied experiment should be run in an expanded design as near as possible to that suggested above for experiment I (psychophysical). Indeed, many similar experiments in varied contexts are recommended, as the construct of distraction is tested far beyond its original domain of driving.

This work used a novel set of methods across multiple contexts to test a fundamental human performance question: Would increasing the signal probability (SP) of a secondary task increase associated performance, as is seen in the prevalence effect associated with vigilance tasks? Would it reduce associated performance, as is seen in driving distraction tasks? A series of experiments weighed these competing assumptions, testing the question in a basic psychophysical task, as well as in the applied contexts of driving, cyberattack and battlefield threat detection. In each, and in line with previous prevalence effect inquiry, presentation of stimuli at higher SP led to higher accuracy. In line with existing distraction work, performance of higher numbers of concurrent tasks tended to elicit slower response times. In all experiments raising either number of concurrent tasks or SP of targets resulted in greater subjective workload, as measured by the NASA TLX, even when accompanied by improved accuracy.

These findings support the view that superior experimental control of signal probability reveals nomothetic patterns of performance that allow the application of the distraction construct in diverse contexts beyond driving. The present data in fact make it very clear that the use of SPS methods, steeped in vigilance methodology, have allowed a much finer-grained view of the performance trade-offs associated with multitasking-based failures of perception and action. These data are the result of the use of methods which can precisely account for experimental variables in time and space by accurately specifying and measuring SP. They are one element implicating such failures as distraction, even in contexts where the construct has historically been unused. Tools of actuarial logic for identifying strong cases of task-primacy are the second element allowing the identification of, for example, battlefield distraction, as a reality in need of intervention. Beyond this need lie further contexts where distraction holds sway and has vital costs. It is hoped that the present work will be a foundation for theory and experimentation allowing identification, intervention, and improved outcomes for related populations.

APPENDIX A: IRB APPROVAL LETTERS

Experiment I



University of Central Florida

University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Human Research

From: **UCF Institutional Review Board #1**
FWA00000351, IRB00001138

To: **Benjamin Sawyer**

Date: **March 10, 2015**

Dear Researcher:

On 3/10/2015, the IRB approved the following human participant research until 03/09/2016 inclusive:

Type of Review:	UCF Initial Review Submission Form
Project Title:	Attention Layer Testbed
Investigator:	Benjamin Sawyer, MS Industrial Engineering
IRB Number:	SBE-15-11064
Funding Agency:	
Grant Title:	
Research ID:	N/A

The scientific merit of the research was considered during the IRB review. The Continuing Review Application must be submitted 30days prior to the expiration date for studies that were previously expedited, and 60 days prior to the expiration date for research that was previously reviewed at a convened meeting. Do not make changes to the study (i.e., protocol, methodology, consent form, personnel, site, etc.) before obtaining IRB approval. A Modification Form **cannot** be used to extend the approval period of a study. All forms may be completed and submitted online at https://iris.research.ucf.edu .

If continuing review approval is not granted before the expiration date of 03/09/2016, approval of this research expires on that date. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

Use of the approved, stamped consent document(s) is required. The new form supersedes all previous versions, which are now invalid for further use. Only approved investigators (or other approved key study personnel) may solicit consent for research participation. Participants or their representatives must receive a copy of the consent form(s).

All data, including signed consent forms if applicable, must be retained and secured per protocol for a minimum of five years (six if HIPAA applies) past the completion of this research. Any links to the identification of participants should be maintained and secured per protocol. Additional requirements may be imposed by your funding agency, your department, or other entities. Access to data is limited to authorized individuals listed as key study personnel.

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

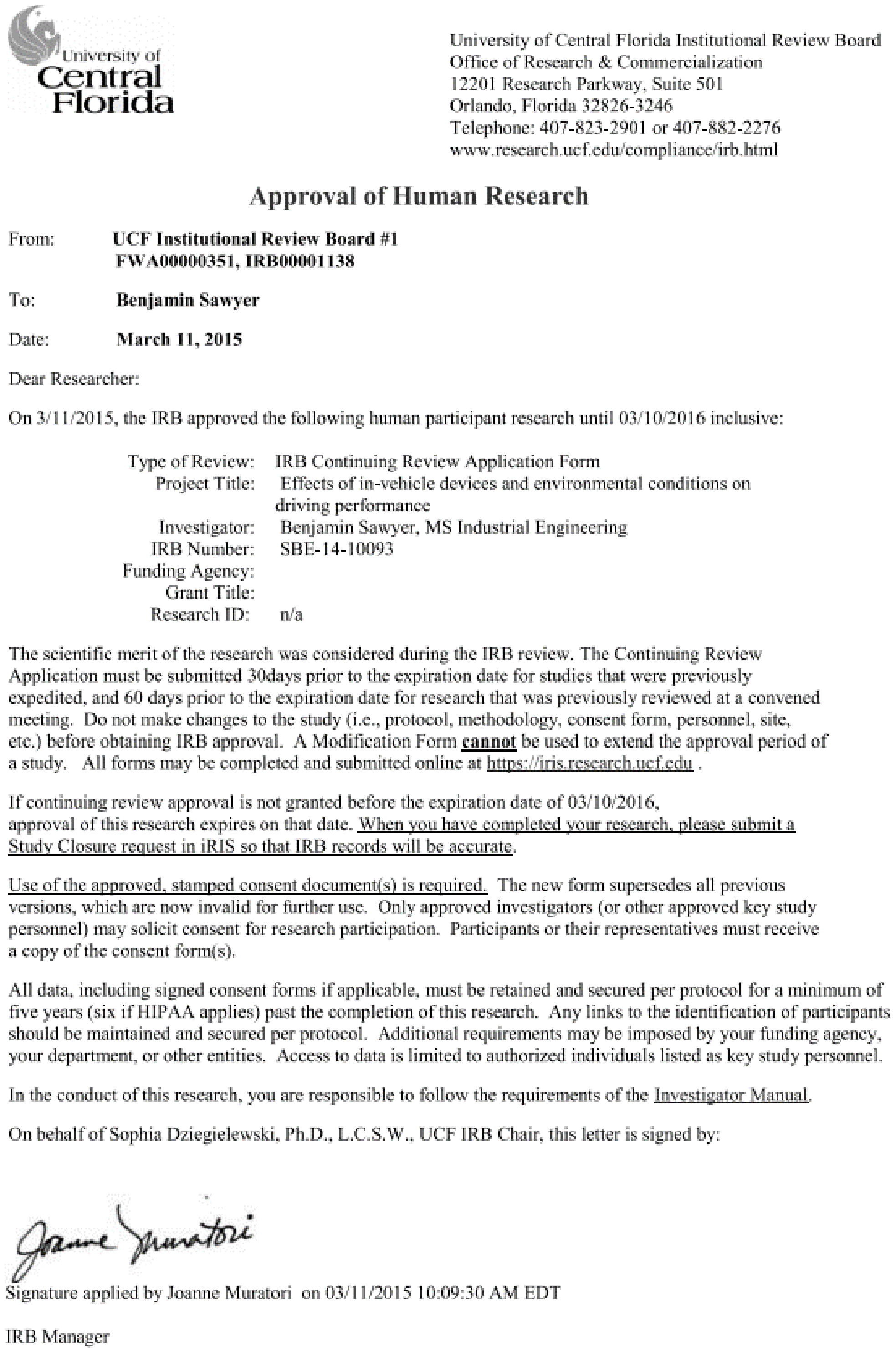
On behalf of Sophia Dziegielewski, Ph.D., L.C.S.W., UCF IRB Chair, this letter is signed by:

Joanne Muratori

Signature applied by Joanne Muratori on 03/10/2015 03:02:13 PM EDT

IRB Coordinator

Experiment II



University of Central Florida

University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Human Research

From: **UCF Institutional Review Board #1**
FWA00000351, IRB00001138

To: **Benjamin Sawyer**

Date: **March 11, 2015**

Dear Researcher:

On 3/11/2015, the IRB approved the following human participant research until 03/10/2016 inclusive:

Type of Review:	IRB Continuing Review Application Form
Project Title:	Effects of in-vehicle devices and environmental conditions on driving performance
Investigator:	Benjamin Sawyer, MS Industrial Engineering
IRB Number:	SBE-14-10093
Funding Agency:	
Grant Title:	
Research ID:	n/a

The scientific merit of the research was considered during the IRB review. The Continuing Review Application must be submitted 30days prior to the expiration date for studies that were previously expedited, and 60 days prior to the expiration date for research that was previously reviewed at a convened meeting. Do not make changes to the study (i.e., protocol, methodology, consent form, personnel, site, etc.) before obtaining IRB approval. A Modification Form **cannot** be used to extend the approval period of a study. All forms may be completed and submitted online at https://iris.research.ucf.edu .

If continuing review approval is not granted before the expiration date of 03/10/2016, approval of this research expires on that date. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

Use of the approved, stamped consent document(s) is required. The new form supersedes all previous versions, which are now invalid for further use. Only approved investigators (or other approved key study personnel) may solicit consent for research participation. Participants or their representatives must receive a copy of the consent form(s).

All data, including signed consent forms if applicable, must be retained and secured per protocol for a minimum of five years (six if HIPAA applies) past the completion of this research. Any links to the identification of participants should be maintained and secured per protocol. Additional requirements may be imposed by your funding agency, your department, or other entities. Access to data is limited to authorized individuals listed as key study personnel.

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

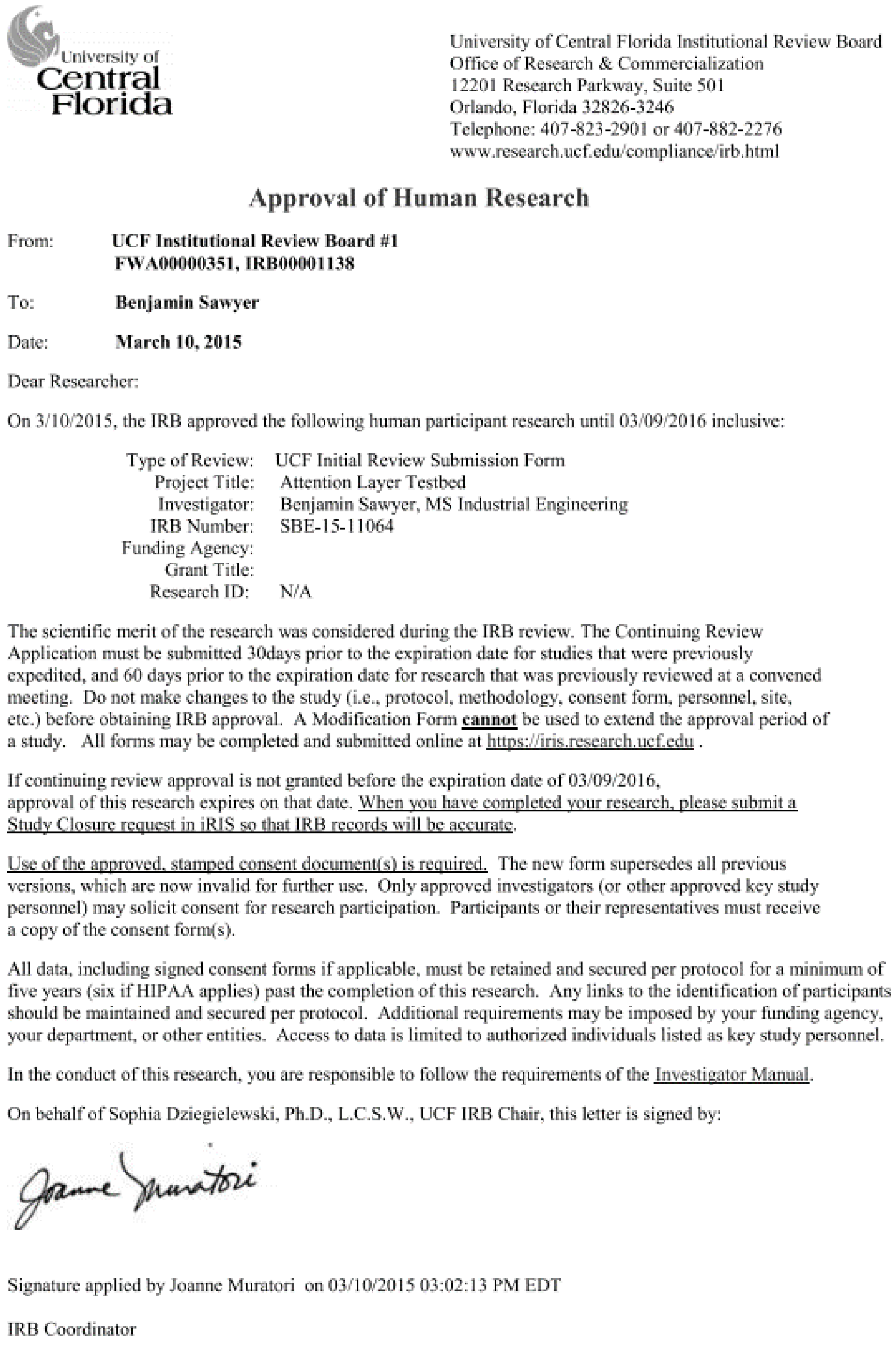
On behalf of Sophia Dziegielewski, Ph.D., L.C.S.W., UCF IRB Chair, this letter is signed by:

Joanne Muratori

Signature applied by Joanne Muratori on 03/11/2015 10:09:30 AM EDT

IRB Manager

Experiment III



University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Human Research

From: **UCF Institutional Review Board #1**
FWA00000351, IRB00001138

To: **Benjamin Sawyer**

Date: **March 10, 2015**

Dear Researcher:

On 3/10/2015, the IRB approved the following human participant research until 03/09/2016 inclusive:

Type of Review:	UCF Initial Review Submission Form
Project Title:	Attention Layer Testbed
Investigator:	Benjamin Sawyer, MS Industrial Engineering
IRB Number:	SBE-15-11064
Funding Agency:	
Grant Title:	
Research ID:	N/A

The scientific merit of the research was considered during the IRB review. The Continuing Review Application must be submitted 30days prior to the expiration date for studies that were previously expedited, and 60 days prior to the expiration date for research that was previously reviewed at a convened meeting. Do not make changes to the study (i.e., protocol, methodology, consent form, personnel, site, etc.) before obtaining IRB approval. A Modification Form **cannot** be used to extend the approval period of a study. All forms may be completed and submitted online at https://iris.research.ucf.edu .

If continuing review approval is not granted before the expiration date of 03/09/2016, approval of this research expires on that date. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

Use of the approved, stamped consent document(s) is required. The new form supersedes all previous versions, which are now invalid for further use. Only approved investigators (or other approved key study personnel) may solicit consent for research participation. Participants or their representatives must receive a copy of the consent form(s).

All data, including signed consent forms if applicable, must be retained and secured per protocol for a minimum of five years (six if HIPAA applies) past the completion of this research. Any links to the identification of participants should be maintained and secured per protocol. Additional requirements may be imposed by your funding agency, your department, or other entities. Access to data is limited to authorized individuals listed as key study personnel.

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

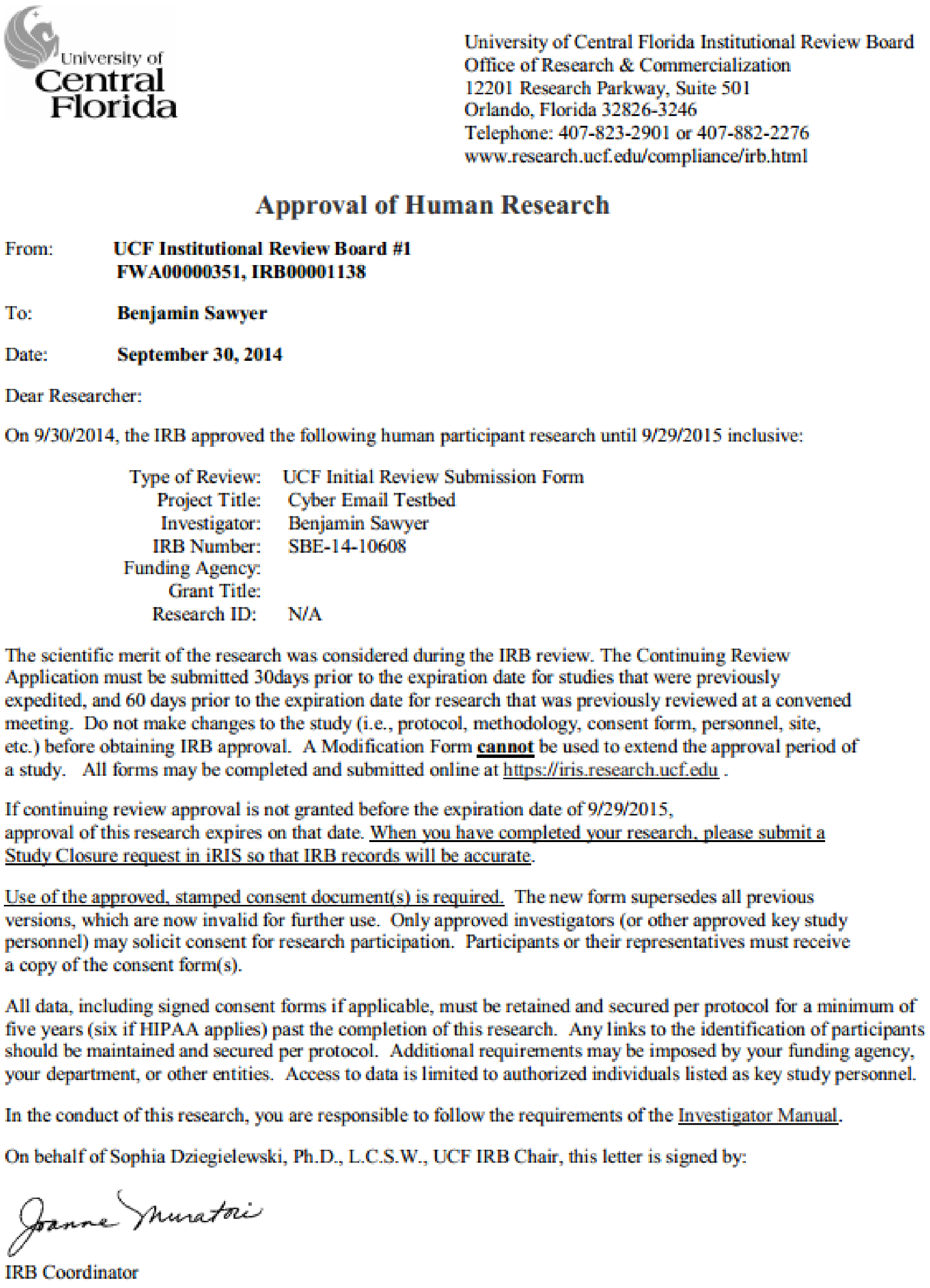
On behalf of Sophia Dziegielewski, Ph.D., L.C.S.W., UCF IRB Chair, this letter is signed by:

Joanne Muratori

Signature applied by Joanne Muratori on 03/10/2015 03:02:13 PM EDT

IRB Coordinator

Experiment IV



University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Human Research

From: **UCF Institutional Review Board #1**
FWA00000351, IRB00001138

To: **Benjamin Sawyer**

Date: **September 30, 2014**

Dear Researcher:

On 9/30/2014, the IRB approved the following human participant research until 9/29/2015 inclusive:

Type of Review:	UCF Initial Review Submission Form
Project Title:	Cyber Email Testbed
Investigator:	Benjamin Sawyer
IRB Number:	SBE-14-10608
Funding Agency:	
Grant Title:	
Research ID:	N/A

The scientific merit of the research was considered during the IRB review. The Continuing Review Application must be submitted 30days prior to the expiration date for studies that were previously expedited, and 60 days prior to the expiration date for research that was previously reviewed at a convened meeting. Do not make changes to the study (i.e., protocol, methodology, consent form, personnel, site, etc.) before obtaining IRB approval. A Modification Form **cannot** be used to extend the approval period of a study. All forms may be completed and submitted online at https://iris.research.ucf.edu .

If continuing review approval is not granted before the expiration date of 9/29/2015, approval of this research expires on that date. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

Use of the approved, stamped consent document(s) is required. The new form supersedes all previous versions, which are now invalid for further use. Only approved investigators (or other approved key study personnel) may solicit consent for research participation. Participants or their representatives must receive a copy of the consent form(s).

All data, including signed consent forms if applicable, must be retained and secured per protocol for a minimum of five years (six if HIPAA applies) past the completion of this research. Any links to the identification of participants should be maintained and secured per protocol. Additional requirements may be imposed by your funding agency, your department, or other entities. Access to data is limited to authorized individuals listed as key study personnel.

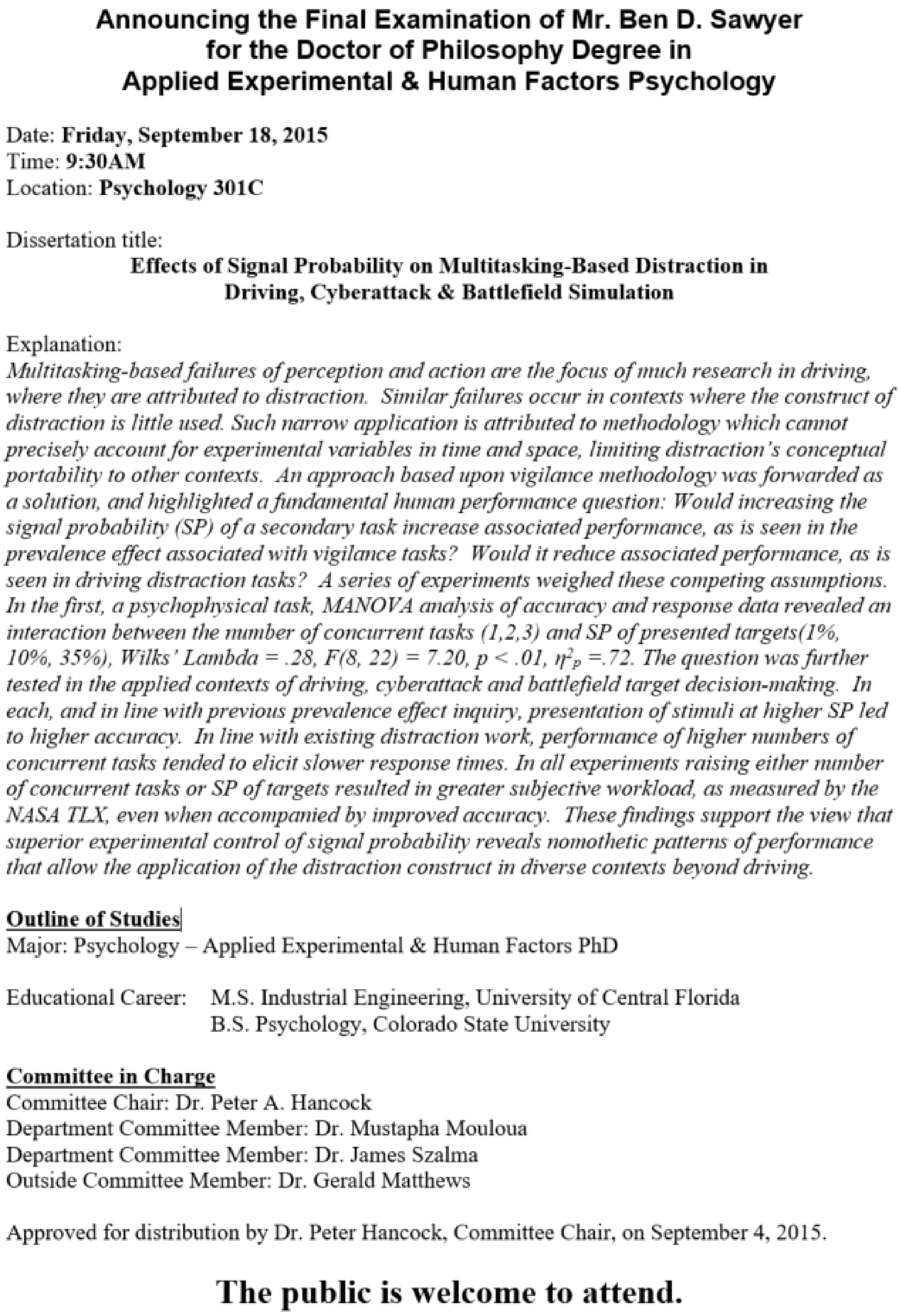
In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

On behalf of Sophia Dziegielewski, Ph.D., L.C.S.W., UCF IRB Chair, this letter is signed by:

Joanne Muratori

IRB Coordinator

APPENDIX B: DEFENSE ANNOUNCEMENT



Announcing the Final Examination of Mr. Ben D. Sawyer for the Doctor of Philosophy Degree in Applied Experimental & Human Factors Psychology

Date: **Friday, September 18, 2015**
Time: **9:30AM**
Location: **Psychology 301C**

Dissertation title:

Effects of Signal Probability on Multitasking-Based Distraction in Driving, Cyberattack & Battlefield Simulation

Explanation:

Multitasking-based failures of perception and action are the focus of much research in driving, where they are attributed to distraction. Similar failures occur in contexts where the construct of distraction is little used. Such narrow application is attributed to methodology which cannot precisely account for experimental variables in time and space, limiting distraction's conceptual portability to other contexts. An approach based upon vigilance methodology was forwarded as a solution, and highlighted a fundamental human performance question: Would increasing the signal probability (SP) of a secondary task increase associated performance, as is seen in the prevalence effect associated with vigilance tasks? Would it reduce associated performance, as is seen in driving distraction tasks? A series of experiments weighed these competing assumptions. In the first, a psychophysical task, MANOVA analysis of accuracy and response data revealed an interaction between the number of concurrent tasks (1,2,3) and SP of presented targets(1%, 10%, 35%), Wilks' Lambda = .28, F(8, 22) = 7.20, p < .01, η^2_p *=.72. The question was further tested in the applied contexts of driving, cyberattack and battlefield target decision-making. In each, and in line with previous prevalence effect inquiry, presentation of stimuli at higher SP led to higher accuracy. In line with existing distraction work, performance of higher numbers of concurrent tasks tended to elicit slower response times. In all experiments raising either number of concurrent tasks or SP of targets resulted in greater subjective workload, as measured by the NASA TLX, even when accompanied by improved accuracy. These findings support the view that superior experimental control of signal probability reveals nomothetic patterns of performance that allow the application of the distraction construct in diverse contexts beyond driving.*

Outline of Studies
Major: Psychology – Applied Experimental & Human Factors PhD

Educational Career: M.S. Industrial Engineering, University of Central Florida
B.S. Psychology, Colorado State University

Committee in Charge
Committee Chair: Dr. Peter A. Hancock
Department Committee Member: Dr. Mustapha Mouloua
Department Committee Member: Dr. James Szalma
Outside Committee Member: Dr. Gerald Matthews

Approved for distribution by Dr. Peter Hancock, Committee Chair, on September 4, 2015.

The public is welcome to attend.

APPENDIX C: PAIRWISE COMPARISONS

Cohen's *d* for all pairwise comparisons is here provided (Cohen, 1988 pp. 276-280).

Table 11: Pairwise Comparisons for Experiment I: Accuracy & Response

DV			Sig. *	*d*
By Task #	vs.			
Accuracy (%)	Single	Dual	0.173	0.188
	Single	Triple	0.084	0.317
	Dual	Triple	0.000*	0.512
Response Time (ms)	Single	Dual	0.357	0.130
	Single	Triple	0.993	0.001
	Dual	Triple	0.245	0.163
By Signal Probability	vs.			
Accuracy (%)	1%	10%	0.001*	0.413
	1%	35%	0.000*	1.335
	10%	35%	0.000*	1.088
Response Time (ms)	1%	10%	0.406	0.593
	1%	35%	0.003*	0.829
	10%	35%	0.011*	1.548

Note. * indicates significance at the .05 level

Table 12: Pairwise Comparisons for Experiment I: NASA TLX

DV			Sig. *	*d*
	vs.			
TLX	Single	Dual	0.000*	0.504
	Single	Triple	0.000*	0.610
	Dual	Triple	0.208	0.128
TLX Mental	Single	Dual	0.000*	0.739
	Single	Triple	0.000*	0.787
	Dual	Triple	0.406	0.066
TLX Physical	Single	Dual	0.003*	0.459
	Single	Triple	0.011*	0.420
	Dual	Triple	0.973	0.003
TLX Temporal	Single	Dual	0.081	0.249
	Single	Triple	0.073	0.339
	Dual	Triple	0.570	0.109
TLX Performance	Single	Dual	0.000*	0.708
	Single	Triple	0.002*	0.597
	Dual	Triple	0.571	0.081
TLX Effort	Single	Dual	0.000*	0.519
	Single	Triple	0.000*	0.673
	Dual	Triple	0.138	0.186
TLX Frustration	Single	Dual	0.019*	0.341
	Single	Triple	0.007*	0.406
	Dual	Triple	0.456	0.070

Note. * indicates significance at the .05 level

Table 13: Pairwise Comparisons for Experiment II: Accuracy & Response

DV	FWSP		Sig. *	*d*
	vs.			
Collision	1%	2%	0.160	0.344
(%)	1%	8%	0.013*	0.548
	2%	8%	0.193	0.198
HRT	1%	2%	0.013*	1.122
(ms)	1%	8%	0.005*	1.324
	2%	8%	0.555	0.101

Note. * indicates significance at the .05 level

Table 14: Pairwise Comparisons for Experiment II: NASA TLX

DV	FWSP		Sig. *	*d*
	vs.			
TLX	1%	2%	0.546	0.393
	1%	8%	0.019*	1.294
	2%	8%	0.060	0.882
TLX Mental	1%	2%	0.646	0.056
	1%	8%	0.010*	1.480
	2%	8%	0.002*	1.661
TLX Physical	1%	2%	0.967	0.129
	1%	8%	0.294	0.545
	2%	8%	0.293	0.501
TLX Temporal	1%	2%	0.527	0.340
	1%	8%	0.041*	0.885
	2%	8%	0.125	0.670
TLX Performance	1%	2%	0.537	0.304
	1%	8%	0.767	0.203
	2%	8%	0.754	0.093
TLX Effort	1%	2%	0.289	0.530
	1%	8%	0.073	1.002
	2%	8%	0.401	0.445
TLX Frustration	1%	2%	0.418	0.451
	1%	8%	0.478	0.455
	2%	8%	0.940	0.058

Note. * indicates significance at the .05 level

Table 15: Pairwise Comparisons for Experiment III: Accuracy & Response

DV	Level		Sig. *	*d*
	vs.			
Accuracy (%)	1T01%	3T03%	0.002*	0.947
	1T01%	3T21%	0.077	0.600
	3T03%	3T21%	0.169	0.412
Response Time (ms)	1T01%	3T03%	0.006*	0.837
	1T01%	3T21%	0.046*	0.656
	3T03%	3T21%	0.891	0.038

Note. * indicates significance at the .05 level

Table 16: Pairwise Comparisons for Experiment III: NASA TLX

DV	Level		Sig. *	*d*
		vs.		
TLX	1T01%	3T03%	0.009*	0.737
	1T01%	3T21%	0.000*	1.259
	3T03%	3T21%	0.065	0.420
TLX Mental	1T01%	3T03%	0.003*	0.980
	1T01%	3T21%	0.000*	1.489
	3T03%	3T21%	0.058	0.446
TLX Physical	1T01%	3T03%	0.482	0.121
	1T01%	3T21%	0.049*	0.459
	3T03%	3T21%	0.065	0.391
TLX Temporal	1T01%	3T03%	0.004*	0.562
	1T01%	3T21%	0.001*	0.723
	3T03%	3T21%	0.298	0.150
TLX Performance	1T01%	3T03%	0.000*	1.766
	1T01%	3T21%	0.000*	1.893
	3T03%	3T21%	0.846	0.051
TLX Effort	1T01%	3T03%	0.004*	0.930
	1T01%	3T21%	0.000*	1.412
	3T03%	3T21%	0.092	0.318
TLX Frustration	1T01%	3T03%	0.001*	1.342
	1T01%	3T21%	0.000*	1.477
	3T03%	3T21%	0.509	0.157

Note. * indicates significance at the .05 level

Table 17: Pairwise Comparisons for Experiment IV: Accuracy & Response

DV			Sig. *	*d*
By Task #	vs.			
Accuracy (%)	Single	Dual	0.173	0.188
	Single	Triple	0.084	0.317
	Dual	Triple	0.000*	0.512
Response Time (ms)	Single	Dual	0.357	0.130
	Single	Triple	0.993	0.001
	Dual	Triple	0.245	0.163

Note. * indicates significance at the .05 level

Table 18: Pairwise Comparisons for Experiment IV: NASA TLX

DV	Level	Sig. *	d	
	vs.			
TLX	1%	5%	0.258	0.446
	1%	20%	0.114	0.771
	5%	20%	0.637	0.206
TLX Mental	1%	5%	0.551	0.238
	1%	20%	0.379	0.417
	5%	20%	0.774	0.124
TLX Physical	1%	5%	0.284	0.623
	1%	20%	0.037*	0.914
	5%	20%	0.286	0.390
TLX Temporal	1%	5%	0.298	0.444
	1%	20%	0.892	0.055
	5%	20%	0.363	0.429
TLX Performance	1%	5%	0.696	0.169
	1%	20%	0.602	0.211
	5%	20%	0.895	0.061
TLX Effort	1%	5%	0.329	0.428
	1%	20%	0.070	0.813
	5%	20%	0.382	0.368
TLX Frustration	1%	5%	0.202	0.528
	1%	20%	0.038*	0.996
	5%	20%	0.395	0.363

Note. * indicates significance at the .05 level

REFERENCES

Adams, J. A., & Boulter, L. R. (1964). Spatial and temporal uncertainty as determinants of vigilance behavior. *Journal of Experimental Psychology*, *67*(2), 127–131.

Agnes, M. (2004). *Webster's new world college dictionary*. Indianapolis: Wiley Publications.

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Backer-Grøndahl, A., & Sagberg, F. (2011). Driving and telephoning: Relative accident risk when using hand-held and hands-free mobile phones. *Safety Science*, *49*(2), 324–330.

Beaman, C. P. (2005). Auditory distraction from low-intensity noise: A review of the consequences for learning and workplace environments. *Applied Cognitive Psychology*, *19*(8), 1041-1064.

Beck, D. M., & Lavie, N. (2005). Look here but ignore what you see: Effects of distractors at fixation. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(3), 592.

Betz, W. (1874). Anatomischer nachweis zweier gehirncentra. *Zentralbl Med Wiss*, *12*, 578-580, 595-599.

Block, N. (1980). Troubles with functionalism. *Readings in philosophy of psychology*, *1*, 268-301. Boston: Harvard University Press.

Broadbent, D. E. (1958). *Perception and communication.* New York: Macmillan.

Caird, J. K., Johnston, K. A., Willness, C. R., Asbridge, M., & Steel, P. (2014). A meta-analysis of the effects of texting on driving. *Accident Analysis & Prevention*, *71*, 311–318.

Chisholm, S. L., Caird, J. K., & Lockhart, J. (2008). The effects of practice with MP3 players on driving performance. *Accident Analysis & Prevention*, *40*(2), 704–713.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New Jersey: Academic press.

Cone, J. D. (1986). Idiographic, nomothetic, and related perspectives in behavioral assessment. *Conceptual foundations of behavioral assessment* (pp. 111-128). New York: Guilford.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in quantitative methods for psychology, 1*(1), 42-45.

Deutsch, J. A., & Deutsch, D. (1963). Attention: some theoretical considerations. *Psychological Review, 70*(1), 80–90.

Dirkin, G. R. (1983). Cognitive tunneling: use of visual information under stress. *Perceptual & Motor Skills, 56*(1), 191–198.

Dirkin, G. R., & Hancock, P. A. (1985). An attentional view of narrowing: The effect of noise and signal bias on discrimination in the peripheral visual field. In: I.D. Brown, R. Goldsmith, K. Coombes, & M. A. Sinclair (Eds.), *Ergonomics International 85: Proceedings of the Ninth Congress of the International Ergonomics Association* (pp. 751-753). Bournemouth, England.

Dressel, J., & Atchley, P. (2008). Cellular phone use while driving: A methodological checklist for investigating dual-task costs. *Transportation Research Part F: Traffic Psychology and Behaviour, 11*(5), 347-361.

Drews, F. A., Yazdani, H., Godfrey, C. N., Cooper, J. M., & Strayer, D. L. (2009). Text messaging during simulated driving. *Human Factors, 51*(5), 762-770.

Dukic, T., Ahlstrom, C., Patten, C., Kettwich, C., & Kircher, K. (2013). Effects of electronic billboards on driver distraction. *Traffic Injury Prevention, 14*(5), 469–476.

Effect Size: Relationship between partial Eta-squared, Cohen's f, and Cohen's d. (n.d.). Retrieved October 19, 2015, from http://www-01.ibm.com/support/docview.wss?uid=swg21476421

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16*(1), 143-149.

Eriksen, C. W. (1995). The flankers task and response competition: A useful tool for investigating a variety of cognitive problems. *Visual Cognition*, *2*(2-3), 101–118.

Field, A. (2009). *Discovering statistics using SPSS*. London: Sage.

Finomore, V. S., Shaw, T. H., Warm, J. S., Matthews, G., & Boles, D. B. (2013). Viewing the workload of vigilance through the lenses of the NASA-TLX and the MRQ. *Human Factors, 55*(6), 1044-1063.

Fischer, E., & Haines, R. F. (1980). *Cognitive issues in head-up displays (NASA Technical Paper 711)*. Washington, DC: National Aeronautics and Space Administration.

Fitch, G. A., Soccolich, S. A., Guo, F., McClafferty, J., Fang, Y., Olson, R. L., Perez, M. A., Hanowski, R. J., Hankey, J. M., & Dingus, T. A. (2013). *The impact of hand-held and hands-free cell smartphone use on driving performance and safety-critical event risk.* (Report No. DOT HS 811 757). Washington, DC: National Highway Traffic Safety Administration.

Fitts, P. M., & Peterson, J. R. (1964). Information capacity of discrete motor responses. *Journal of Experimental Psychology*, *67*(2), 103–112.

Fitts, P. M., & Posner, M. I. (1967). *Human performance*. Oxford, England: Brooks/Cole.

Fleck, M. S., & Mitroff, S. R. (2007). Rare targets are rarely missed in correctable search. *Psychological Science, 18*(11), 943–947.

Forster, S., & Lavie, N. (2011). Entirely irrelevant distractors can capture and captivate attention. *Psychonomic Bulletin & Review, 18*(6), 1064-1070.

Geer, J. H., & Fuhr, R. (1976). Cognitive factors in sexual arousal: The role of distraction. *Journal of Consulting and Clinical Psychology*, *44*(2), 238–243.

Gibson, E. J. (1969). *Principles of perceptual learning and development*. New York: Appleton-Century-Crofts.

Gibson, J. J. (1979). *The ecological approach to visual perception*, New York: Houghton-Miffin.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika, 24*(2), 95–112.

Guralnick, M. J. (1973). Effects of event rate and signal difficulty on observing responses and detection measures in vigilance. *Journal of Experimental Psychology*, *99*(2), 261–265.

Hallowell, E. M. M., & Ratey, J. J. (2011). *Driven to distraction: Recognizing and coping with attention deficit disorder from childhood through adulthood*. Anchor Books.

Hancock, P. A. (1986). Sustained attention under thermal stress. *Psychological Bulletin, 99*(2), 263-281.

Hancock, P. A. (1997). Hours of boredom, moments of terror, or months of monotony, milliseconds of mayhem. *Proceedings of the Ninth International Symposium on Aviation Psychology, April, Columbus, OH.*

Hancock, P. A. (2013). In search of vigilance: The problem of iatrogenically created psychological phenomena. *American Psychologist, 68*(2), 97-109.

Hancock, P. A. (2014). Automation: How much is too much? *Ergonomics*, *57*(3), 449–454.

Hancock, P. A., & De Ridder, S. N. (2003). Behavioural accident avoidance science: Understanding response in collision incipient conditions. *Ergonomics, 46*(12), 1111-1135.

Hancock, P. A., & Meshkati, N. (1988). *Human mental workload* (pp. 329-382). Amsterdam: Elsevier.

Hancock, P. A., Mouloua, M., & Senders, J. W. (2008). On the philosophical foundations of driving distraction and the distracted driver. In K. L. Young, J. D. Lee, & M. A. Regan (Eds.), *Driver distraction theory: Effects and mitigation* (pp. 11–30). Boca Raton, FL: CRC Press.

Hancock, P. A., Sawyer, B. D., & Stafford, S. (2015). The effects of display size on performance. *Ergonomics*, *58*(3), 337-354.

Hancock, P. A., & Szalma, J. L. (2003). The future of neuroergonomics. *Theoretical Issues in Ergonomics Science*, *4*(1-2), 238–249.

Hancock, P. A., & Warm, J. S. (1989). A dynamic model of stress and sustained attention. *Human Factors, 31*(5), 519-537.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human Mental Workload* (pp. 139-183). Amsterdam: Elsevier.

He, J., McCarley, J. S., & Kramer, A. F. (2013). Lane keeping under cognitive distraction: Performance changes and mechanisms. *Human Factors*, *56*(2), 414-26

Helton, W. S., Weil, L., Middlemiss, A., & Sawers, A. (2010). Global interference and spatial uncertainty in the Sustained Attention to Response Task (SART). *Consciousness & Cognition*, *19*(1), 77–85.

Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, *4*(1), 11–26.

Hitchcock, E. M., Dember, W. N., Warm, J. S., Moroney, B. W., & See, J. E. (1999). Effects of cueing and knowledge of results on workload and boredom in sustained attention. *Human Factors, 41*(3), 365–372.

Hitchcock, E. M., Warm, J. S., Matthews, G., Dember, W. N., Shear, P. K., Tripp, L. D., … Parasuraman, R. (2003). Automation cueing modulates cerebral blood flow and vigilance in a simulated air traffic control task. *Theoretical Issues in Ergonomics Science*, *4*(1-2), 89–112.

Hyman, R. (1953). Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology, 45*(3), 188-196.

Ibrahim, J. K., Anderson, E. D., Burris, S. C., & Wagenaar, A. C. (2011). State laws restricting driver use of mobile communications devices: Distracted-driving provisions, 1992–2010. *American Journal of Preventive Medicine*, *40*(6), 659–665.

Jerome, C. J., Ganey, H. N., Mouloua, M., & Hancock, P. A. (2002). Driver workload response to in-vehicle device operations. *International Journal of Occupational Safety and Ergonomics*, *8*(4), 539–548.

Kircher, K., Ahlström, C., Gregersen, N. P., & Patten, C. (2013). Why Sweden should not do as everybody else does. In *3rd International Conference on Driver Distraction and Inattention. September 4-6, 2013, Gothenburg, Sweden*.

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(3), 451-468.

Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in Cognitive Sciences*, *9*(2), 75–82.

Lavie, N. (2010). Attention, distraction, and cognitive control under load. *Current Directions in Psychological Science*, *19*(3), 143–148.

Luck, S. J. (2014). *An introduction to the event-related potential technique*. Boston: MIT press.

Mackworth, J. (1969). *Vigilance and habituation: A neuropsychological approach*. Penguin Books, 1970.

Mackworth, J. (1970). *Vigilance and attention: A signal detection approach*. Harmondsworth, England: Penguin Books.

Mackworth, N. H. (1948). The breakdown of vigilance during prolonged visual search. *Quarterly Journal of Experimental Psychology*, *1*(1), 6-21.

Mackworth, N. H. (1950). Researches on the measurement of human performance. *Med. Res Council Spec Rep Ser*, (268).

Matthews, G., & Desmond, P. A. (2002). Task-induced fatigue states and simulated driving performance. *The Quarterly Journal of Experimental Psychology Section A*, *55*(2), 659–686.

Maybury, M. T. (2012). *Cyber Vision 2025: United States Air Force Cyberspace Science and Technology Vision 2012-2025.* SAF/PA Public Release Case No. 2012-0439/460/715.

McCallum, M. C., Campbell, J. L., Richman, J. B., Brown, J. L., & Wiese, E. (2004). Speech recognition and in-vehicle telematics devices: Potential reductions in driver distraction. *International Journal of Speech Technology*, *7*(1), 25–33.

McCaul, K. D., & Malott, J. M. (1984). Distraction and coping with pain. *Psychological Bulletin*, *95*(3), 516–533.

Mitroff, S. R., & Biggs, A. T. (2014). The ultra-rare-item effect visual search for exceedingly rare items is highly susceptible to error. *Psychological Science, 25*(1), 284-289.

Miyake, S., Hancock, P. A., & Manning, C. M. (1992). Effects of prior task load level on subsequent workload and performance. *Presented at the Annual Human Factors Society Conference.*

Moore, S. F., & Gross, S. J. (1973). Influence of critical signal regularity, stimulus event matrix, and cognitive style on vigilance performance. *Journal of Experimental Psychology, 99*(1), 137-139.

Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology*, *11*(1), 56–60.

Moray, N. (Ed.). (1979). *Mental workload: Its theory and measurement (Vol. 8)*. New York: Plenum Press.

Morgan, J. F., & Hancock, P. A. (2011). The effect of prior task loading on mental workload an example of hysteresis in driving. *Human Factors, 53*(1), 75-86.

Mouloua, M., Hancock, P. A., Rinalducci, E., & Brill, C. (2003). Effects of radio tuning on driving performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *47*(8), 1044–1047.

Natsoulas, T. (1967). What are perceptual reports about?. *Psychological bulletin, 67*(4), 249.

Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology.* WH Freeman/Times Books/Henry Holt & Co.

Owens, J. M., McLaughlin, S. B., & Sudweeks, J. (2011). Driver performance while text messaging using handheld and in-vehicle systems. *Accident Analysis & Prevention, 43*(3), 939–947.

Parasuraman, R. (1979). Memory load and event rate control sensitivity decrements in sustained attention. *Science, 205*(4409), 924-927.

Parasuraman, R., & Nestor, P.G. (1991). Attention and driving skills in aging and Alzheimer's disease. *Human Factors, 33*(5), 539–557.

Paulsen, C., McDuffie, E., Newhouse, W., & Toth, P. (2012). NICE: Creating a Cybersecurity Workforce and Aware Public. *IEEE Security & Privacy, 10*(3), 76–79.

Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics, 2,* 10.

Qualtrics (2013) Version 37892. Provo, UT.

Roberts, D. C., Snarski, S., Sherrill, T., Menozzi, A., Clipp, B., & Russler, P. (2012). Soldier-worn augmented reality system for tactical icon visualization (Vol. 8383, pp. 838305–838305–12).

Room, A., Buchanan-Brown, J., & Pratchett, T. (2000). *Brewer's Dictionary of Phrase and Fable* (16th Revised edition). New York: HarperCollins.

Sawyer, B. (2010*). Impact of components of text messaging on simulated driving performance*. (Honors Thesis).

Sawyer, B. D. (2014). *Applied error related negativity: single electrode electroencephalography in complex visual stimuli.* (Masters thesis).

Sawyer, B. D., Calvo, A., Finomore, V. S., & Hancock, P. A. (2015a). Serendipity in simulation: Building environmentally valid driving distraction evaluations of Google Glass and an Android smartphone. *Paper accepted for presentation at the 19th Triennial Congress of the International Ergonomics Association, Melbourne, Australia. August*.

Sawyer, B. D., Finomore, V. S., Calvo, A. A., & Hancock, P. A. (2014a). Google glass: A driver distraction cause or cure? *Human Factors, 56*(7), 1307-1321.

Sawyer, B. D., Finomore, V. S., Funke, G., Mancuso, V. F., Miller, B., Warm, J. S., & Hancock, P. A. (2015b). Evaluating cybersecurity vulnerabilities with the email testbed: Effects of training. *Paper accepted for presentation at the 19th Triennial Congress of the International Ergonomics Association, Melbourne, Australia. August*.

Sawyer, B. D., Finomore, V. S., Funke, G., & Warm, J. S. (2014b). Cyber vigilance: effects of signal probability and event rate. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, *58*(1), 1771-1775.

Sawyer, B. D., & Hancock, P. A. (2012). Assisted entry mitigates text messaging based driving detriment. *Work, 41*, 4279–4282.

Sawyer, B. D., & Hancock, P. A. (2013). Performance degradation due to automation in texting while driving. *Proceedings of the 7th International Driving Symposium on Human Factors in Driving Assessment, Training and Vehicle Design, No. 68, 446-452. Bolten, NY.*

Sawyer, B. D., & Hancock, P. A. (2014). An evaluation of drivers using an ignition interlock breath test while driving. *Proceedings of the 2014 Human Factors and Ergonomics Society Annual Meeting.* , *58*(1), 2098-2101.

Sawyer, J. M. (1985). *The effect of the progressive relaxation/concentration unit on reading speed and reading comprehension of high anxious college students.* (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses A&I; ProQuest Dissertations & Theses Global. (8516787).

Schmidt, E. A., Kincses, W.E., Schrauf, M., Haufe, S., Schubert, R., & Curio, G. (2007). Assessing drivers' vigilance state during monotonous driving. *Presented at the Driving Assessment, Stevenson, Washington, USA.*

Shannon, C. E. (1948). A mathematical theory of communication. The *Bell System Technical Journal*, *27*(3), 379–423.

Shaw, T. H., Warm, J. S., Finomore, V., Tripp, L., Matthews, G., Weiler, E., & Parasuraman, R. (2009). Effects of sensory modality on cerebral blood flow velocity during vigilance. *Neuroscience Letters*, *461*(3), 207–211.

Simon, H. A. (1969). *The sciences of the artificial*. Boston, MA: MIT Press.

Simons, D. J., & Ambinder, M. S. (2005). Change blindness theory and consequences. *Current Directions in Psychological Science*, *14*(1), 44–48.

Skinner, B. F. (1977). Why I am not a cognitive psychologist. *Behaviorism*, *5*(2), 1–10.

Skulmowski, A., Bunge, A., Kaspar, K., & Pipa, G. (2014). Forced-choice decision-making in modified trolley dilemma situations: a virtual reality and eye tracking study. *Frontiers in Behavioral Neuroscience*, *8*, 426.

Smith, K., & Hancock, P. A. (1995). Situation awareness is adaptive, externally directed consciousness. *Human Factors*, *37*(1), 137–148.

Sternberg, S. (1969). Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, *57*(4), 421–457.

Strayer, D. L., Cooper, J. M., Turrill, J., Coleman, J., Medeiros-Ward, N., & Biondi, F. (2013). *Measuring cognitive distraction in the automobile*. Washington, DC: AAA Foundation for Traffic Safety.

Strayer, D. L., Drews, F. A., & Johnston, W. A. (2003). Cell phone-induced failures of visual attention during simulated driving. *Journal of Experimental Psychology: Applied*, *9*(1), 23–32.

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

Strayer, D. L., Turrill, J., Coleman, J. R., Ortiz, E. V., & Cooper, J. M. (2014). *Measuring cognitive distraction in the automobile II: Assessing in-vehicle voice-based interactive technologies*. Washington, DC: AAA Foundation for Traffic Safety.

Szalma, J., Schmidt, T., Teo, G., & Hancock, P. A. (2014). Vigilance on the move: Video game-based measurement of sustained attention. *Ergonomics*, *57*(9), 1315–1336.

Thomas, L. C., & Wickens, C. D. (2001). Visual displays and cognitive tunneling: Frames of reference effects on spatial judgments and change detection. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *45*(4), 336–340.

Tippey, K. G., Sivaraj, E., Ardoin, W. J., Roady, T., & Ferris, T. K. (2014). Texting while driving using Google Glass Investigating the combined effect of heads-up display and hands-free input on driving safety and performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*(1), 2023-2027.

Törnros, J., & Bolling, A. (2006). Mobile phone use – effects of conversation on mental workload and driving speed in rural and urban environments. *Transportation Research Part F: Traffic Psychology and Behaviour*, *9*(4), 298–306.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive psychology, 5*(2), 207-232.

Veltman, J. A., & Gaillard, A. W. K. (1996). Physiological indices of workload in a simulated flight task. *Biological Psychology, 42*(3), 323–342.

Von Neumann, J. (1953). A certain zero-sum two-person game equivalent to the optimal assignment problem. In K. J. Arrow (Eds), *Contributions to the theory of games* (p.p 5-12). New Jersey: Princeton University Press.

Warm, J. S., Finomore, V. S., Vidulich, & Funke, M.E. (2015). Vigilance: A perceptual challenge. In R.R. Hoffman, P.A. Hancock, R. Parasuraman, J.L. Szalma, & M. Scerbo (Eds.), *The Cambrige handbook of applied perception research*. New York: Cambridge University Press.

Warm, J., & Jerison, H.J. (1984). The psychophysics of vigilance. In J.S. Warm (Eds.), *Sustained attention in human performance* (pp. 15-60). Chichester, UK: Wiley.

Warm, J. S., Parasuraman, R., & Matthews, G. (2008). Vigilance requires hard mental work and is stressful. *Human Factors, 50*(3), 433–441.

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

Wickens, C. D., & Long, J. (1994). Conformal symbology, attention shifts, and the head-up display. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 38* (1), 6-10.

Wolfe, J. M. (2012). Saved by a log how do humans perform hybrid visual and memory search? *Psychological Science 23*, 698-703.

Wolfe, J. M., Horowitz, T. S., & Kenner, N. M. (2005). Cognitive psychology: Rare items often missed in visual searches. *Nature, 435*(7041), 439-440.

Wolfe, J. M., Horowitz, T. S., Van Wert, M. J., Kenner, N. M., Place, S. S., & Kibbi, N. (2007). Low target prevalence is a stubborn source of errors in visual search tasks. *Journal of Experimental Psychology: General, 136*(4), 623.

XKCD. (2011). How Standards Proliferate. https://xkcd.com/927/

Yeh, M., Wickens, C. D., & Seagull, F. J. (1999). Target cuing in visual search: The effects of conformality and display location on the allocation of visual attention. *Human Factors, 41*(4), 524–542.

Young, K., Lee, J. D., & Regan, M. A. (2008). *Driver distraction: Theory, effects, and mitigation*. FL: CRC Press.

Young, M. S., Mahfoud, J. M., Walker, G. H., Jenkins, D. P., & Stanton, N. A. (2008). Crash dieting: The effects of eating and drinking on driving performance. *Accident Analysis & Prevention, 40*(1), 142–148.