UNDERSTANDING DIVIDED ATTENTION AND ITS RELATION TO AGING AND COGNITION THROUGH UNOBTRUSIVE CONTINUOUS MEASURES

By

James “Zephy” McKanna

A DISSERTATION

Presented to the Department of
Medical Informatics and Clinical Epidemiology
And the Oregon Health & Science University
School of Medicine
in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

April 2013

School of Medicine

Oregon Health & Science University

CERTIFICATE OF APPROVAL

This is to certify that the PhD dissertation of

James McKanna

"*Understanding divided attention and its relation to aging and cognition through unobtrusive continuous measures*"

has been approved

Holly Jimison, PhD

Misha Pavel, PhD

Aaron Cohen, PhD

Jayashree Kalpathy-Cramer, PhD

Joel Nigg, PhD

Table of Contents

List of Figures v

List of Tables vii

Acknowledgements viii

Abstract ix

Chapter 1: Background 1

Introduction: America and the Elder Care Crisis 1

Attention: Utility and Necessity 3

Attention: Current Understanding 4

Divided Attention 6

Current Testing Practices 8

Testing Divided Attention 11

Modeling Divided Attention 12

The Useful Field of View Test 13

Driving and the Elderly 17

Benefits of Continuous Monitoring 19

Computer Games for Cognitive Health 20

Letter Confusability 23

Learning and Fatigue Curves in Visual Search 25

The Psychometric Function 27

Previous Research 30

Chapter 2: Pilot Study 33

Overview 33

Overall Procedure 33

Participant Selection 34

Methods 34

Enhanced Useful Field of View 34

21 Tally Puzzle 37

EEG for UFOV and 21 Tally 40

Results 41

Discussion 44

Improving the enhanced UFOV 45

Dual Attention versus Task Switching 47

Learning and Fatigue: follow-up to the pilot study 48

Chapter 3: Main Study 54

Overview 54

Research Question 54

Specific Aims 54

Methods 55

Participant Selection 55

Administering the experiment 56

Analyzing results 59

Results 63

Accounting for Learning and Exhaustion 63

Separation of divided attention from speed of processing in standard UFOV 64

Modeling of performance 66

Measurement of divided attention skill from Tally performance 76

Qualitative measures and self-report 79

Chapter 4: Discussion 82

Summary of Major Findings 82

Aim 1: Separating divided attention from speed of processing in UFOV 82

Aim 2: Prediction of divided attention from Tally performance 86

Implications 87

Limitations 91

Testing and Time of Day 92

Cognitive versus Perceptual Load 92

Future Directions 93

Conclusion 96

References 97

Appendices 107

Appendix I: Surveys 107

Survey 1: Pilot Study Pre-Experiment Survey 107

Survey 2: Pilot Study Post-UFOV Survey 108

Survey 3: Pilot Study Post-Tally Survey 109

Survey 4: Pilot Study General post-participation questions 110

Survey 5: Main Study Pre-Experiment Survey 111

Survey 6: Main Study Post-UFOV Survey 112

Survey 7: Main Study Post-Tally Survey and general questions 113

Appendix II: Instructional Text 115

Pilot Study UFOV Instructions 115

Main Study UFOV Instructions 119

Tally Puzzle Instructions 123

Appendix III: Screenshots 126

All Enhanced UFOV Focal Point and Luminance Equalization 126

Pilot Study UFOV Screenshots 127

Learning Curve and Main Study UFOV Screenshots 134

Example Tally Screenshots for All Studies 135

Appendix IV: recruitment emails 137

Appendix V: glossary of terms 139

List of Figures

Figure 1. Standard UFOV screenshot. 15

Figure 2. Psychometric function examples with steep and mild slopes 28

Figure 3. Screenshot from the 21 Tally game 31

Figure 4. Screenshots from the enhanced UFOV portion of the pilot study 35

Figure 5. Two screenshots from the Tally Puzzle 38

Figure 6. Central “difficulty” and 75% accuracy duration 42

Figure 7. Peripheral difficulty and 75% accuracy duration 43

Figure 8. Number of trials and 75% accuracy duration 45

Figure 9. Two screenshots from the enhanced UFOV experiment 48

Figure 10. Trial set number and 75% accuracy duration 51

Figure 11. Trial vs threshold duration for hard enhanced UFOV trials 51

Figure 12. Comparison of subjects’ Enhanced UFOV learning curve 52

Figure 13. Comparison of subjects’ Tally learning 53

Figure 14. Age of participants and their standard UFOV scores 60

Figure 15. Comparison of trial sets with normal and non-normal distributions 62

Figure 16. Comparison of participants’ 75% accuracy threshold durations 63

Figure 17. Understanding instructions and the learning curve 65

Figure 18. Example of temporal relationships determining response time 67

Figure 19. Examples of performance produced by a parallel process 68

Figure 20. Examples of performance produced by a serial process 69

Figure 21. UFOV serial process example 70

Figure 22. UFOV parallel process example 71

Figure 23. UFOV results showing no interaction and a large interaction 73

Figure 24. Tally results, showing no interaction and a large interaction.................. 73

Figure 25. Surface visualization of a large interaction.. 75

Figure 26. Surface visualization of a small interaction.. 75

Figure 27. F-statistic of the interaction as calculated by Tally versus UFOV........... 76

Figure 28. F-statistic of interaction as calculated by Tally versus UFOV, detail...... 77

Figure 29. Correlating enhanced UFOV to Tally Puzzle interaction......................... 78

Figure 30. Participants who play FPS games have shorter standard UFOV............ 80

Figure 31. Performing well on 21 Tally does not mean high divided attention......... 81

Figure 32. Standard UFOVs2 and easy enhanced UFOV correlation....................... 83

Figure 33. Standard UFOV subtest 2 and Tally Puzzle correlation........................... 87

List of Tables

Table 1. Example Letter Confusability Matrix.. 24

Table 2. Conditions for the enhanced UFOV. 36

Table 3. Letter Confusability in this Experiment.. 47

Table 4. Conditions for Enhanced UFOV in the Learning Curve Experiment. 49

Table 5. Conditions for Enhanced 21 Tally in the Learning Curve Experiment. 49

Table 6. Conditions for UFOV in the main experiment. 58

Table 7. Conditions for 21 Tally in the main experiment. 58

Table 8. Process model fitting of observed participant behavior. 71

Table 9. Observed threshold values for one subject in the main experiment. 74

Table 11. Correlation of standard UFOV to enhanced UFOV.. 84

Table 12. Correlation of standard UFOV to Tally Puzzle. 88

Acknowledgements

I would like to acknowledge that this work would not have been possible without the following: Holly, Misha, Aaron, Jayashree, Joel, the passage of time, my NLM training fellowship, and caffeine. You have my deepest gratitude.

I owe a huge debt to my family for their thorough and undeserved love and support: Vivien, Jim, and Austy, Liz and Jade, Candy and Bruce, Alison, Dan, and Nathan, Fey, Jana and Genya, and Satya. Each of you has done more to help this process than I can believe. I would have quit long ago if not for you.

Thank you also to my beta testers and proofreaders, particularly Krystal and Felicity who were the first and therefore had it worst. Every iteration afterward owes you. Thank you to Payton and Taggart for their inspiration and limitless philosophical inquiry. Thank you to Vanessa for reminding me how to take time for what I enjoy. Thank you to all those who listened to my practice talks, particularly Steve. Thank you to those who offered good advice: Kathy, Diane, Katrina, and many others. Thank you to Andrea, Diane, and Lynne for keeping me on course for so many years. Thanks also to the professors of DMICE and the rest of OHSU who gave me the ability to ask the right questions. Thank you to Devin and Spry Learning for the opportunity to program fun cognitive games and the continuing support of allowing their use in this research. Thank you to the Cognitive Monitoring group, particularly Daniel, Stuart, and Max, for all their help understanding the computational modeling aspects of this work. Thank you to my participants, particularly those in the pilot study who put up with electrode gel and a complete lack of feedback for a very long procedure.

This could never have been accomplished alone. Thank you all.

Abstract

One of the most important aspects of cognitive health is the ability to control attention; this is also one of the aspects of cognition that typically declines with age. New methods for cognitive health coaching for older adults require new techniques for unobtrusive continuous monitoring: gathering cognitive data from everyday activities performed in the home, rather than forcing elders to leave their natural environment to undergo a battery of tests. The studies discussed herein investigated a method for unobtrusively monitoring divided attention ability through analysis of performance on a cognitive computer game called 21 Tally.

These studies demonstrate that divided attention ability can be estimated through the statistical analysis of the interactions between difficulty levels in a dual-task test. This method is used to enhance one standard dual-task cognitive test of divided attention, the Useful Field of View. The enhanced Useful Field of View (eUFOV) can predict performance on the standard test as accurately as repeated measures of that test can; eUFOV also allows measurement of divided attention without including processing speed, through analysis of the interaction between difficulty levels of the two tasks. Comparing participants' interactions between the different difficulty levels on 21 Tally to those observed in eUFOV performance, we found a correlation of 0.98 between the two ways of measuring divided attention.

This work helps to satisfy two important goals in cognitive testing for older adults. First, it presents a potentially more useful test of divided attention than the current standard, allowing the scientific community to investigate this cognitive skill without including confounding speed of processing measurement. Second, it helps to validate the use of a computer game to monitor divided attention over time, with applications for early detection of cognitive problems as well as monitoring and evaluation of interventions designed to improve divided attention capabilities.

Chapter 1: Background

Introduction: America and the Elder Care Crisis

One of the largest problems facing the country is that we, as a nation, are growing older. This is partially due to the "Baby Boomer" generation, an estimated 76 million Americans who have just passed retirement age(1) and are now struggling with decisions about when, or if, they should enter an assisted living facility. This decision has huge implications for their independence and quality of life – not to mention their likelihood of becoming depressed or even suicidal.

The danger to overall quality of life among elders is, ironically, exacerbated by our modern medical practices. For the past several decades, medical science has been concerned with extending life without taking the quality or independence of that life into account; as might be expected under such circumstances, the average quality of Americans' final years has decreased dramatically. Fries suggests(2) that the solution to this problem lies in the idea of "rectangularization": that is, keeping the average length of life the same, but turning our focus towards keeping the quality of life as high as possible throughout. We Americans have difficulty admitting that we will ever die; once that hurdle is overcome, however, Fries' idea of the perfect life – a high quality only dropping off very sharply at the end – holds great appeal.

One of the keys to a high quality of life among American elders is independence. Not only is independence considered an ideal in our culture, but the simple disruption in routine, for people who have spent decades living in their own homes and dealing with their own affairs, can be devastating. Elders who are able to stay in their own homes have been shown to remain at a higher QoL than their counterparts in assisted living facilities.(3) From a societal standpoint, the resources consumed by someone in a facility with permanent medical and nursing staff are vastly greater than those used by someone living at home.

However, it's unusual for elders in this country to be able to remain in their homes through the end of their lives, because the onset of cognitive decline can cause independent living to be dangerous. Normal activities like cooking, driving, etc., can be difficult or dangerous for a cognitively impaired person, and the adult children (or other caregivers) often feel compelled to preemptively move elders to assisted living facilities for fear of the onset of impairment.

The "gold standard" way to determine whether someone is experiencing cognitive decline is through neuropsychological testing. However, there are several drawbacks to this approach. First, the person is forced to enter an unfamiliar environment (the psychologist's office) in order to take these tests, and can experience all sorts of abnormal symptoms simply due to their surroundings (i.e., White Coat Syndrome).(4;5) Second, one major symptom of cognitive impairment is increased variability in functioning; a person may be perfectly aware on "good days" and completely impaired on "bad days." (6-8) There's simply no way to know, in a traditional office visit, whether the given day is representative. Nor can the tests be taken more frequently – most are required to have at least six months between testing, because the tests themselves can be learned.(9) Further, the results of standard neuropsychological tests are compared to national standards rather than to a person's own baseline ability. Thus, a highly intelligent person won't be diagnosed until they've declined much further than a less intelligent one. Finally, these tests are expensive, in terms of both time and money. This is an important factor when considering the enormous number of people approaching potential cognitive decline in the near future in this country.

One excellent potential solution to all of these issues is in-home monitoring. This practice involves taking data from the normal daily activities of an elder (e.g., walking speed around the home, computer usage, etc.) in order to predict whether the individual is experiencing cognitive decline.(10-14) This obviously keeps people in their usual environment, as well as allowing them to be compared to their own baseline of activity. Additionally, they can be compared across different amounts of time (months, days, or even hours) to allow their own variability to be assessed as part of the diagnosis. This gives peace of mind to caregivers, which in turn allows elders to continue to age in place, keeping their quality of life high and fulfilling the goals of rectangularization.

This study is an attempt to enhance the practicality of in-home monitoring through investigation of divided attention via computer game metrics. In the following pages I hope to show that divided attention is among the most important of cognitive skills with regard to independence among elders, and that certain types of computer games are becoming more and more a part of the daily activities of America’s aging population. Thus, the ability to discern changes in divided attention through metrics derived from computer game use would be a huge step forward in our ability to care for elders in their homes, rather than forcing them to move to places where more expensive care must be provided.

Attention: Utility and Necessity

Attention is one of the key cognitive abilities that allows us to interact successfully with our environment. We have a huge amount of sensory input bombarding us virtually all the time. Think about it: how many objects, people, etc. can you see from your current location? How many photons are bringing you information about each object every second? Visual cells are being excited into firing by each of those photons, and some structure or process must deal with each one of those firings. Cognition could quickly be overwhelmed if each tiny bit of information about the environment actually reached your conscious mind.

For a good illustration of the utility of attention, imagine that we wanted to build a robot to behave like a human. We could, theoretically, attach a separate processor to every single one of the visual-cell-analogue sensors, allowing the robot to examine every single photon detected, and react to every visual change in the environment. However, in order to react *appropriately* – in this instance, like a human being – at some point all of that information, every note about how the current photon detected by sensor XYZ is different from the last one, must be combined into an understanding of the whole. Without some method of filtering the data for relevance, highlighting information likely to help determine the robot's next action and removing the rest, the search space for determining what to do becomes enormous, making real-time reaction impossible even for our theoretical vast system of parallel processors. The reaction to such a large feature space would be something akin to autism: an inability to respond appropriately to many situations, exacerbated by increases in the number of simultaneous stimuli.

The only way the problem becomes remotely tractable is if we throw out a huge amount of data – and that is precisely what neural mechanisms controlling attention have evolved to do.(15) The moment we start classifying input, even in the most basic "important" / "not important" way, we need attention. This obviously extends into the other cognitive domains of learning (noting which input is relevant to a certain situation, which data translate to other similar situations and which don't, etc.), memory (storage space is limited, so we need to give preference to the patterns that are most relevant to survival, or those which occur most often), planning (placing attended information together to accurately predict outcomes), to name but a few.

The importance of attention to the elderly is even greater than it is to the general population, as elderly are often dealing with novel situations in which their previously-formed and –useful response patterns no longer apply. For example, a gentleman who has had a telephone connected via wire in his kitchen for most of his life is not experientially primed to attend to the ringing of his cell phone when he is out for a walk. In the present era of technological advancement, a person might experience several such dramatic shifts in the "normal" way of doing things throughout a lifetime; responding appropriately to these requires a cognitively healthy attention. Unfortunately, attention-affecting disorders rise in prevalence as the population gets older.(16-18)

Attention: Current Understanding

Scientists have been concerned with the study of attention for over a century. As William James put it in 1890, "Everyone knows what attention is. 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… It implies withdrawal from some things in order to deal effectively with others…"(19) Attention can be thought of as focus on a particular set of stimuli and excluding other stimuli from that focus. It can be extremely short-lived or it can be extended over a period as long as many hours (though this becomes more difficult unless the stimuli become progressively more engaging or evolutionarily relevant).

Unlike most other psychological phenomena, attention involves several behavioral dichotomies, suggesting that it stems from at least two (and possibly several) distinct internal processes.(15) For example, attention has a conscious component, in that we can choose to focus on certain things and ignore others. However, it also has several less-conscious components: some involve an inability to attend to things we would like to (like long-term focus), while others involve our inability to ignore things we might like to (like unexpected high-contrast or high-intensity stimuli). This distinction is often termed top-down (that is, attention generated from the conscious mind and directed toward some goal) versus bottom-up (attention generated by environmental stimuli, presumably with the implicit evolutionary goal of individual survival).(19;20)

Another dichotomy in our understanding of attention is that attention can either be perceptual (i.e., external) or cognitive (i.e., internal). That is, attention can either be focused through our senses, our external perception of the world, or it can be focused inwardly, involving only our cognition. For example, one could direct a movable organ designed to sense the external environment (eyes, skin, etc) toward a stimulus as part of directing our attention toward it. Alternatively, one might simply shift "mental focus" toward a stimulus, changing attention with (potentially) no externally discernible change in our physical state – though it should be noted that in almost all cases some measurable change in physiological state occurs, even with a purely cognitive attention shift.(21;22) Adding even more complexity to the mix, we can even direct our attention toward stimuli that aren't physically present, as when remembering or imagining.

Any combination of these disparate aspects of attention can occur – we could be startled by a memory which consumes our attention for a few moments, or we could focus our eyes upon an external stimulus. However, none of these definitions really assist us in building a model of attention that encompasses all that we know about the neural underpinnings and the externally observable behaviors. For the purposes of this paper, attention can be defined as a set of mechanisms that control and allocate cognitive and perceptual processing resources to support goal-directed behaviors in the face of competing tasks and distractions.(23)

In humans, vision is the most relied-upon sense for perceiving the world; as such, vision is closely tied with attention, both in terms of the way people typically direct their attention in a top-down method, and in the number of stimuli that can elicit a bottom-up, non-consciously-directed attention shift.

Divided Attention

Most research on attention deals with focused, or selective, attention, which can be thought of as the ability to focus on a single stimulus, or a single stream of stimuli, at the exclusion of other stimuli. Humans are extremely good at this sort of filtering, particularly in our most highly developed senses of vision and hearing; this filtering can occur far before the stimulus ever reaches the conscious stage.(24) An excellent example of this is the fact that we don't see the blood vessels in our eyes, despite the fact that they constantly block our vision.(25) Since they always block the same visual space, they are not considered part of our relevant visual map of the universe, and are removed before we can consciously perceive them. Other filtering can occur at higher levels of processing – for instance, seeing only one of two possible images in an optical illusion – and, finally, some amount of filtering occurs on the conscious level, as when intentionally attend to only a single conversation in a room full of talking people.

Divided attention describes the ability to attend to multiple tasks, or multiple parts of a task, either simultaneously or by switching back and forth fast enough that the measurable effect is the same. This cognitive skill is highly associated with driving ability, so much so that many stereotypical roadside tests of sobriety in police handbooks, like walking a straight line while counting backwards from 100, are divided attention tests.

Divided attention is often presented as the opposite of selective attention; however, this is a misconception, as both types of attention can and do operate simultaneously. During a divided-attention-intense task, such as driving, we are required to attend to a great many things: the speed of the vehicle, other vehicles in the surrounding roadway, pedestrians and animals on the sidewalk or in the roadway, the volume of the radio, the temperature of the air conditioning, the conversation of our passengers, etc. All of these must be taken into account, each has a different priority – and these priorities shift based on external stimuli like weather, road conditions, and the actions of other drivers and pedestrians. However, drivers are certainly applying selective attention as well, since they must filter out a huge amount of additional information which is generally irrelevant – the color of the fire hydrants they pass, the birds passing overhead, etc. Thus, both sorts of attention are required to perform the normal daily activities of an independent person.

Like most cognitive abilities, divided attention typically increases during childhood, plateaus in the years following puberty, and then gradually decreases with continued aging. The decline in this cognitive skill is seen as one of the primary facets of decreased driving ability in the elderly.(26) When in familiar surroundings and typical conditions, older drivers often have no trouble maneuvering a vehicle. However, good driving is generally measured by drivers' reactions to unusual conditions; otherwise, we would have no issues with robot cars taking us everywhere we need to go. When unexpected events – snow or ice, say, or an animal in the road – occur, people with lower divided attention skill lack the ability to track the motions of other drivers as well as their own speed and direction, which can put them and those around them in danger.

Divided attention is one of the skills which declines earliest at the onset of cognitive decline;(27) this makes it ideal to use as a screening tool for more generalized cognitive decline which might lead to Alzheimer's or other forms of dementia. The loss of divided attention skill is also highly associated with other age-related risks, including falls(28) and the onset of Mild Cognitive Impairment (MCI), often a precursor to Alzheimer's Disease.(29;30) Although there are numerous experimental approaches to the assessment of divided attention, few tools exist to measure it in real life. Accurate measurement of variability in this skill over time, relative to a patient-specific baseline, would allow inference of other forms of cognitive decline, most importantly those leading to Mild Cognitive Impairment or Alzheimer's Disease.

Current Testing Practices

Cognitive assessment is an important aspect of care for elders, as unanticipated cognitive decline can have serious negative consequences for independence. Ideally, cognitive assessment would be incorporated into normal daily activities in a home environment so as to detect trends and variability rather than just a single test outside of the larger context. However, current standard testing practices for MCI and AD still involve a patient traveling to a neurologist or neuropsychologist, being administered a battery of tests, and then receiving the results of these tests. This standard of care may make sense for certain somatic conditions like pneumonia or hypertension; however, they are insufficient and may even be counterproductive when applied to cognitive issues.(31-33)

First, the tests themselves are extremely coarse and their scoring is highly subjective. For example, one of the questions on the Mini-Mental State Exam (MMSE), one of the most common tests for cognitive impairment, is "Where are you right now?" The instruction manual states that when scoring this question, "Any appropriate answer is okay." (33;34) Nor is this an unusual example of a question; by the time patients fail enough of these questions to warrant diagnosis with MCI or AD, friends and family have in many cases known for years that cognitive decline was taking place. When the doctor – or worse yet, the patient – is the last to know about his or her condition, that is a sad state of affairs. This is especially true when knowledge of the onset of the disease could provide so much benefit to the patient, as is the case with cognitive decline. In a worst-case scenario, this knowledge would at least allow a patient to formalize documents and wishes that require "sound mind" – wills, conditions for resuscitation or not, etc. In a more optimistic outlook, this knowledge might allow targeted cognitive exercise like vocabulary or spatial reasoning practice in order to combat a particular area of decline.

The subjectivity inherent in tests that allow "any appropriate answer" is obvious. This is problematic both for improvement – or, heaven forbid, automation – of the test, and for standardization. Better test administrators will pick up on subtle cues of answering these test questions and will be able to give a more nuanced diagnosis than a novice or than a computer could hope to give. That's great if you're being tested by the world's leading neuropsychologist, but terrible if you're getting this test from a less experienced person. In terms of providing good care to all, the subtleties of ideal scoring must be enumerated so that every administrator can follow them. Unfortunately, even the best administrators may not be capable of providing that insight for these tests. These tests have been standardized and validated, and a generation or more of psychologists have been trained on how to administer and score them. Sadly, this does not make them good tests; it merely makes the transition to better tests more difficult.

A second problem with the current testing paradigm is that the environment and other contextual variables may have a large effect upon the symptoms of a cognitive disorder. One of the primary characteristics of cognitive decline is an increased variability in cognitive function (as well as motor control and emotion).(35) This means that a person slipping into MCI may be completely unimpaired on a given day, or on a given task, but may have extreme difficulty on a different day or task. A diagnosis of MCI or AD has extreme negative connotations within our culture, both legally and socially: it is incurable, and may be seen as a death sentence; it involves mental disorder, which is rarely talked about and can be seen as embarrassing for a heretofore independent person; it may be grounds for a charge of legal incompetence, which can allow family members to change a will, force a person out of his or her home and into assisted living, or dictate any number of things about the diagnosed person's life and future.

Thus, while it is unclear exactly what factors contribute to the variability in cognitive function, there is strong incentive for the patient to perform well to whatever extent he or she can. This incentive, combined with the subjective nature of much of the test scoring, calls to mind the old joke about the elderly man who, when asked who the president was at his yearly cognitive checkup, always answered "Oh, that same old asshole," – and was always counted correct. The gentleman described arguably employed an optimization strategy for the given test, and as such was able to bypass its diagnostic abilities completely. While naturally intended for humor, this joke points out two of the fundamental flaws of the current test structure: in many situations it pits the test giver against the test taker in a battle to determine the elder's future. Obviously this is not ideal; an objective measure of cognitive decline should be computationally understandable, given data on the individual baseline and trends of a particular patient.

Another arguably greater flaw in the current paradigm is that the tests are invariate, and can therefore be learned. This was always a problem in that many of the tests had a required "forgetting" period – for example, the Useful Field of View test can only be given once every six months to avoid a learning effect.(36) In terms of following the trend of a worsening condition, this is a terrible limitation. However, in the Information Age, this problem becomes drastically worse. Virtually every test that is administered in a normal cognitive battery is described in detail on easily accessibly websites like Wikipedia. As people familiar with finding and consuming internet-based information grow older, there is nothing to prevent them from finding and practicing tests before they go in to see their doctors. One might argue that a person able to accomplish this multi-step plan in order to foil a cognitive test is, in fact, cognitively healthy; however, it's important not to underestimate the motivational power of a healthy diagnosis, when the alternative is loss of independence. Even a little bit of practice, or reading over the questions ahead of time, would certainly alter the expected results from those that have been so carefully validated in the past – as well as delaying the start of treatment at the earliest stages when it may be most helpful.

Testing Divided Attention

In terms of clinical testing for divided attention, the test used by most neuropsychologists is the Trail Making Test (TMT) B.(37) The TMT consists of a piece of paper with 25 labeled circles, upon which participants must draw a line from one circle to the next in a particular order (connect-the-dots, in other words), until each circle has been reached once. Part A is simply the numbers 1 to 25, and the participant must connect them in order as fast as possible. TMTB is a set-switching task in which participants must draw a line from the circle labeled “1” to the circle labeled “A”, continue on to “2” and then “B”, “3”, “C”, and so forth – again, as fast as possible. If participants make a mistake, the test administrator must correct them before they move on to the next circle. The “score” for the test is the time it took the participant to complete the test; errors are not recorded (though obviously they affect the speed at which the test can be accomplished).(38) A trained neuropsychologist can determine a great deal from watching the way someone performs this test; it allows insight into not just set switching but also processing speed, executive function, planning, short-term and working memory, visual and auditory comprehension, and divided attention. (39) This makes the TMTB a useful and versatile test for neuropsychologists, but it also makes it an extremely subjective one, in which the official “score” tells only a minor part of the story and in which experienced test-givers have an extreme advantage over novices or computer analyses.

Thus, while TMTB does contain “aspects of divided attention, ” (37) it is clearly not specifically a test of divided attention, and in fact the neuropsychologists who use it admit that it’s not the *best* test for that aspect of cognition. Presumably, its popularity is based on other factors, such as the fact that it can be administered on paper rather than on a computer, the cost (free), and the versatile but subjective overall scoring scheme discussed above.

Modeling Divided Attention

Like all cognitive abilities, divided attention exists as part of a system, rather than in a vacuum. Thus, when divided attention declines, many other things are also affected – either as a result of the reduction in divided attention or due to whatever caused the divided attention decline. Thus, correlations can be made between standard tests of divided attention and age, walking speed, motor speed, car crashes, typing speed, depression level, and so forth.(10;40) In order to model divided attention, however, we must incorporate our understanding of the underlying neural mechanisms as well as the observed behaviors.

The simplest model of divided attention as relates to a dual task would be as follows:

$$A = \alpha_1 + \alpha_2 + \varepsilon \qquad (1)$$

… where A represents total attention, α_1 and α_2 are the attention paid to task 1 and task 2, respectively, and ε is some error. However, in order to relate this to observable behavior, we must include some measure of performance on each task as well. Assuming each task is binary (that is, performance is either correct or incorrect, with some time factor attached), we can say several things:

1. The probability of success on each task is related to the attention paid to that task and the time taken to respond: $P_1 \approx \alpha_1 + 1/t$
2. At some very low level of load, the probability of success on task 1 is unrelated to the probability of success on task 2: $P_1 \neq P_2$
3. At some very high level of load, the probability of success on task 1 is completely dependent on the success of task 2, as attention is required for both and the user must choose which to focus upon: $P_1 \approx 1/P_2$

Thus, if we minimize the duration of the stimulus, and discover the probability of success at that duration, we should be able to estimate the attention paid to that task. Estimating the total attention being paid to both tasks is somewhat more difficult, though measurements with EEG may give some insight. Finally, the relationship between the probability of success on task 1 versus task 2 at high load relative to low load should give a measure of divided attention.

The Useful Field of View Test

Many lab tests of divided attention have been developed for both visual and auditory systems, involving such disparate methods as using a keyboard to respond to differences between tones while listening to recorded conversation,(41) tracking an object across a screen with a joystick while responding with a button press to novel visual stimuli,(42) or vocally responding to different tones and letters appearing on a screen.(43) However, none of these designs have been widely used to test real-world behavior relevant to the aging process outside of the laboratory. For this reason, we chose to focus on one test which has been widely endorsed by many researchers and driving organizations, including AAA,(44) as an offroad test of driving ability for older drivers: the Useful Field of View test.(45-47)

The Useful Field of View test (UFOV) was introduced by Ball and Owsley in 1993, and was intended as an easy, off-road, vision-centered test of driving ability among elders. With it they attempted to address the disparity between observed difficulties among elderly "problem" drivers, which seemed chiefly visual (e.g., not obeying road signs, problems when turning, not giving the right of way, etc.), and standard visual tests, which had extremely low correlations with actual road performance.(48) They were successful in this goal, producing a test which has a relatively high correlation ($r > 0.45$) with each of crashes, at-fault crashes, and on-road driving errors. This is at least partially due to the fact that this "vision" test actually tests speed of processing and divided attention, as well as central and peripheral visual acuity.

The standard UFOV consists of four subtasks given in order. In each subtask, the participant is shown a focal point (a dot or cross in the center of the screen), followed by one or more images on the screen for a certain short period of time (between 16 and 500ms), followed by a 500ms screen of “snow” (random black and white pixels) to remove any afterimages the participant may have. The images shown after the focal point are considered the stimuli for each task, and the duration for which these stimuli are presented is the primary metric of the UFOV. These stimuli generally involve cartoon representations of either a truck or a car (see Figure 1 for an example). After the “snow” screen, the participant is asked to give a response based on these stimuli. If the participant’s response is correct, then the next trial’s stimulus duration is shorter; if the response is incorrect, the duration is longer. The durations are shortened and lengthened using a staircase algorithm such that the participant is always approaching the point at which he or she is 75% accurate at the given task. This duration at 75% accuracy is the “score” for that subtask of the UFOV. Typically, 75% accuracy duration scores will increase for each successive subtask (that is, the first subtask is the easiest, requiring the least duration, and the last is the hardest, requiring the most processing time to achieve 75% accuracy). The total UFOV score is the sum of all subtask scores. In the first test, the “speed of processing” subtask, the participant is shown either a car or a truck in the center of the screen, and is asked to identify the image by clicking on the matching image during the response period. This is intended to measure the participant’s base processing speed, and to familiarize him or her with the car and truck images which will be used throughout.

In the second test, the “divided attention” subtask, the participant is shown either a car or a truck in the center of the screen, just as in the first subtask. However, the participant is also shown a car image on the periphery of the screen, at one of the eight cardinal or semicardinal directions. During the response, the participant must first identify what image appeared in the center of the screen, and then must identify where the car image was located (given the eight possible locations) on the periphery of the screen. Because the set of experiments laid forth in this paper deal primarily with divided attention, this is the subtask that is most relevant to our work.

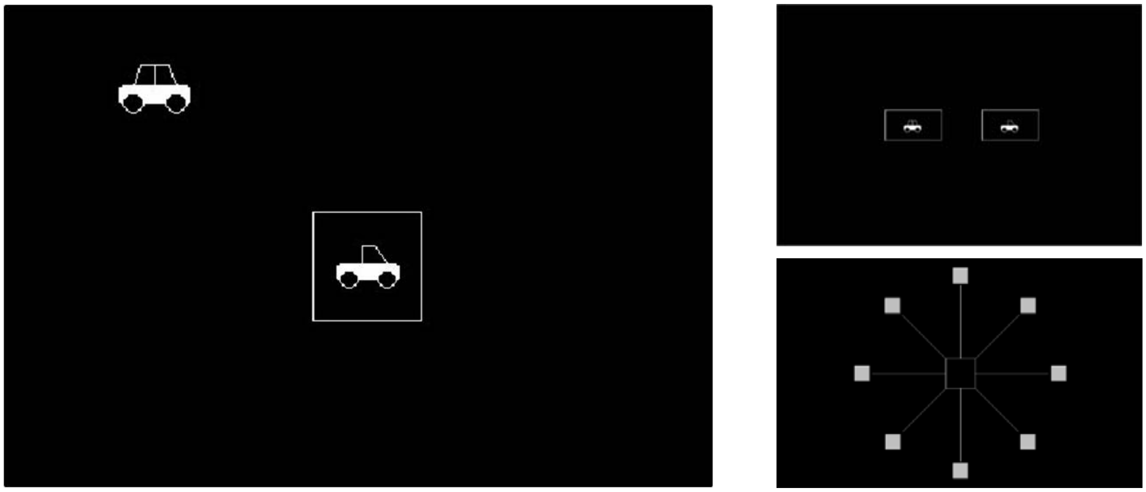


Figure 1. Standard UFOV screenshot. Left: stimulus, which will appear for a duration determined by the staircase algorithm, and will approach the 75% accuracy duration. Top right: query page waiting for a response as to which shape was in the center. Bottom right: query page waiting for a response as to the location of the car on the periphery.

In the third test, the “selective attention” subtask, the participant is again shown either a car or a truck in the center and a car at one of eight possible locations on the periphery, and again is asked to determine what was shown in the center and where the peripheral image appeared. In this subtask, however, distractor images are shown throughout the peripheral space at the same time as the central and peripheral targets. These distractor images are white triangles.

In the fourth test, the “same-different” subtask, the peripheral task remains locating a car image within a field of white triangles. The central task, however, now involves two images (each still either a car or a truck); the participant is now asked to determine whether the images in the center are the same or different, rather than simply identifying them. Thus, there are still only two responses required, but the central task is now more difficult.

The UFOV was originally designed for touchscreen, but has since been shown to work equally well with a mouse pointer for responses.(49;50)

There have been several criticisms of the UFOV and its use as a divided attention test. First, the car and truck images seem arbitrary. As it was intended as a driving test, vehicles might make some sense, but these images are so loosely tied to their intended meaning that they could have been more carefully crafted. There are only two lines' difference between the car and the truck. This makes them easily confusable – which might be an intentional design consideration, except that the fact that there are only two images to choose from means that someone who never saw either image still has a 50% probability of guessing correctly. This makes the determination of the 75% accuracy duration difficult at best. Further, the fact that the car has two more white lines means that it has a higher luminance; from a research standpoint, it is difficult to tell whether the participant is responding to the image itself or just the difference in brightness. This criticism also extends to the distractor triangles on subtests 3 and 4; they have a different number of white pixels than either car or truck image.

In terms of its widespread use in research, the fact that these images are proprietary also causes difficulty. This is a commercially available test, and as such any research or improvements on it not performed by its creators must either be licensed or completely recreated, as we have done here.

In terms of subtask 2 and its use as a divided attention test specifically, the criticism is that the only thing that changes from trial to trial is the duration of the stimulus presentation. Thus, while divided attention is certainly being utilized during the test – in order to observe the identity of the image in the center simultaneously with the location of the image on the periphery, one must use divided attention – the participant's divided attention ability is not really being *measured*. Only the speed of processing during a divided attention task is being measured. In order to measure the divided attention of a participant, as compared to others or to some personal baseline, one would have to know how varying the difficulty of one task affects performance on the other task. This would allow, as is shown in the following sections of this paper, a calculation of the interaction between the two task difficulties, which would indicate the ability (or lack thereof) of a participant to divide attention in order to successfully perform both tasks at the given difficulty level.

Nevertheless, the UFOV has been highly correlated with both age and driving-related tasks.(50;62) As divided attention is such a major factor in driving, and tends to decline earliest during age-related cognitive impairment, the UFOV is also seen as one of the best tests of divided attention currently available.(48) Specifically, subtask 2 (UFOV-s2) is far better than the tests most neuropsychologists give for divided attention (see Background: Current Testing Practices), and is utilized as the current best standard test for divided attention in our study.

Driving and the Elderly

The ability to drive is a key component in the independence of an adult in our society. It is celebrated as one of the first concrete ways of pulling away from our parents during teenage years, and is, after being unable to stay in one's home, arguably the most difficult independence-related privilege to lose in one's old age. Many elders see losing driving privileges as "the beginning of the end" in terms of losing their autonomy, and this has been linked to depression, loss of social structure, increased isolation, and hastening of physical (as well as cognitive) decline.(51) One of the primary cognitive skills involved in driving is divided attention.(26;29;39;52) It is vital, when on the road, to be able to process information from multiple sources at once, and combine these into a prioritized list indicating how the driver should react. The driver's attention is generally focused outside and in front of the vehicle; however, this focus is constantly shifting to provide a composite picture of the surrounding driving conditions. For example, when crossing an intersection, it is just as important to know who is entering the intersection from either side as it is to know who is currently occupying the intersection.(23) Especially when unexpected situations occur, the ability to attend to simultaneous stimuli and respond appropriately equals the ability to stay safe while driving.(53;54) Thus, it is incredibly important that driving tests incorporate this skill; on-road tests incorporate it somewhat automatically, as it requires divided attention to successfully navigate the road. However, any off-road tests for driving ability must incorporate an assessment of this skill, and indeed, it is a major component of the current best off-road driving test available (the Useful Field of View test).(36;48)

Driving is by far the preferred travel method for older Americans (as well as Americans in general); this is especially crucial to those living in rural areas where public transport is impractical.(4;55) As the population of elderly in this country is currently growing faster than any other age group, it should be unsurprising that the number of elderly drivers is also growing extremely fast. According to the US Census Bureau, there are currently more than 35 million Americans over the age of 65; 27 million of these are licensed drivers. If the same percentage (~80%) continue to drive, by 2030 it is estimated that there will be more than 57 million elderly on the road.(56)

While licensed, these drivers are not necessarily roadworthy. Many states do not require an on-road test for license renewal, and most states allow several years to pass between renewals; in Oregon, for instance, drivers only have to renew licenses once every eight years. Croston, et al. found that 28% of elders with dementia surveyed were still driving.(57) Perkinson, et al., found that family members and other stakeholders generally do not think that diagnosis of mild Alzheimer's should prevent people from driving.(58) Older drivers do self-regulate, tending to stay off of the road at night and during inclement weather. However, with no objective, non-dangerous way to measure their driving ability, they are reduced to regulation based on their confidence, a notoriously poor measure of actual ability in humans.(59) Further, elderly drivers are associated with over 150,000 injury-producing accidents each year;(60) clearly, older drivers are an important public safety issue. More and more frequently, these elderly will be "aging in place" – living in their own homes for as long as possible, possibly utilizing part-time nursing staff or other caregivers, rather than enter into assisted living facilities. About 7.4% of Americans aged 65 and older lived in nursing homes in 2006, compared with 8.1% in 2000 and 10.2% in

1990;(61) this trend is expected to continue, with only 4% living in assisted living facilities by 2030.(1) This suggests that an in-home monitoring system will become more and more crucial to the ongoing cognitive health of our aging population.

As described above, currently the off-road test most highly correlated with on-road driving ability, particularly among elders, is the Useful Field of View test.(36) However, this test can only be taken once every six months at most, because it can be learned.(62) Thus, the best solution for an in-home monitoring system of driving ability would be an unobtrusive, continuous measure of naturalistic behavior in the home environment which nevertheless can predict performance on the UFOV. This would allow a personalized baseline of performance to which future performance could be compared, daily or weekly “spot checks” of performance relative to that baseline, and longitudinal data to identify trends and cycles over longer time periods. Our current work is to develop just such a system using metrics based on casual computer game play.

Benefits of Continuous Monitoring

The ability to monitor changes to cognition over time, identifying trends and outliers, is a vital aspect of care for elders – particularly those at risk for, or already diagnosed with, Mild Cognitive Impairment (MCI).(63) The best methods for gathering data on these changes generally involve some form of continuous monitoring device – that is, something which takes information from the daily activities of the individual. This is achieved through either mobile monitoring devices which can travel with the elder, or stationary devices that exist within a physical or cognitive domain that the elder visits often, such as the home, the car, or a favorite website.

Some of the most compelling reasons to use continuous monitoring stem from its efficacy at diagnosing cognitive decline relative to current techniques; these are discussed above (see Current Testing Practices). Briefly, continuous monitoring allows patients to be compared to their own baselines rather than national averages, enables analysis of the variability in behavior and cognition rather than diagnosis from single-point testing, and encourages affected parties to view the decline as a continuous trend rather than a binary “pre-diagnosis” vs “post-diagnosis” set of cognitive abilities.

However, it is important to note that continuous monitoring offers unique advantages over simple diagnostic improvement. First, early detection of cognitive decline offers a huge boost to available treatments, as drug and other therapies designed to slow decline can extend good cognitive health rather than just prolonging the decline.(64;65) Second, it allows researchers a much better chance to understand the onset of MCI, vastly improving the chances of finding a true preventive therapy or cure. Third, early detection gives all those affected by cognitive decline – certainly the patient, but also family, friends, and other caregivers – the time to discuss the future and make the difficult decisions such a diagnosis necessitates.(4) Further, the community resources available to those learning to live with Alzheimer’s Disease or dementia, including physician teams, websites and facebook groups, and in-person support groups, are all more effective when routines can be established prior to the onset of severe decline.(66)

The final reason that in-home continuous monitoring is becoming increasingly important is less cheerful: it is estimated that mild to moderate remains undiagnosed in up to 75% of cases.(63;66) As the elderly population more than doubles over the next twenty years, this “hidden” problem is going to come more and more into the spotlight, and assisting these individuals without the benefit of a clear diagnosis will put considerably more strain on our medical system. It is vital that in-home continuous assessment tools be developed and implemented on a wide scale, so that elders can be informed at the start of cognitive decline and get the preventive care they need, rather than waiting for a single-time diagnosis (on a 6 to 12 month schedule) and depending on far more expensive remediative care in the interim.(67)

Computer Games for Cognitive Health

Computer games have been utilized in research for education, training, and cognitive remediation for the elderly since the early 1980s. The earliest studies investigated the efficacy of improving processing speed and reaction time through computer game play.(68) These quickly led to further studies about the possibility of computer games improving general cognition, motor skills, and activities of daily living among the elderly.(69) In more recent years, this poorly-understood connection between video games and general cognition has been the source of incredible claims in some sections of the video game industry. For example, Nintendo claims its Brain Age game can "help keep your mind young."(70) Similarly, online "brain game training" programs like lumosity.com are becoming ever more popular. The evidence for significant cognitive remediation (or even postponement of decline) is tenuous at best. No longitudinal data regarding these games exists, and most claims are based on daily pen-and-paper training sessions, which are more akin to human-interactive cognitive health coaching solutions than computerized game-like activities.(71) This hasn't stopped these companies from making a large amount of money from people who would like to fend off cognitive decline.

One system which deserves to be mentioned in this context is the Posit Science program started by Michael Merzenich. Based on Dr. Merzenich's work on neural plasticity, the program seeks to train the brain in daily tasks which often cause trouble for elders – especially those beginning to suffer from cognitive decline. For example, the program has elders practice noting whether one tone played after another is slightly higher or lower than the first. As these tones get closer together within the human vocal frequencies, this practice helps elders to distinguish between similar sounds in normal everyday speech. This program has been shown effective at short-term remediation of the symptoms associated with cognitive loss.(72) The only drawback to this work is that it's extremely boring, especially compared to video games. Nevertheless, according to PositScience.com, several hundred people have been willing to go through the system in order to achieve the results.

The relationship between divided attention and computer games is also well established: for certain specific games, or for the broad category of "video games", more play seems to correspond to higher divided attention. For example, in 1994, Greenfield et al. showed that college students who played more video games had a higher divided attention ability.(73) Koepp, et al. found that video games cause increased dopamine production, which has implications for attention, learning, and memory, as well as reward and addictive behaviors.(74) Castel, Pratt, and Drummond studied visual attention tasks and showed that video game players had faster reaction times.(75) Green and Bavelier showed that college students playing First-Person Shooting (FPS) games improved their divided attention relative to those that played non-FPS games or those who played no games.(76) Belchior extended this finding to elders as well.(77)

Green and Bavelier have also investigated several different types of games (FPS, multiplayer teamwork, platforming, etc.) to discover what characteristics are most relevant to player improvement in cognitive skills. They found that the most important game features for training visual attention are a fast game and unpredictable stimuli.(78) However, it is important to note that none of the games tested were casual games. The closest game to a casual game that was investigated was Tetris, which really has no aspects of divided attention.

Casual games are a new market that has emerged in gaming over the last decade, primarily due to the spread of internet usage worldwide. These games are characterized by being “easy to learn, but hard to master.”(79) They typically minimize the “barrier to entry” of play – they are single-player, require no skill to begin, and require a tiny per-game time commitment relative to other genres.(80) Though arguably the first casual games were card games like Solitaire, the average casual game is now hosted on an internet site and played in a web browser; again, this minimizes the barrier to entry in that the game is typically available from any computer hooked up to the internet.(79)

Considering the unique focus of casual games, it should not be surprising that the demographics of their players also differ from the stereotypical gamer. Rather than the typical 18-25y.o. male “gamer,” the average casual game player is 35-50, female, and has children.(81;82) Another surprising aspect of these games is how *much* people play them. Each individual game commands less devotion from its fans than a typical “hardcore” console game; however, collectively, the average player spends about fifteen hours a week on these games.(81) This is partly due to the large number of casual games available, but also due to the fact that they can increasingly be played anywhere – at home, at work, on a plane, on a phone, etc. These games are everywhere, they take very little time to complete, and they give a small sense of satisfaction or distraction in the middle of an increasingly overscheduled, technology-driven life.

The reason that casual games are so vital to the study of cognitive decline, and divided attention in particular, is that elders love casual games. The entire point of using games for health and educational purposes is to harness the self-reinforcing (i.e., fun) nature of the game for a higher purpose.(83) Studies of the theoretical value of playing games that people will not, in fact, play seems to negate the purpose for using games in the first place. Nevertheless, previous studies of computer or video games with regard to cognitive detection or remediation were generally performed with games that elders would not play on their own. For instance, elders do not typically seek out "platforming" games like Mario Bros. or Donkey Kong, nor do they tend to independently play First-Person Shooting games like Medal of Honor; however, each of these games has featured prominently in studies of the effects of computer game play on the cognitive abilities of elders.(68;77) These studies are beneficial to scientists trying to understand human attention and cognitive decline; however, their applicability to the real lives of elders is arguably very small. This is why the advent of casual gaming is so compelling to researchers hoping to have a positive impact on this population.

Letter Confusability

Considerable work has been done by different scientists over the past century to characterize the confusability (or, alternately, recognizability) of letters.(84;85) Largely this work was in the pursuit of better font creation, as unique and distinguishable letters are one of the primary metrics by which fonts are judged, or visual acuity tests, as Snellen charts and other such tests utilize different sizes of letters to estimate patient vision. In more recent years, these data sets have been applied more generally in vision research, as letters are a set of unique items known to the majority of participants (i.e., no additional training is needed when using them).

	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**…**	**Z**	**Total**
A	72	1	1	0	1	1	0	0	0	0	2	…	0	92
B	0	51	0	4	5	3	2	5	1	2	1	…	2	105
C	1	2	63	0	3	2	3	0	0	1	0	…	2	95
D	0	5	1	66	1	0	2	0	1	2	0	…	2	96
E	0	1	1	1	43	1	0	0	0	1	0	…	4	59
F	1	0	1	1	8	28	0	0	3	1	2	…	2	58
G	1	6	10	0	1	0	70	0	1	3	0	…	1	117
H	0	1	0	0	1	2	0	65	1	0	1	…	1	87
I	1	3	0	1	10	18	0	3	56	22	4	…	3	199
J	2	1	0	0	2	1	0	1	9	49	0	…	1	95
K	3	2	1	0	4	4	0	1	2	0	58	…	4	117
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
Z	3	0	1	0	0	0	0	1	2	1	0	…	54	69
Total	100	100	100	100	100	100	100	100	100	100	100	…	100	

Table 1. Example Letter Confusability Matrix. Letters shown are in the columns (100 of each), and letters identified are in the rows; correct identification is on the diagonal, and anything else is an error. So, for example, the cell at (G,B) = 6; this means that out of the 100 times that ‘B’ was displayed, it was identified as ‘G’ six times. This matrix comes from “Identification confusions among letters of the alphabet.”(85) The letters investigated in this study were uppercase Helvetica, presented very small.

The general method for testing letter confusability is to show a letter to a participant and then ask him or her to identify that letter. Repeating this process enough times, with enough different participants and different letters, will create a letter confusability matrix like the one shown in Table 1. This gives an idea of the likelihood that each letter will be confused for each other letter on any given presentation. Several of these studies also presented letters under adverse conditions, such as very distant or very brief letter presentation;(84;86;87) naturally this causes a greater number of errors, which allows characterization of the confusability with fewer trials.

Learning and Fatigue Curves in Visual Search

A great deal of work has been done showing that the time needed to find a target within a field of distractors is linearly proportional to the number of distractors present. (88) Most current theories agree that fast, efficient visual search is done in parallel (searching multiple locations simultaneously). However, theories differ on whether relatively slow search is parallel or serial (searching only one location at a time, moving attention like a spotlight from one portion of the search field to another) in nature.(88-90) The parallel theories suggest that increasing the number of distractors simply increases the amount of processing power required to search all locations simultaneously.(91;92) What is less debatable is that these sorts of visual search improve with practice; in other words, at least portions of the process can be learned. Some research suggests that this learning is in fact moving the formerly serial process into a parallel processing space.(93) Naturally, in any task-optimization-based experiment, the participants must understand how to perform the task (i.e., what to do) in order to optimize their performances. Visual search tasks are no exception. The contribution of learning to the variability of results is often minimized by including a training period; however, this does not entirely eliminate the improvements gained due to practice. Note that not all learning is conscious; the difference between being able to perceive a stimulus in 30ms versus 50ms, for example, is not one that is appreciable to the conscious mind. Nevertheless, this sort of learning can occur through repeated exposure. Schiffrin and Dumais suggest that after extensive practice with a similar set of targets and distractors, the targets tend to “pop out” of the visual field, and distractors are automatically skipped over, with far less need for active attention.(94)

Learning one task often transfers to other, related tasks. Indeed, Sireteanu and Rettenbach found that under certain conditions, learning of one visual search task can translate into learning “from one task to another, from one location in the visual field to another, and between the two eyes of a given subject, even if the subject has reduced stereopsis.”(95) However, Fahle and

Morgan found that perceptual learning did *not* generalize across two extremely similar tasks (three-dot bisection and three-dot vernier discrimination) in the same location in the visual field.(96) This suggests that learning on certain visual tasks will transfer to others, while learning on other tasks will not transfer beyond that specific task; thus, experimentation may be needed in order to determine which tasks will generalize, and the extent of that generalization, for specific novel stimuli.

In terms of this experiment, two factors are critical to the learning of visual tasks: first, the difficulty of the search task, and second, the age of the participants. Leonards, et al., confirmed that perceptual learning can take place during tasks involving several different basic features, including distribution of local brightness and angle size of nonparallel lines.(97) In each of these cases the learning curve plateaued after 15 to 80 trials for naïve subjects (and considerably faster for experts who had taken these sorts of tests numerous times before). However, when the features were combined (simultaneously searching for a particular angle size *and* brightness, for example), no learning was observed – that is, while reaction times to certain trials was very quick, overall variability in reaction time remained high and researchers never saw the drop-off and then plateau of response time which characterizes a learning curve, even after several hundred trials or with experts performing the tasks. (97) Thus it is reasonable to hypothesize that easier tasks in the current experiment, such as the standard UFOV, might be fully learned in 80 or fewer trials.(98) Difficult tasks, such as an enhanced UFOV with large numbers of peripheral distractors and high blur factor for the central letter, might never be learned.

These findings may be exacerbated when the subjects are elderly. Gilbert showed that older (mean age 72, range 65-80) adults do not experience the “pop out” effect even after extensive practice (though this practice certainly improved their performance on the given task). This lack of “pop out” allows elders to switch more easily between target and distractor sets, but requires constant effortful attention in order to perform the search tasks.(99) Thus, age may play a role in how, and how quickly, learning of a given search task is accomplished.

The Psychometric Function

Psychometrics is, as the name implies, the study of the measurement of psychological phenomena. Naturally this is extremely important to experiments attempting to define and measure cognitive skills like divided attention. The psychometric function (also known as the psychometric curve) is a function often used to describe the results of psychometric experiments that have a varying-intensity stimulus to which a participant can respond correctly or incorrectly for each trial. Typically, this function is graphed with the percentage of correct responses on the Y axis and the intensity of the stimulus on the X axis.

The graph proceeds in an s-curve from the lower left to the upper right (see Figure 2): theoretically, at sufficiently low intensities of the stimulus (on the left side of the graph) the participant will always respond incorrectly, while at sufficiently high intensities (on the right) the participant will always respond correctly. Of course, in reality, the participant will generally respond correctly by chance some portion of the time even with no stimulus at all. Likewise, being human, the participant will sometimes make mistakes even when he or she recognizes the stimulus correctly. Thus, the left and right ends of the graph never really reach 0% or 100%, though they may come close. In between those two extremes, its shape is sigmoid, having the greatest slope at the stimulus intensity point corresponding to 50% correct response likelihood, and becoming increasingly shallow towards the left and right endpoints.

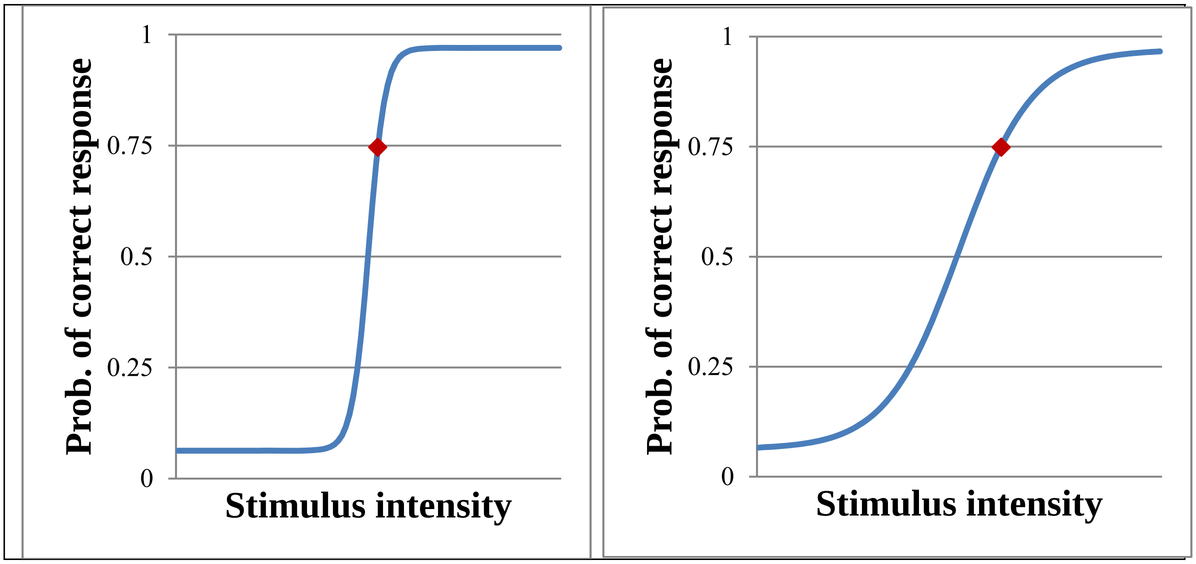


Figure 2. Psychometric function examples with steep (left) and mild (right) slopes. Note that the X axis is left without units, as these examples could apply to any psychometric data. As the intensity of the stimulus increases, the probability of responding correctly also increases. The 75% accuracy threshold value (the value estimated by UFOV) has been marked. The probability of correct response goes from some calculable minimum (the "guess" probability) greater than zero to some estimable maximum (the "mistake" probability) less than one. The slope is determined by the variability of the data; a steep slope (left) means less variability around the threshold, while a mild slope (right) means much more variability.

Much like any logistic regression, the utility of this function in dealing with psychometric experiments lies in its predictive property. It's rarely possible to measure participant responses to every interesting level of stimulus intensity. Even if one had the time and resources, it might turn out later that an unmeasured intensity level was of interest. Once the responses have been mapped to a psychometric curve, researchers can use the function describing that curve to estimate the probability of a correct response at any given stimulus intensity, whether an actual measurement happened there or not.

Thus, in certain experiments like this one, the Y axis represents the *probability* of the participant responding correctly to the given stimulus intensity, rather than the actual observed percentage of correct responses at that intensity. Luckily, this can be calculated with far fewer measurements. Wichmann and Hill describe a method for fitting a psychometric function from collected data that include participant responses at given stimulus intensities.(100;101) In brief, the function is found by maximizing Θ in the log-likelihood function

$$
\begin{aligned}
l(\theta; y) = \sum_{i=1}^{K} \log \binom{n_i}{y_i n_i} + y_i n_i \log \psi(x_i; \theta) \\
+ \left((1 - y_i)\right) n_i \log\left[1 - \psi\left(x_i; \theta\right)\right].
\end{aligned}
\qquad (2)
$$

In this equation, which is equation 2 from Wichmann-Hill 2001a,(100) *i* iterates through each unique stimulus intensity that was tested in the experiment, with *K* being the total number of unique intensities tested. *x* represents the value of the stimulus intensity (x_i the value of intensity *i*). *n* is the number of individual trials run at intensity *i*. *y* is 1 if the participant responded correctly and 0 if incorrectly; thus, y_i is the total number of times that the participant responded correctly at intensity *i*, and $y_i n_i$ is the percentage of correct responses out of the total number at intensity *i*. Θ is the vector involving the parameters which entirely describe the psychometric function: α, the shape parameter of the psychometric function; β, the scale parameter of the psychometric function; γ, the minimum value representing guessing correctly by chance; and λ, the distance between 1 and the actual maximum value representing selecting the incorrect answer despite having correctly seen the stimulus. These are the parameters that will be optimized by maximizing this log-likelihood function. *ψ* represents the value of the psychometric function itself, evaluated at the point described by input values x_i and Θ. Thus, the first part of the equation (before the first "+") represents the number of different ways one *could* respond correctly exactly the number of times that this participant *did* respond correctly. The second part (between the two "+"s) is the number of times the participant responded correctly times the probability of responding correctly with the current intensity and Θ parameters, and the third part (after the final "+") represents the number of times the participant responded incorrectly times the probability of responding incorrectly.

In terms of this experiment, each trial set (that is, a specific set of task conditions, presented with stimulus durations determined by staircase algorithm, until a 75% accuracy threshold duration can be determined) describes a unique psychometric function. In this experiment, the “stimulus intensity” is the duration of the stimulus presentation. With very short durations, participants are reduced to guessing in order to respond; with very long ones, they will theoretically respond correctly except by mistake. In our experiment, i iterates through durations, and x_i is the length (in milliseconds) of a particular duration. Thus, maximizing this function will give us the parameters α, β, γ, and λ, which together define the psychometric function for a particular participant for a particular set of conditions at a particular time.

As can be seen in Figure 2, the slope of the psychometric function is determined by the variability of the data: if a participant might respond correctly or incorrectly to a wide range of stimulus intensities, then the slope will be quite shallow; by contrast, if the participant always responds correctly above the threshold and incorrectly below it, the slope will be nearly vertical. Thus, the slope of the function can be estimated from the variance of the raw data around the threshold, as will be described in the analysis of the main experiment.

Previous Research

In 2005, OHSU teamed up with Spry Learning Co. to develop a suite of cognitive computer games. These games were intended to monitor and help remediate cognitive decline as a small part of a holistic unobtrusive in-home monitoring solution. Some games incorporated variants of standard cognitive tests; others were completely novel. In total, nine games were developed, each intended to address a specific part of the cognitive spectrum. Targeted cognitive abilities included executive function/planning, working memory, set switching, verbal fluency, word formation and recall, and of course divided attention.(12)

The specific game designed to assess and monitor divided attention ability was called 21 Tally (see Appendix III for screenshots), and can be thought of as blackjack in two dimensions. In this game, the board consists of sixteen open spaces (four rows and four columns) in which a player can play a card. The next card to be played appears in the lower left hand corner, and when it is placed on the board the card beneath it becomes the next card. Players attempt to place cards in such a way as to make a row or column (or both) add up to 21 without going over. Card values are equal to their number, except for facecards, which count as 10, and aces, which count as 1 or 11 (whichever is better for the player). Getting exactly 21 in a row or column is known as a "tally" and scores points, while going over 21 is known as a "bust" and removes points. Each time a player gets a bust or tally, the cards involved are removed from the board, so that the board never fills up with cards completely. Additionally, getting enough tallies on a given round will allow the player to proceed to the next round, while getting too many busts on a round causes the player to lose the game.

Figure 3. Screenshot from the 21 Tally game currently being played by cognitive health coaching participants. This game is essentially blackjack in two dimensions at once; players attempt to make cards add to 21 (but not more) in either rows, columns, or both. Players who fail to pay attention to both directions, thereby demonstrating less divided attention ability, make more errors in the less-attended direction.(102)

The reason that this game in particular deals with divided attention is that players must attend to both rows and columns simultaneously or risk a bust in the unattended direction. Additionally, while high divided attention skill may help in getting good game scores, skill at the game is not completely dependent upon divided attention skill, meaning that players with lower divided attention skill can also play and enjoy the game. For example, players with high game skill and low divided attention will score lots of tallies and very few busts, but typically only in one direction. Conversely, players with high divided attention and low game skill will score similar numbers of tallies in each direction but will also score large numbers of busts in each direction.(102)

A one-year proof-of-concept study showed that elders motivated to stay cognitively healthy would play these games several times per week. This study also administered a battery of neuropsychological tests to participants, allowing researchers to show relationships between several of the games and the cognitive skills that they were intended to monitor. (12;12;33) Unfortunately, divided attention was not included in these test batteries. This may be due in part to the fact that there currently is no widely-used, standardized test of divided attention; the proprietary and somewhat controversial Useful Field of View test is as close as we have come. The lack of a good divided attention test and the need for validation for 21 Tally as such a test formed the impetus for the work contained herein.

Chapter 2: Pilot Study

Overview

This pilot study was conducted to determine the feasibility of using the particular central and peripheral difficulty changes that we intended, as well as to rectify any obvious usability complaints and per-trial time commitment. This last was vital due to the fatiguing nature of repetitive cognitive tasks, and the increasing likelihood of this interfering with results as experiment duration increased.

Our results suggested that the central task would have to be modified from an increased number of possible letters to a task that would produce a measurable increase in threshold duration correlated with difficulty. Additionally, we found that learning might have a significant impact on threshold duration. To determine how to best account for this, we ran a small follow-up study investigating the learning curve, the results of which showed that learning for each condition could be roughly approximated by noting the learning rate on a single, "standard" condition.

Overall Procedure

Participants for the pilot study were recruited through email and phone calls from among friends and family of the researchers. Participants were pre-screened over the phone to ensure minimum participation requirements were met before scheduling them to come in to the lab. Once participants arrived, they were run through a training procedure for both UFOV and the 21 Tally game. They then performed a series of trials with each without EEG, followed by a repetition of a select set of trials with EEG. Participants were debriefed after their EEG participation. Finally, participants were given a $10 gift card to Starbucks as thanks for their participation in the pilot study.

Participant Selection

Participants for the pilot study were recruited from among friends and family. Potential participants were screened for baseline computer ability (ability to move a mouse pointer to a point on the screen and click, or verbal assertion of familiarity with computers and computer mice) and visual acuity (self-report of vision no worse than 20/70 and a vision test at least as recently as five years ago). No participants were screened out of the study due to these requirements.

This provided a pool of ten participants, ages 28-69, six of whom were female. One participant had completed some college, three had four-year college degrees, two were pursuing advanced degrees, and five had completed terminal degrees (MFA, JD, PhD, etc.).

Methods

Participants were invited back to the lab and were given an initial survey (see Appendix I) to determine eligibility, demographics, and gaming preferences. No participants reported significant uncorrected vision problems; thus, none were removed from the study at this stage.

Enhanced Useful Field of View

Participants were given a suite of different variants on the Divided Attention subtest of the Useful Field of View (UFOV-s2). These included the standard UFOV-s2, as well as several enhanced versions. Each followed the same procedure (described in Ball & Owsley, 1993).(48) First, the subjects were seated approximately 24 inches from a 26-inch LCD monitor with a refresh rate of 60Hz. The subjects then went through a series of training exercises to familiarize them with the general procedure of detecting visually-presented targets in a sequence of self-paced trials. Instructions were given in text form on the screen (see Appendix II), followed by opportunities for participants to practice the directions before “official” testing began. Responses on the practice session were recorded. Participants were required to answer correctly on three examples of each type of task. They were allowed to practice until they felt comfortable performing the tasks before moving on. At the beginning of each trial, subjects were asked to fixate upon a central point; after a randomized waiting period, a test image appeared for a limited amount of time. The test image consisted of one target presented at the center and another presented with zero or more distractors on the periphery. After a brief period (beginning with 400ms), the stimuli were replaced by a masking image with the same average intensity as the stimulus image. The subject was then asked to indicate both the identity of the central stimulus and the position of the stimulus on the periphery. The presentation duration was adjusted via a staircase method, reducing presentation duration after correct responses and increasing it after incorrect responses, in order to ascertain the duration that produced approximately 75% correct responses.

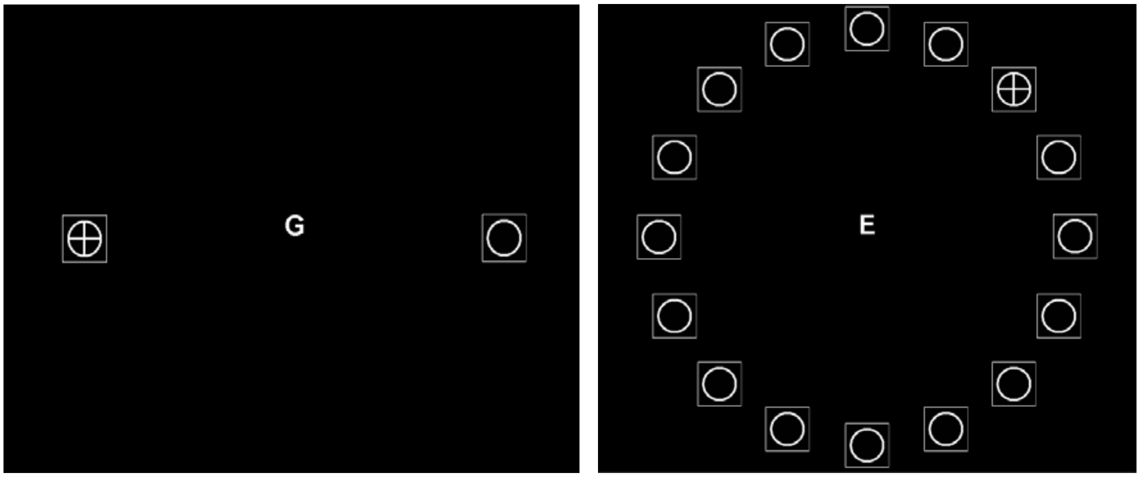


Figure 4. Example screenshots from the enhanced UFOV portion of the pilot study. These demonstrate the range of difficulties used in the experiment. For screenshots of interim difficulties, see Appendix III.

Each set of trials (beginning with a 400ms stimulus presentation duration and ending with the duration at 75% accuracy) had a particular configuration of stimuli for each of the two tasks. Though described verbally here, example screenshots can also be found in Appendix III. The standard UFOV-s2 configuration has two possible choices for the central task (car or truck) and eight possible locations for the peripheral stimulus, with no distractors on the periphery. In addition to that configuration, we had conditions involving letters as the central identification task and conditions using a peripheral task with a cross as the target and empty circles as distractors in all other peripheral locations. The number of items to choose from in the central task varied between 2, 8, and 16. The number of peripheral locations varied between 2, 4, 8, 12, and 16. Configurations were presented in random order, though the standard UFOV-s2 was given first and repeated at the end. The full list of sets of conditions given can be found in Table 2, and examples of the difficulty extremes of the new tasks are displayed in Figure 4.

Center Type	# Centers	Peripheral Type	# Peripherals
car and truck	2	car	8
car and truck	2	car	8
car and truck	2	car	12
car and truck	2	X and O	2
car and truck	2	X and O	4
car and truck	2	X and O	8
car and truck	2	X and O	12
car and truck	2	X and O	16
Letter	2	X and O	2
Letter	2	X and O	4
Letter	2	X and O	8
Letter	2	X and O	12
Letter	2	X and O	16
Letter	8	X and O	2
Letter	8	X and O	4
Letter	8	X and O	8
Letter	8	X and O	12
Letter	8	X and O	16
Letter	16	X and O	2
Letter	16	X and O	4
Letter	16	X and O	8
Letter	16	X and O	12
Letter	16	X and O	16

Table 2. Conditions for the enhanced UFOV. Note that configurations with “car” as the peripheral type had no distractors in the periphery, whereas those with “X and O” had distractors in each non-target peripheral location. Letters used in the “letter” central task were the first letters in the alphabet up to the given number (e.g., “letter: 2” would have the participant choosing from the letters ‘A’ and ‘B’, whereas “letter: 8” would have the letters ‘A’ through ‘H’). All letters were uppercase. The first row represents the conditions for the standard Useful Field of View subtask 2.

After completing trials determining 75% accuracy on each of these sets of conditions for the standard and enhanced UFOV, participants were given the “Pilot Study Post-UFOV survey” (see Appendix I). This survey explores both the participants’ familiarity with the UFOV and their subjective experience of taking the test (e.g., did they feel like performing these tasks improved their divided attention skill, would they be willing to perform these sorts of tasks on a daily or weekly basis, etc). No subjects reported that they had previously encountered or taken any version of the UFOV.

21 Tally Puzzle

Participants then returned to the computer to perform a similar series of tasks based on the 21 Tally game. Differences between these tasks and the actual game were as follows. First, boards were not contiguous; that is, play on board X did not affect the cards displayed or available on board X+1. All boards had eight cards and eight empty spaces, to avoid changes in cognitive load based on the number of cards. No aces were used, to avoid the possible confusion of points (1 versus 11). Each board had a certain number of different decisions to be made (0, 1, or 2) regarding the score in each of the different directions (rows and columns). For instance, a board which had 0 decisions in the rows would have all moves making 21 in the rows, or all moves making a bust, or all moves making neither bust nor 21 – regardless, all moves would be the same in the row direction. That same board might have 2 decisions in the columns, meaning that at least one move would bust, at least one would 21, and at least one would neither bust nor 21, in the column direction. A direction with 1 decision might have some moves which caused a bust and some which 21, or some which bust and some which do neither, or some moves which do neither and some which 21. Boards with 0 decisions in both directions were not tested, as all moves would have been equivalent and participants could presumably have responded correctly with a stimulus duration of 0ms.

Additionally, each board had a single “best” move, though most had several equivalent less-than-best moves; for instance, if the single best move had a bust in one direction and a 21 in the other direction, there might be several places that had a bust in one direction and neither bust nor 21 in the other direction.

Figure 5. Two screenshots from the Tally Puzzle. Left: stimulus, which will be displayed for a certain duration based on the participant’s correct or incorrect responses. Right: awaiting response, in which the participant can see card placement but nothing else. Note that there are equal numbers of cards and blank spaces, and that there is only one best answer (in this case, the bottom right corner, which will not bust in either direction).

The final important difference between these tasks and the normal 21 Tally game is that participants were only allowed to look at card values and row/column sums for a certain period of time. After this period, cards turned over (so that their position could be seen, but not their value) and row and column sums were hidden (see Appendix III for a screenshot). Thus, participants could take as long as they liked to pick the best place to move, but the actual information on the board was only displayed for a certain duration. This duration was varied according to the same staircase algorithm used in the enhanced UFOV, to arrive at the duration which gave 75% accuracy in terms of selecting the best spot to play.

Just as in the UFOV trials, participants first went through an instruction and practice session (see Appendix II for the actual text of the instructions). During this session, participants were instructed that both rows and columns were important in making 21s and avoiding busts (22+). They were given several boards to practice, and then were introduced to the concept of only viewing the board for a limited time. No participants seemed particularly confused by this aspect, possibly because they had just been performing Enhanced UFOV tasks which employed the same limited-stimulus-duration concept. During each practice step, participants were required to select the best move on three unique boards; they were given the option to practice further until they felt comfortable performing the tasks.

After the instruction and practice session, participants were asked to perform a series of trials on each possible combination of 0, 1, and 2 decisions in each of the row and column directions (for example, 0 row decisions and 1 column decision, or 2 row decisions and 2 column decisions). As no “best” move could exist on such a board, 0 row decisions and 0 column decisions was considered an impossible combination; this meant that there were 8 unique sets of conditions. The condition of 1 row decision and 1 column decision was given at the start and repeated at the end to give an indication of how learning had affected each participant; thus, there were a total of 9 trial sets administered during this part of the experiment. Presentation for each other set of conditions was randomized. Each series of trials continued until the staircase algorithm could estimate the stimulus duration that produced 75% accuracy.

After completing trials determining 75% accuracy on each of these sets of conditions for the Enhanced 21 Tally, participants were given the “Pilot Study Post-Tally survey” (see Appendix I). This survey questions participants about their previous familiarity with 21 Tally and their experience of it as a game, including specific questions about the possibility of playing this game repeatedly as part of a cognitive health solution. No subjects reported that they had played 21 Tally before.

EEG for UFOV and 21 Tally

We investigated electrical signals originating in the brain through electroencephalogram (EEG) to support our behaviorally-based assumptions about cognitive load during the different parts of the experiment. If the EEG showed evidence for increased cognitive load during tasks with increased difficulty, then this would give further evidence that difficulty was in fact changed, even for participants who showed no significant difference in 75% accuracy stimulus duration threshold for those tasks.

Participants were hooked into an EEG involving a cloth cap with 32 leads and running BioSemi ActiView software for data collection. Electrolytic gel was used for conductivity enhancement; participants were not requested to shave their heads.

While recording the EEG, participants were asked to repeat certain select trials of the Enhanced UFOV and Enhanced 21 Tally. For UFOV, these included the standard UFOV-s2 for reference, as well as four trials representing the combinations of an easy central task (2 central letters), an easy peripheral task (2 peripheral locations), a difficult central task (16 central letters), and a difficult peripheral task (16 peripheral locations). Likewise for the 21 Tally, a reference condition was given (1 row decision and 1 column decision), as well as two easy tasks, two difficult tasks, and one difficult with one easy task. In the format (row decisions, column decisions), the conditions given were: (1,1), (0,1), (1,0), (1,2), (2,1), and (2,2). In both UFOV and Tally, the reference condition was given at the start and again at the end, to give an idea of the vector of learning throughout the process; the order of all other conditions was randomized.

After completing the EEG portion of the experiment, participants were given the “General Post-Participation Questionnaire” (see Appendix I). This survey focused on qualitative questions like subjective length of the experiment, designed to assist in modification of our main experiment to accommodate elders in terms of participant experience, if required.

Unfortunately, due to restructuring of faculty and student positions in the department during this portion of the experiment, we currently lack the expertise to isolate a signal corresponding to cognitive load. Thus, these data have been stored in the repository until such time as they can be adequately analyzed.

Results

As expected, all participants were able to complete both the UFOV and Tally portions of the experiment after brief training sessions; each participant took an average of 28.6 trials to complete the training portion (total training trials ranged from 15 to 84). Participants had a reasonably large range of UFOVs2 scores (15 – 315), suggesting a range of both divided attention ability and speed of processing.

As mentioned previously, when attempting to dissociate speed of processing from divided attention ability, it's important to be able to vary the difficulty of both tasks so as to be able to determine the effect that changing one task's difficulty has on performance of the other task. The magnitude of this change is the best measure we have of divided attention ability, as opposed to speed of processing or any of the other things that might influence performance on a particular trial set like the standard UFOVs2.

Thus, one of the primary goals of this pilot study was to show that our methods for varying difficulty in each task did in fact produce changes in performance commensurate to the change in difficulty. Figure 6 shows the results of varying difficulty for the central task (through increasing or decreasing the number of letters from which the center image might be drawn), while Figure 7 shows the same for the peripheral task (with difficulty adjusted through increasing or decreasing the number of locations that the peripheral image might have appeared).

As it can be seen, increasing the number of peripheral locations does seem to increase the stimulus duration needed for participants to achieve 75% accuracy (Figure 7). Especially when the central task was at its least difficult (i.e., when faced with the minimum amount of interference from the central task), adding peripheral targets seems to increase the difficulty of the task in a somewhat linear fashion. This is as expected from previous research.(88)

By contrast, adding to the number of letters from which the central image might be drawn does not seem to particularly affect the stimulus duration needed for 75% accuracy (Figure 6). Indeed, when the peripheral task is kept at minimum difficulty (four peripheral targets), there is almost no difference between a central letter chosen from two possibilities or sixteen.

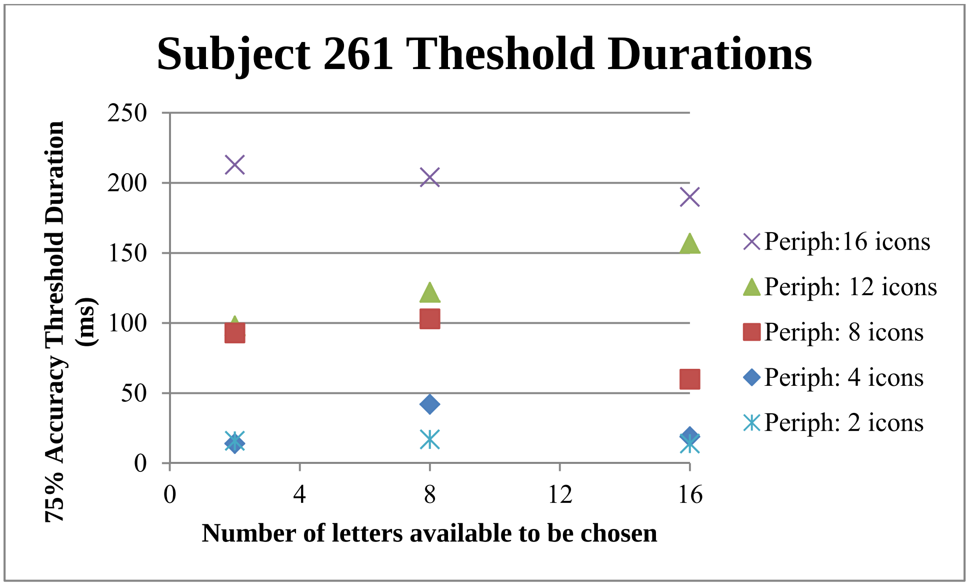


Figure 6. Central "difficulty" and 75% accuracy duration. Note that the stimulus duration needed to achieve 75% accuracy generally does not increase when adding more items into the "bag of letters" for the central task. This suggests that the difficulty of the task is largely independent of the number of different images the central letter *could* have been. In other words, the participant either saw the central letter or not; the only time that the non-displayed letters affect the difficulty is when the participant did not see it and is guessing. Thus, this is not an effective method for controlling the difficulty of the central task.

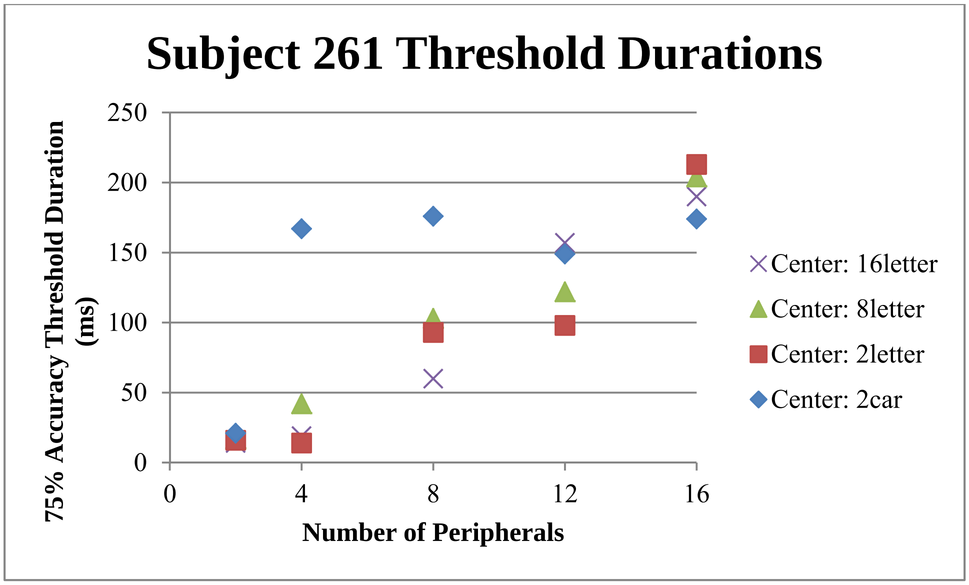


Figure 7. Peripheral difficulty and 75% accuracy duration. Note that the stimulus duration needed to achieve 75% accuracy generally increases linearly with the number of items on the periphery. This suggests that we are in fact increasing difficulty by increasing the number of distractors on the periphery, as expected.

We also built a software tool to simulate a participant performing the UFOV and Tally portions of the experiment. This piece of software had the same variability characteristics as the average participant in the pilot study; that is, given a "true" UFOVs2 score, it was as likely to respond correctly or incorrectly as the average participant. The true UFOVs2 score for human participants was the 75% accuracy duration for the first trial set with standard UFOV conditions; for the simulator, it was an input parameter.

This tool allowed us to computationally determine the statistical power we would have to detect trends with our sample size of 60, as a complementary "sanity check" to our statistical calculations.(103) By generating simulated participants who have the same true UFOV scores as each of the participants in the pilot study and have the average variability of all the pilot study participants, we can determine how many 'identically performing' participants we need in order to estimate the true UFOV score at a given confidence interval. Or, alternatively, the size of the likely confidence interval with the 60 participants we're allowed by the IRB, assuming that participants in the main study have similar variability. As each subject was treated individually without regards to other subjects' data in calculating interaction terms in the main study, this did not affect the number of subjects needed to test divided attention this way. However, for a correlation of ~0.82 between interaction magnitude generated by eUFOV versus 21 Tally, as our simulator suggested we might find, we determined that having 50 participants in the main study would yield a confidence interval of slightly less than 0.2.

Discussion

The fact that adding extra letters to the mix didn't change performance numbers substantially suggests that this didn't make the task more difficult (see Figure 6). Clearly when guessing the subjects would have a more difficult time getting the correct answer when there are more possible letters from which to guess; however, subjects typically made errors on the peripheral task rather than the central one. When not guessing, a larger number of letters from which to choose is a minor addition to the cognitive load at best. Thus, in the main study, we intend to use blurring to control the difficulty of central letter recognition, and keep the number of letters constant (see Improving the Enhanced UFOV). This blurring idea comes from letter confusability tests, which have made letters more difficult to see by making them further away, making them smaller, and blurring them.(84;86;87)

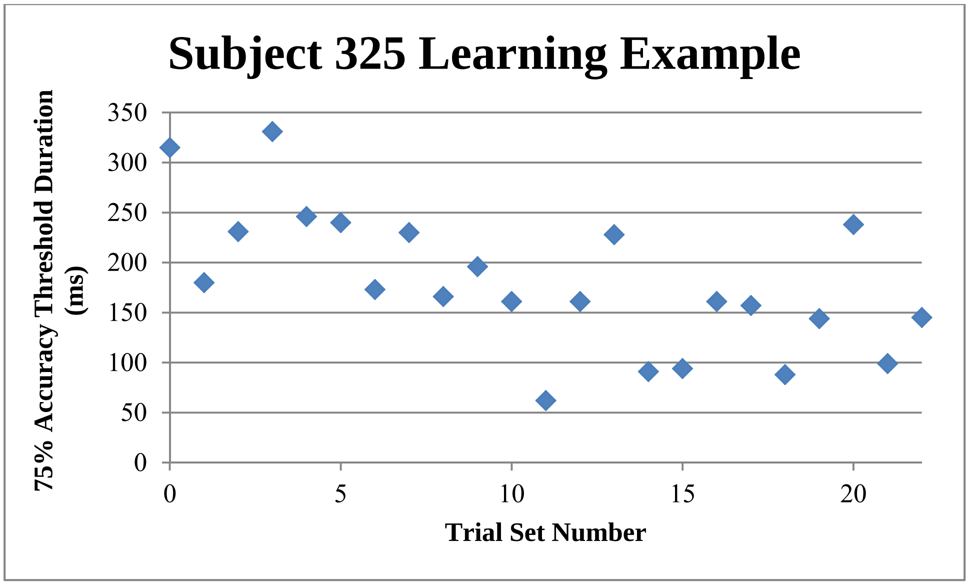


Figure 8. Number of trials and 75% accuracy duration. Note that the stimulus duration needed to achieve 75% accuracy generally decreased in this participant based on the number of trials they had performed, and regardless of the difficulty of the tasks. This suggests learning is taking place.

One other issue that should be taken into account is learning. Figure 8 shows the threshold values for a single participant over the course of the experiment; the order of difficulty conditions was randomized, but one can still see a clear trend toward shorter durations, suggesting that learning is taking place. As noted above, all pilot participants performed 23 different trial sets for the enhanced UFOV, though this number was substantially reduced for the main study (due to fatigue reported by participants). The standard UFOV is used for training of driving skill, so obviously learning takes place there.(36;77) The UFOV is also used for testing; however, this can only occur once every six months to avoid being tainted by learning of the test.(62) Thus, in addition to improving the difficulty settings of the two UFOV tasks, we must investigate whether learning can be estimated without adding a great deal to the overall experiment duration.

Improving the enhanced UFOV

As a result of our pilot study experiment, we realized that the eUFOV needed further improvement in order to fulfill its function of accurately measuring divided attention. Figure 9 shows the three major improvements: equal luminance for peripheral distractors, eight equally-confusable letter possibilities for the central task, and blurring as the primary difficulty adjustment in the center.

First, we realized that the target and distractors on the periphery were not of equal luminance; the cross in the center of the target made it brighter than the empty circles used for distractors. Thus, the task remained merely a luminance detection task rather than a true search or recognition task. To rectify this, we made the peripheral distractors have an “X” shape to match the luminance of the target’s “+” shape.

Second, the number of letters from which the central task might be chosen proved not to affect the difficulty of the task. We still wanted guessing to be relatively difficult while not overwhelming the subjects with the number of choices available when picking the letter that they saw. Thus, we chose the median number used in the pilot study: eight.

However, this still left the issue of which eight letters to choose. Since this is an identification task, it is important that no pair (or subset) of that group be any more or less confusable than any other pair (or subset). Thus, we chose to use eight letters that were shown to be similar on the confusability matrix (see Table 3). See the Background: Letter Confusability for more information on the work that led to this matrix.

Finally, the difficulty adjustment of the central task needed to cause measurable difference in performance in order to be useful for measuring difficulty interaction. Blurring has two advantages over adding more letters from which to choose. First, the algorithm for Gaussian blur is well known(104) and can decrease the recognizability of the letter (and thus, increase the difficulty of the task) by a measurable amount. In this case, the algorithm takes a weighted average of the surrounding X pixels in the original letter in order to choose the shade of each given pixel in the blurred letter. Thus, a letter blurred with a radius of 5 pixels is half as blurry as one blurred with a 10-pixel radius. Second, the use of a constant number of letters (in this case, eight) allows us to choose letters that have a constant amount of confusability between them. In other words, no letter is more likely to be confused with another than any other letter, in this particular set (see Background: Letter Confusability).

	W	**A**	**G**	**U**	**V**	**S**	**P**	**I**	**Total**
W	88	0	0	0	4	0	2	0	94
A	0	72	0	0	1	1	0	0	74
G	1	1	70	2	1	2	0	1	78
U	0	1	2	64	0	0	0	0	67
V	4	0	0	2	61	0	3	0	70
S	0	0	2	0	0	60	1	1	64
P	1	1	0	0	0	2	59	3	66
I	0	1	0	1	1	1	11	56	71
Total	94	76	74	69	68	66	76	61	

Table 3. Letter Confusability in this Experiment. These are the uppercase Helvetica letters chosen for the enhanced UFOV in this experiment, taken from Gervais, et al., 1984 (see Table 1). While not the eight most recognizable letters, they have relatively equal confusability with each other. Out of the 100 presentations of each letter, none of these letters was confused with any other more than 4 times.

Dual Attention versus Task Switching

Note that in this experimental design we are attempting to elicit dual attention, but not task switching. In the literature on dual task testing, much has been written about the temporal and accuracy costs associated with attempting to perform two tasks simultaneously.(105;106) In brief, many researchers postulate a “central processor” which controls the order of task processing,(107) and several tests have shown a “psychological refractory period” (PRP), or pause between the processing of the first task and the processing of the second.(108;109) In these experiments, switching between the tasks is hypothesized to occur a maximum of once, eliciting the PRP.(110) These effects of a dual-task experiment are expected and do not interfere with our analysis of the interaction between the difficulty levels of the two subtasks as a proxy for divided attention.

However, many experimental designs have elicited a task switching behavior in which the central processor re-prioritizes tasks (either due to explicit researcher instruction or implicit experimental changes), causing a larger amount of switching between tasks.(111;112) Since each switch has an associated cost, in terms of time, this causes the overall reaction time measured to be considerably greater. Since these effects are seem primarily in response to a changing configuration of the experimental tasks, we expect that switching between the dual tasks in this experiment will be kept to a minimum, and will not require a significant change to our analysis.

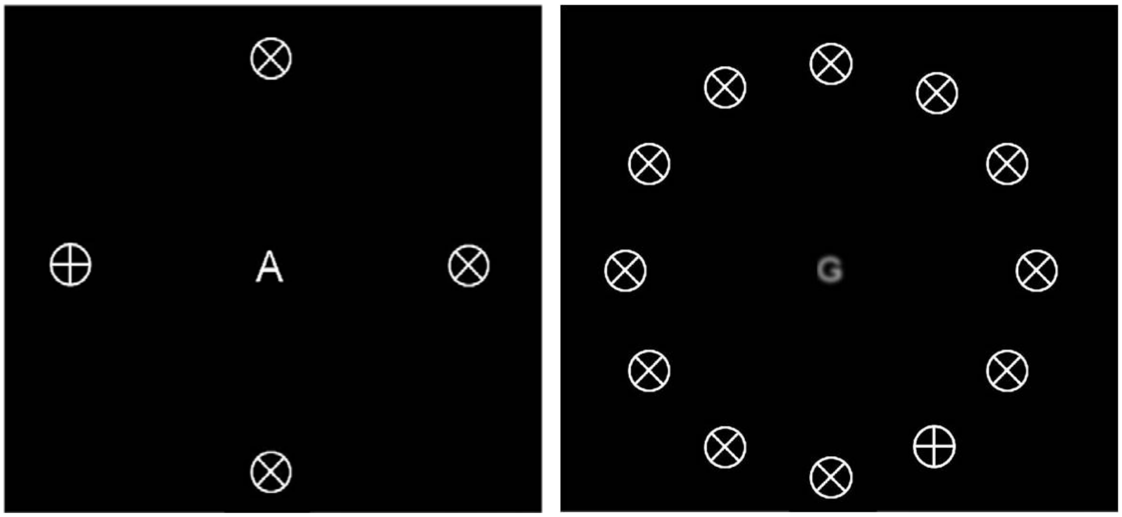


Figure 9. Two screenshots from the enhanced UFOV experiment, showing the range of difficulties. Left: easy central task (no blur) and easy peripheral task (three distractors). Right: hard central task (10 pixel blur) and hard peripheral task (eleven distractors).

Learning and Fatigue: follow-up to the pilot study

Only three elderly participants were included in this follow-up, as their data were similar enough to generalize to other subjects. In this study, the methods were very similar to those in the pilot study. However, the number and order of trial sets were altered in order to better understand how learning and fatigue affected performance, particularly of elderly participants. Additionally, the stimuli presented were modified to those intended for the main study, in order to provide further exploratory analysis before beginning the main study.

Center Type	# Centers	Central Blur Radius	Peripheral Type	# Peripherals
car and truck	2	0	Car	8
car and truck	2	0	Car	8
car and truck	2	0	Car	8
letter	8	5	+ and X	4
letter	8	5	+ and X	4
letter	8	5	+ and X	4
letter	8	15	+ and X	12
letter	8	15	+ and X	12
letter	8	15	+ and X	12
letter	8	5	+ and X	4
letter	8	15	+ and X	12
car and truck	2	0	Car	8

Table 4. Conditions for Enhanced UFOV in the Learning Curve Experiment. Note that configurations with "car" as the peripheral type had no distractors in the periphery, whereas those with "+ and X" had distractors in each non-target peripheral location. Letters used in the "letter" central task were uppercase Helvetica font and were drawn from the set {W, A, G, U, V, S, P, I}. The first row represents the conditions for the standard Useful Field of View subtask 2. Conditions were presented in the order seen here, with the first presentation at the top and the last at the bottom.

Row Decisions	Column Decisions
1	1
1	1
0	2
0	2
2	0
2	0
2	2
2	2
0	2
2	0
2	2
1	1

Table 5. Conditions for Enhanced 21 Tally in the Learning Curve Experiment. Each board contained eight cards, eight open spaces, and one "best" play. The number of decisions in each direction determined the difficulty of each task, and thus the overall complexity of the board.

Specifically, in the enhanced UFOV portion of the experiment, the central letters were chosen based on equality of confusability (see Background: Letter Confusability). Specific letters chosen were W, A, G, U, V, S, P, and I, in uppercase Helvetica font. The relevant confusion matrix was taken from Gervais, et al., 1984. These letters were blurred using radii of 0 (no blurring), 5, 10, 15, or 20 pixels. Distractor images on the periphery were marked with an X inside a circle, while the target was marked with a + inside a circle (see Appendix III for a screenshot); this ensured that the target and distractors had identical brightness, making the visual search task an object recognition task rather than simply a brightness detection task. In both the Enhanced UFOV and Enhanced 21 Tally portions of the experiment, the BestPEST staircase algorithm(113) was used to quickly assess 75% accuracy on the given set of conditions. Finally, the number of direction changes used by the staircase algorithm was increased to 15, to ensure that participants spent a greater amount of time in the 75% accuracy range than they had during the pilot study.

The trial conditions given were the same for all subjects, and were presented in the same order for all subjects (see table 4). Likewise, the conditions for 21 Tally were identical and presented in the same order for all subjects (see table 5).

As shown in Figures 10 and 11, participants demonstrated learning through the three identical-conditions trials, for each set of conditions, and either had a plateau or continued the decrease in threshold value for the final set of identical conditions. Further, learning rates in each set of conditions seem to hold a very strong correlation (see Figure 12); as such, we can reasonably estimate the learning of each set of enhanced UFOV conditions by measuring each participant's learning in the standard UFOV. This has the added benefit of allowing us to understand how quickly these participants learn the standard test, and how they would perform under test-retest conditions.

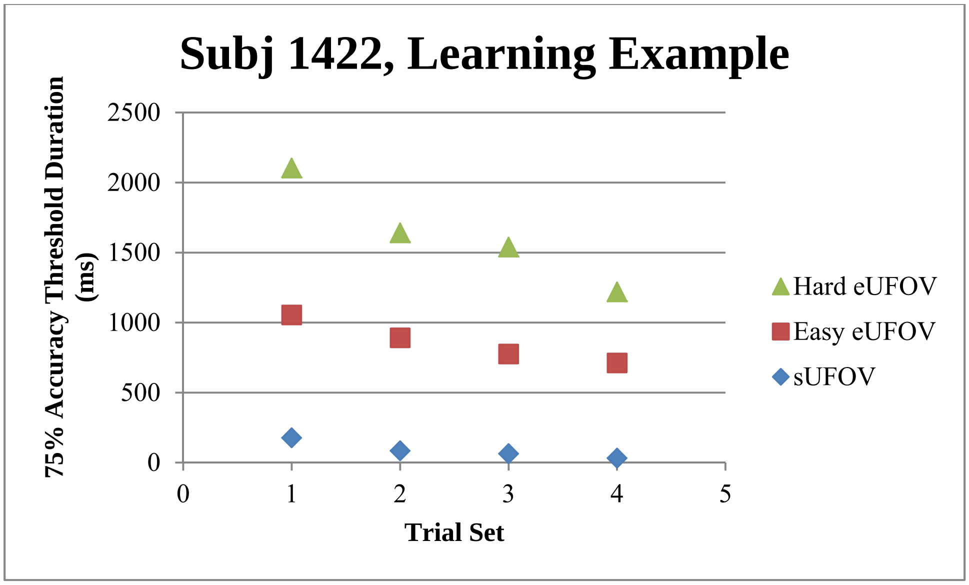


Figure 10. Trial set number and 75% accuracy duration. Note that the stimulus duration needed to achieve 75% accuracy decreased in this participant from the start of a given set of conditions to the end (three trials later), showing a learning curve. Note also that the same conditions given at the end of the experiment had either a similar or slightly slower 75% accuracy duration, suggesting that fatigue or some form of extinction is also operating within this amount of time.

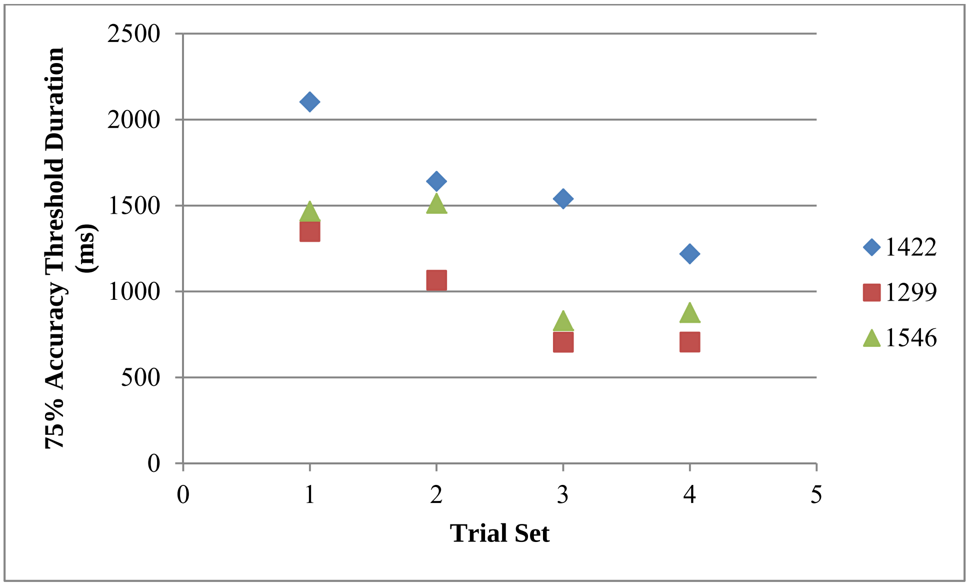


Figure 11. Trial number vs threshold duration for hard (15 pixel blur in central task, 12 peripheral locations) enhanced UFOV trials. Legend indicates the subject number for each set of points. Note that each participant has a different learning curve slope, but each is similarly decreasing (and possibly reaching a plateau due to fatigue towards the end of these trials, for subjects 1299 and 1546) .

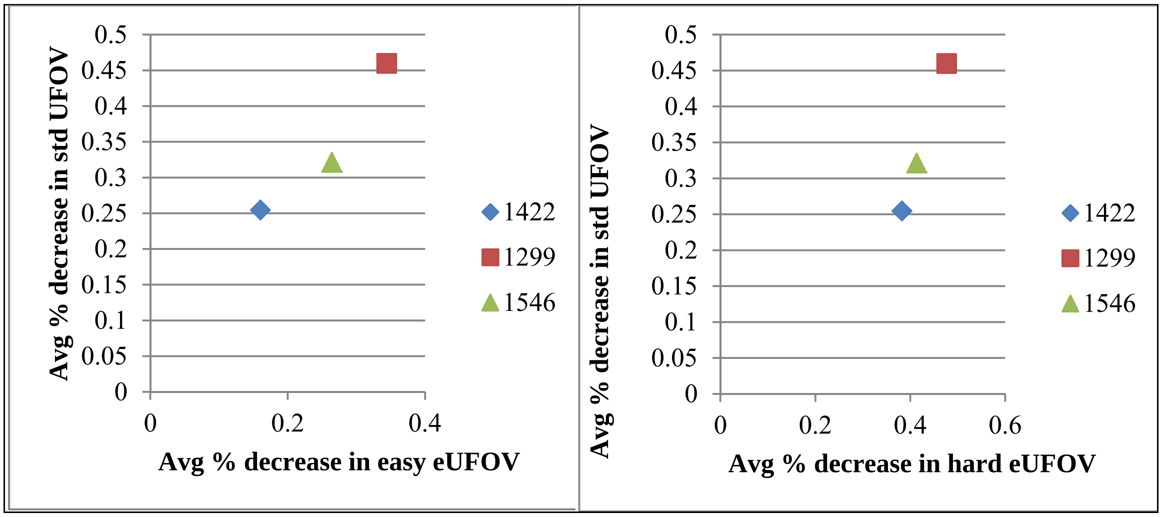


Figure 12. Comparison of subjects' Enhanced UFOV learning curve based on Standard UFOV trial performance. Each axis represents the average percent decrease in 75% accuracy threshold value from one trial set to the next; the Y axis is standard UFOV, and the X axis is enhanced UFOV. The left graph is easy enhanced UFOV (no central blur, four peripheral icons), and the right is hard (15-pixel central blur, 12 peripheral icons). Note that average decrease in threshold value seems to vary between subjects but is relatively constant within a subject. This suggests that the learning curve for the Enhanced UFOV trials can be roughly calculated based on standard UFOV trial performance.

Tally Puzzle trials showed the same trend (see Figure 13), though did not show an identical learning rate within a single participant. Therefore we cannot predict Tally learning rate using the standard UFOV learning rate; we must choose an arbitrary set of conditions to act as the "standard" Tally conditions. So as not to favor one direction over another in terms of difficulty, and to minimize the interaction on the standard Tally test, we chose the minimum equal difficulty (one decision in each direction) to act as the standard Tally condition set for learning purposes.

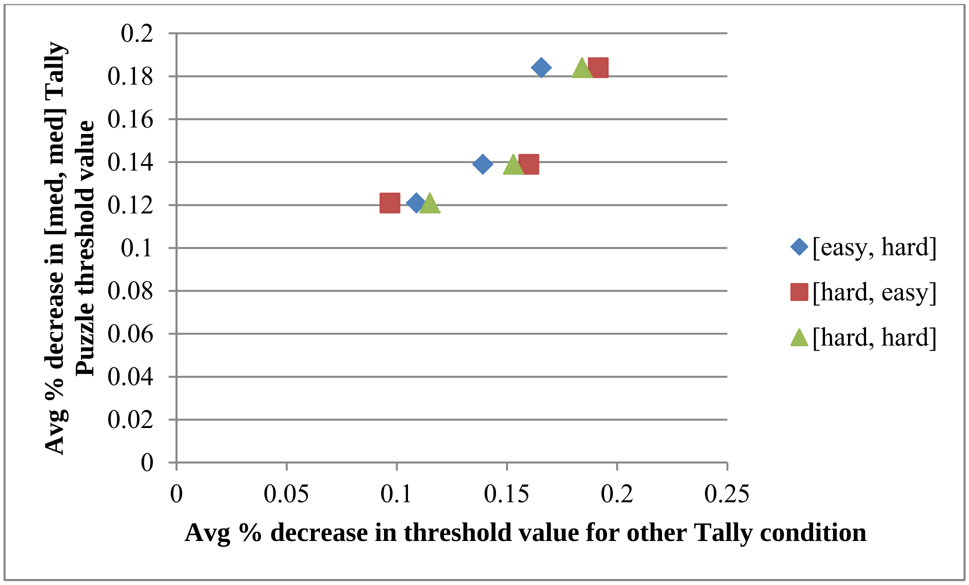


Figure 13. Comparison of subjects' Tally learning based on a standard set of Tally conditions. Each axis represents the average percent decrease in 75% accuracy threshold value from one trial set to the next trial set of identical conditions: the Y-axis is difficulty [1,1], and the X axis the other set of conditions (see legend). For example, if a participant had threshold values of 100, 90, 81, and 73 for the four [med, med] trial sets, his average percent decrease shown on the Y-axis would be 0.1 (ten percent). Since each participant has only one average percent decrease for the [med, med] set of conditions, each Y-axis value on the graph represents a different participant (i.e., the three participants in the learning experiment). The bottom (lowest on the Y-axis) set of values are from subject 1422, the middle set are 1546, and the top are 1299. Note that these are the same positions as learning in UFOV for these three participants, suggesting that relative learning rate may be constant across the two experiments. Note also that average decrease in threshold value seems to vary between subjects but is relatively constant within a subject. This suggests that the learning curve for the Tally trials can be approximated by learning on a single set of conditions, such as [med, med].

Chapter 3: Main Study

Overview

This study was conducted in order to investigate the relationship between divided attention and speed of processing in the UFOV, and then to use that knowledge to determine whether the Tally Puzzle could give similar information about divided attention. Our results support the hypotheses that the standard UFOV conflates speed of processing with divided attention, and that the enhanced UFOV, by analyzing the interaction between difficulty levels, is a more accurate indicator of a participant's ability to divide attention. Further, the Tally Puzzle shows a similar discriminant potential when we examine the significance of the interaction between the difficulty levels in the two directions.

Research Question

Can we accurately measure divided attention using naturalistic gameplay performance in a clinically meaningful way?

As has been discussed in the Background section, current clinical gold standard tests of divided attention are insufficient for capturing day-to-day variability, which is the hallmark of imminent cognitive decline. As such, our intent was to test whether a normal daily activity viable for repeated continuous unobtrusive measures (21 Tally Puzzle) can be used to predict performance on a clinical standard test of divided attention (Useful Field of View).

Specific Aims

Aim 1. Build a working model of divided attention that differentiates divided attention from speed of processing, and develop a test specifically for divided attention.

A model of divided attention, in this context, means a set of consistent theories about how simultaneously-presented visual stimuli are perceived and processed to gain information, in some combination of parallel and serial neural mechanisms. Such a model would describe natural human pre-conscious and conscious optimization strategies, which explain how players are likely to react to game boards presented under time pressure and with high cognitive load, allowing us to relate those same mechanisms to subjects taking gold-standard tests.

Aim 2. Validate the predictive power of the game to similarly measure divided attention.

The ability to predict something of clinical import, like driving ability, is essential for the utility of any clinical test. One way to show that a test is clinically valid is to show its ability to predict scores on another test which is correlated with the cognitive or physical ability in question. In this case, we used the Useful Field of View test, which has been significantly correlated with at-fault crashes (Ball, 1993; $r = 0.52$), on-road driving errors (Wood, 1995; $r = 0.55$), and driving simulator errors (Roenker, 2003; $r = 0.422$), as our clinical standard. Though the UFOV may not be the ideal test for divided attention, it certainly measures aspects of this cognitive skill, since divided attention is required for successful and safe driving. Thus, the ability to predict divided attention scores on the divided attention subtest of the UFOV, or a test developed from it that further distills divided attention measurement, would show that divided attention can indeed be measured through performance on a computer game.

Methods

Participant Selection

Participants were informed of this experimental opportunity through email (see Appendix IV for recruitment emails) and word of mouth. Note that volunteers informed by either method self-screened into a group of participants who used the computer regularly, and tended towards those who play computer games, as those people stand to gain the most from game research. Participants in the Cognitive Health Coaching study (OHSU IRB #: 7466) were targeted for recruitment of elders, as these individuals were known to be proficient in computer use and supportive of cognitive research. Other participants were recruited from the general adult population around Portland, OR.

We were able to recruit twelve people age 75+, and at least five people in each of the other age groups: 30-39, 40-49, 50-59, and 60-74. This range was by design, as this research requires as wide a range of UFOV scores as possible, and age has previously been shown to correlate to UFOV score.(62) Participant screening was by self-report only, and involved only two criteria:

1. Participants must be comfortable enough using a computer to have played at least one computer game during the last year, and
2. Participants must have no currently uncorrected vision problems.

Both of these criteria were entirely ascertained through self report, and one participant's data were excluded after admitting that he did have "trouble seeing things on the left side" though it "wasn't generally a problem."

Participants were given a $20 gift card to a local bookstore or coffee shop as thanks for their participation.

Administering the experiment

This experiment was administered similarly to the pilot study, in that participants were brought into the lab, asked to complete questionnaires, and given computerized assessments that included standard UFOV, enhanced UFOV, and the Tally puzzle. Similar to the Learning Curve study, no EEG was administered; the cost for elder participants in terms of time and discomfort was deemed larger than the added benefit of complementary EEG information. The order of trial sets was randomized (other than the standard condition sets, which were repeated at fixed intervals throughout the study), unlike the fixed order of the learning curve study.

The pre- and post-test questionnaires were altered from those used in the Pilot study, adding questions based on feedback from the Learning Curve study and removing all mention of EEG. See Appendix I for the specific questions asked.

In order to utilize the information gleaned from the Learning Curve study, we repeated the standard UFOV conditions four times throughout the UFOV portion of the experiment: once at the start, to determine the "real" standard UFOV score, then twice more at the 33% and 66% complete stages, and finally once at the end. This allows us a 75% accuracy duration for comparable conditions throughout the experiment, which lets us calculate the learning/exhaustion contribution at any given point. Similarly, we used the Tally condition of (1 row decision, 1 column decision) as the "standard" set of conditions, repeating it at set intervals throughout the Tally portion of the experiment to understand how learning and exhaustion affected Tally performance for each individual participant.

Feedback was provided to participants after every trial set (every distinct set of conditions), describing how many times they got each task correct out of the total number of presentations. In the UFOV portion of the experiment, tasks were described as "central" and "peripheral," whereas in the Tally portion they were described as "row" and "column." See Appendix III for an example screenshot of the feedback. During the feedback screen, the experiment was paused and nothing was recorded. Participants were encouraged to use this time to stretch, use the restroom, or ask questions as needed; once they were ready to proceed with the experiment, they returned their heads to approximately the same location and then pressed the "continue" button.

Snacks were made available (but were not required) for participants during each Feedback break. These snacks included Pringles chips, M&Ms, and soda (diet, regular, caffeinated and caffeine-free). Snack consumption was recorded but did not otherwise influence the administration of the experiment.

Center Type	# Centers	Central Blur Radius	Peripheral Type	# Peripherals
car and truck	2	0	Car	8
letter	8	10	+ and X	4
letter	8	10	+ and X	8
letter	8	10	+ and X	12
car and truck	2	0	Car	8
letter	8	5	+ and X	4
letter	8	5	+ and X	8
letter	8	5	+ and X	12
car and truck	2	0	Car	8
letter	8	0	+ and X	4
letter	8	0	+ and X	8
letter	8	0	+ and X	12
car and truck	2	0	Car	8

Table 6. Conditions for UFOV in the main experiment. Just as in the learning curve experiment, trial sets using standard UFOV conditions (those with "car" as the peripheral type) had no distractors in the periphery, whereas the enhanced UFOV trial sets had distractors in each non-target peripheral location. Trial sets were presented in random order, other than the standard conditions that were presented at the regular intervals seen here, in order to account for learning throughout the study.

Row Decisions	Column Decisions
1	1
2	1
1	2
1	1
0	2
2	0
1	1
0	1
1	0
1	1

Table 7. Conditions for 21 Tally in the main experiment. As in the learning curve experiment, each board contained eight cards, eight open spaces, and a single correct (best) move. Only the number of decisions in each direction (i.e., the difficulty of each task) contributed to the complexity of the board. Trial sets were presented in random order, other than the arbitrarily-chosen-standard [1,1] condition, which was repeated four times in the intervals seen here, in order to account for learning throughout the study.

When participants were recruited, they were informed that the experiment would take no more than two hours. In the event that the experiment exceeded this duration, the unfinished Tally

Puzzle trial sets were abandoned, and the participant simply completed the post-experiment questionnaire before departing.

Analyzing results

Preliminary analysis

As shown in Figure 14, the variability of 75% accuracy duration threshold values for standard UFOV tended to increase with participants' age, though some participants in each age group had threshold values below 30ms (essentially perfect). According to the UFOV User's Guide, any subtest score of longer than 500ms would cause the test to be discarded as invalid. Specifically, scores in this range for subtest 2 indicate "severe difficulty with divided attention. Subtest 3 [is] not administered because the examinee displayed severe difficulty completing Subtest 2."(114) Three of our subjects had standard UFOV values meeting this criteria. These subjects' data were not removed from the rest of our experiment, however. People fitting this description may still be playing cognitive monitoring games, and their behavior should ideally be incorporated into our models.

The errors in each task formed an interesting pattern which may speak to the relative difficulty of each task (an important consideration when analyzing interactions between difficulty levels). We observed more errors on the central task than in the peripheral task for almost all participants performing under standard UFOV conditions. This may be because the car and truck icons are difficult to distinguish from one another (there is a single line, or approximately 12 pixels, of difference between the two). The peripheral task, by contrast, is just a luminance recognition task with no distractors. On the other hand, most participants had more errors on the peripheral task than the central task in enhanced UFOV. This is likely because letters are overlearned, trivially easy to distinguish when not blurred and still fairly distinguishable even when quite blurred. The peripheral task, on the other hand, has equal-luminance distractors in all non-target locations, making it essentially an increasingly difficult search task.

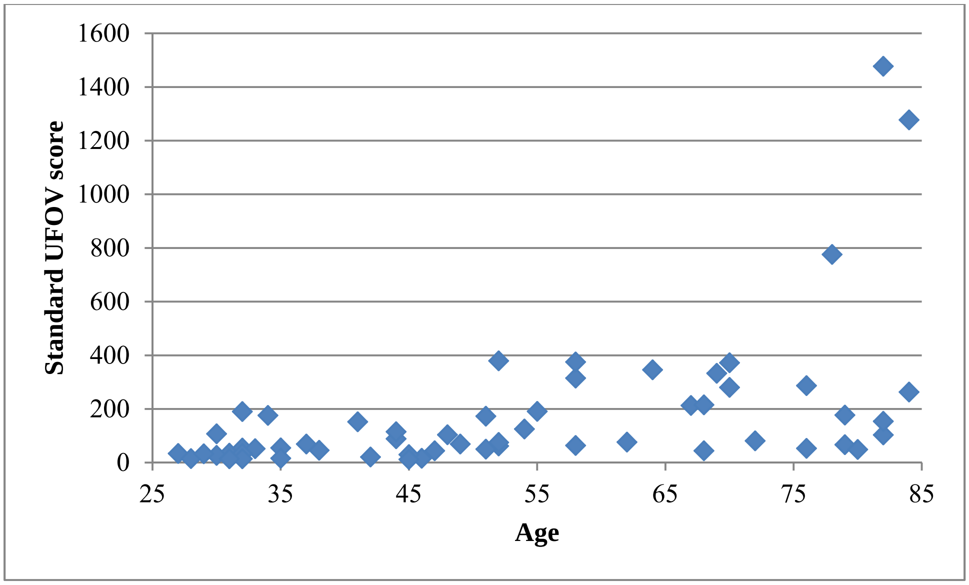


Figure 14. Age of participants and their standard UFOV scores. Note that almost all participants have "acceptable" – that is, under 500ms – UFOV scores (above that number would be thrown out of the real UFOV). Though individuals may score just as well as younger counterparts, as a whole variability increases with age.

UFOV score computation

Once a UFOV trial set has ended, based on the termination criteria of the staircase algorithm (as described in the Background UFOV section), the 75% accuracy duration score for that trial set must be calculated. In the UFOV test, this score is calculated by taking the average of the final X trial durations, where X = 15 unless the participant either performs at the bottom (15ms) or top (500ms) of the range for ten consecutive trials, in which case X = 10. Due to the nature of the staircase algorithm, the final several values will circle around the true 75% accuracy value, making this a reasonable way to calculate the value.

However, there are other ways to calculate the score, some of which are more useful to further analysis. For example, by using the results of all of the trials to maximize the log-likelihood function mentioned in the Background: Psychometric Function section, we can find the psychometric function best fitted to the data in a particular trial set. Then we can simply find the duration on the X axis where the graph passes 75% on the Y axis; this is the psychometric curve estimation of the 75% accuracy duration.

Another way of estimating the threshold value would be to take the mean or the median of a larger section of the trials than the UFOV uses. Using all of the durations in a trial set would introduce considerable noise into the calculation, as the initial trial's duration is arbitrary and the subsequent few trials may cover a great range. For example, if the true threshold were 480ms, the first few trial durations might be 500ms, 250ms, 375ms, 438ms, 469ms – all of these would inappropriately lower a mean or median calculation. However, if we use all of the durations after seeing two changes in direction of the staircase algorithm, this will give us a reasonable idea of the variance of the data without improperly weighting it higher or lower. Essentially, after two direction changes, we are likely to be much closer to the threshold value than the initial arbitrary duration, even if one of the direction changes was due to a mistake on the participant's part (e.g., guessing a too-quick stimulus correctly or missing one longer than the true threshold duration). For example, if the participant initially responded to the stimulus correctly (indicating that the threshold duration was likely less than the current duration), we would then use only trials recorded after the first mistake and the first correct response after that first mistake. Figure 15 shows the 75% accuracy duration thresholds as estimated by this method and the standard UFOV method for all participants' trial sets; the correlation between the two methods is greater than .99.

One key reason to calculate the threshold value this way is that it allows estimation of the within-condition variability (that is, the slope of the psychometric function for this trial set) by computing the variance of the individual durations about their mean. As long as we are able to minimize the effect of the starting point, which is arbitrary, this is a convenient way to estimate the slope of the probability function for this set of conditions, because for a sufficiently small step size, the variance is proportional to the slope of the underlying function in the neighborhood of the threshold point estimate. Additionally, we can analyze the variance about the threshold with ANOVA to determine the significance of each of the tasks' difficulties, as well as the interaction between those two difficulties. These measurements are not independent; however, using a repeated-measures ANOVA, the measurements need only follow a normal (Gaussian) distribution, rather than requiring independence as well. We tested each of the post-two-direction-changes sets of durations using the Shapiro-Wilk test, which has a null hypothesis that the data follow a Gaussian distribution. Out of the 1200 total trial sets (50 participants; 4 standard UFOV, 9 enhanced UFOV, 4 [med, med] Tally, 7 other Tally), 1181 could not reject the null hypothesis of normality at alpha = .05. These nineteen data sets appear to have taken more than two direction changes to reach proximity to the threshold, causing them to have large "upper" tails and removing their normal distribution (see Figure 15).

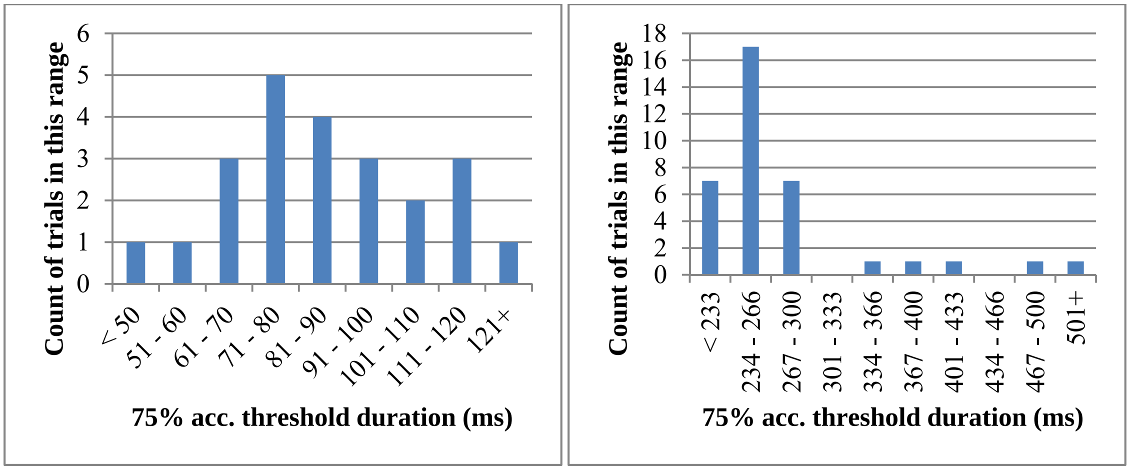


Figure 15. Comparison of trial sets with normal distribution (left) and non-normal distribution (right). Both of these trial sets come from [easy, easy] eUFOV trials; the left graph is from participant 3123, while the right is 3826. Data from the left graph have a Shapiro-Wilk p-value of 0.97, while those from the right have a p-value less than .001. The calculated threshold value for each comes from the highest-counted range (left = 73, right = 254), as expected. Note the long "upper" tail on the right graph; this suggests that the participant made more than one "mistake" when moving from the arbitrary starting value of 500ms to the actual threshold value of 254; thus, taking all data points after two direction changes included data from 300-530ms, causing the values used to be less normal (Gaussian).

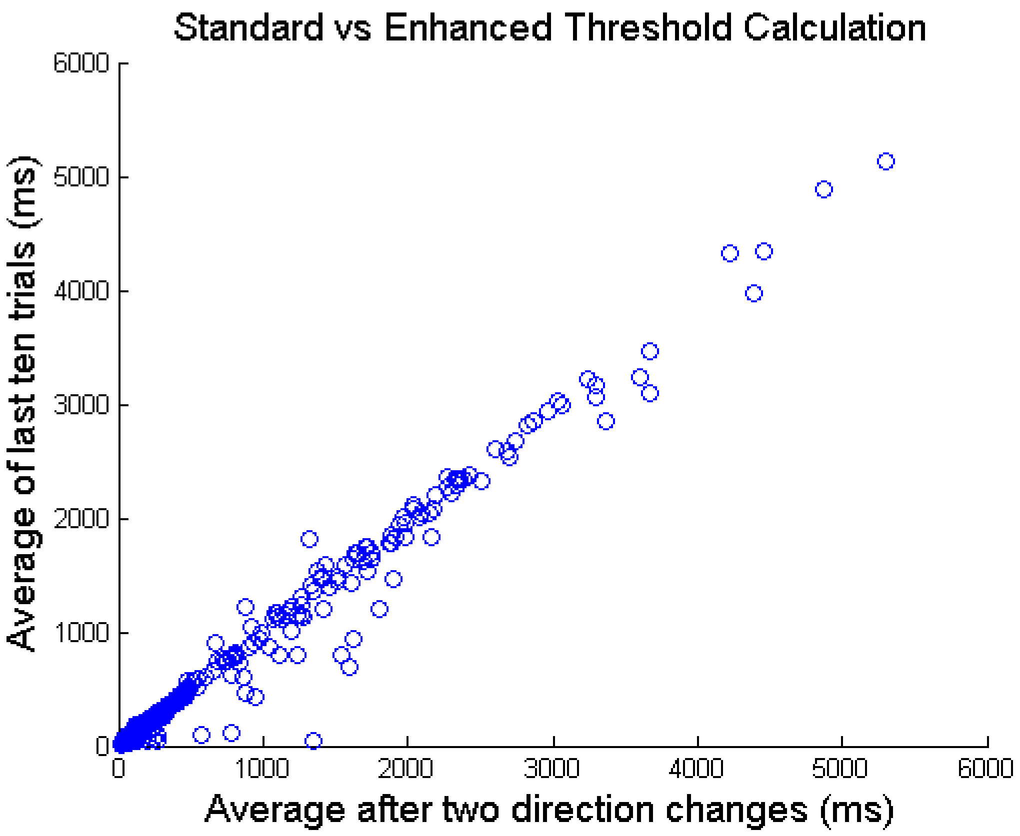


Figure 16. Comparison of participants' 75% accuracy threshold durations as calculated by standard UFOV methods and by taking all trials after two direction changes (as we did for ANOVA). In order to have an accurate picture of the variance of the performance, we calculated the threshold value as the average duration after two direction changes of the staircase algorithm. This figure shows that the vast majority of threshold values are not substantially altered by this calculation change. Pearson's correlation = 0.9934.

Results

Accounting for Learning and Exhaustion

One important difference between our experiment and the normal course of the Useful Field of View test is that in standard UFOV, participants have received verbal instructions for the general test and for subtest 1 (UFOVs1), done a trial demo of UFOVs1, taken UFOVs1, received verbal instructions for subtest 2, and done a trial demo of UFOVs2 before they begin actually taking UFOVs2. In our experiment, they only received verbal instructions and did a trial demo of subtest 1 and 2 before doing the UFOV part of the experiment. This may account for some portion of the increased variability that we saw compared to the reported variability for UFOVs2 in other experiments. However, the most likely reason is that the standard UFOV removes subjects who exceed 500ms on any subtest threshold.(48;50)

In our experiment, some people clearly understood the UFOV task more than others following the instructions. The people with a larger number of errors in the instruction/demo portion of the experiment tended to have a longer duration for the 75% accuracy standard UFOV score, and also tended to have a greater slope to the learning curve (see Figure 17). Those with few errors in the instruction / training period, by contrast, tended to have less of a learning effect and even showed some deterioration in performance which we attribute to exhaustion. This suggests that the difficulty was in understanding the task rather than in divided attention or speed of processing. This may be a contributor to high standard UFOV threshold rates in general, and is likely one of the reasons that the standard test removes data from people over 500ms.

Separation of divided attention from speed of processing in standard UFOV

One of the major problems with the UFOV is that it claims to measure divided attention, but only varies the duration of the stimulus. Without measuring the subject's performance at different difficulty levels for the different tasks, it's impossible to determine what effect changing the difficulty of one task has on performance of the other task, which is an essential aspect of divided attention.

By measuring a subject's performance on each task alone, and then varying the difficulty of each task independently, we can demonstrate the change in performance based on increasing one task's difficulty and leaving the other constant. The difference between people of high and low divided attention ability then becomes measurable as the difference between 1) the combined contributions that both tasks *would* have to the final duration if done independently, versus 2) the observed final duration score for the participant. Participants with high divided attention ability will have duration thresholds that closely approximate their predicted values if performing both tasks independently, while those with lower divided attention ability will have higher thresholds since they require more time to account for the division of attention between the two tasks. Note that this also removes the aspect of speed of processing, since it is independent of the speed in which either of the two tasks was performed.

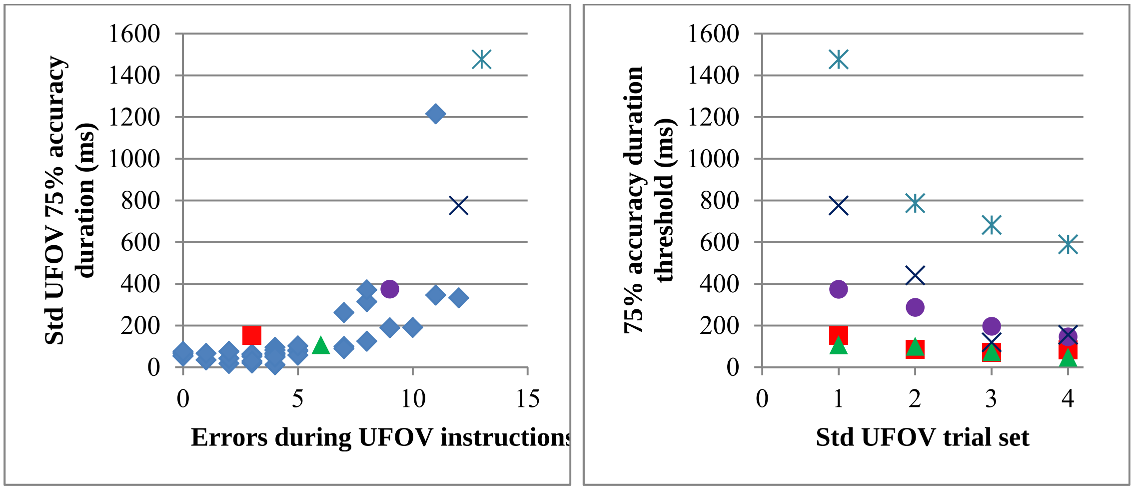


Figure 17. Understanding instructions and the learning curve. Left: standard UFOV score paired against errors made during the demo portions of the UFOV instructions; each point is a participant, and some points have been picked out for examples on the right. Right: threshold values for selected participants for each of the four trial sets with standard UFOV conditions. Note that subjects who had high numbers of errors on the instruction portion of the test generally started out with a higher 75% accuracy duration on the standard UFOV conditions (left, Pearson's coefficient = 0.71). This would correspond to a worse divided attention score on the actual UFOV test. However, many of those subjects had a sharper learning curve throughout the experiment (right), suggesting that they had more difficulty understanding the instructions rather than having lower divided attention per se.

If we consider the results of a single subject for the main experiment, we have three difficulty levels in each of the two tasks for both UFOV and Tally. These difficulty levels are physically equidistant (for example, the peripheral UFOV task has an additional 4 distractors per level of difficulty), but subjectively they may not be equal increases in difficulty; as such, they can be treated as monotonically increasing difficulty levels of Easy, Medium, and Hard. Note that it would be impossible to measure the 75% accuracy duration for the [Easy, Easy] set of conditions in the 21 Tally Puzzle, since the Easy condition involves zero decisions. As such, an

[Easy, Easy] trial set would have no decisions at all, and the participant could choose any available space – even without ever seeing the stimulus – and be correct.

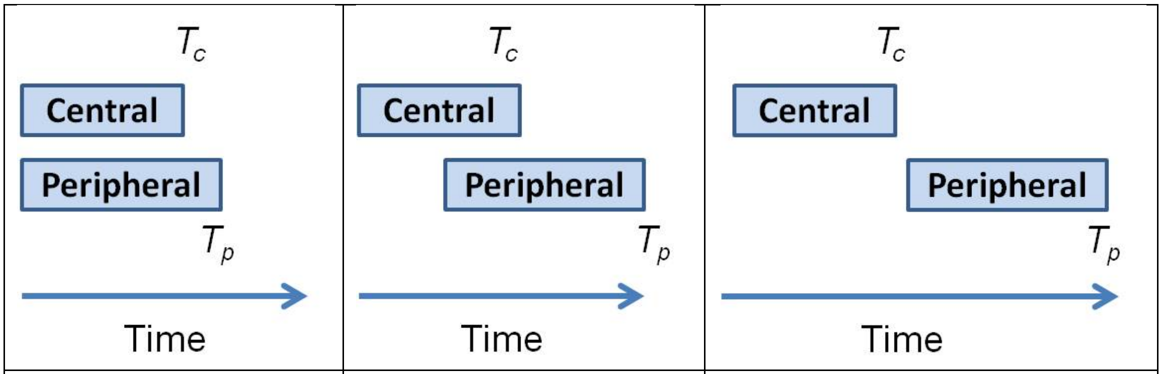
Modeling of performance

The ultimate goal of this work was to relate the observable performance measures to the underlying, unobservable attentional processes so that the observable results can be used to make inferences regarding these hidden processes. In order to make such inferences in a rigorous manner it is necessary to develop mathematical (computational) models describing the relationship between characteristics of the attentional processes and the observable quantity – in this case, 75% accuracy duration threshold, as a function of the difficulties of the tasks. The main question regarding divided attention concerns the mutual interference between the two tasks. More specifically, there are two questions:

1. To what extent are the two processes independent?
2. Are the two tasks performed in a sequential manner, or are they executed simultaneously?

Consistent with the first question, the null hypothesis is that the processes can be executed independently; that would translate into statistical independence between the difficulties of the two tasks as measured by the display duration. In particular, under the independence assumption, the effects of the task difficulties can be represented as an additive function of the two effects, without an interaction term.

The second question concerns the temporal relationship between the processes and is illustrated in Figure 18, ranging from completely simultaneous to completely sequential executions. We note that there are many versions of the sequential model (i.e., there may be several different versions of the mixed process that switches between the tasks, emphasizing one task or the other), but here we focused on the simplest extreme versions of the model represented in Figure 18-A and -C: namely, parallel and serial. This choice was based on a large volume of prior research addressing the differences between parallel and serial processes.(115-119)



A – Parallel **B – Mixed** **C – Serial**

Figure 18. Example of temporal relationships between the underlying processes determining response time in a dual task experiment. Note that tasks may be able to be completed simultaneously (left), have some components that can be completed simultaneously but other components that require less divided processing power (center), or may require focused attention on each task (right).

In the completely parallel execution model the task is completed when the longest process is finished. Thus the completion time on each trial is equal to the longer processing time, i.e., to the maximum of the two processes

$$T = \max(T_c, T_p) \tag{3}$$

where T_c and T_p are the processing times for the central and peripheral tasks, respectively. In general, to compute prediction of T from Equation (3) would require the knowledge of the distributions of T_c and T_p that would be difficult to estimate without overfitting. On the other hand, in cases when the distribution of T_c and T_p do not have significant overlap, the distribution of the maximum is the distribution of the greater component. In that case, the expected value of T is equal to the expected value of the process with a dominating distribution function. Figure 19 shows an example of the results that are consistent with the case where the peripheral task takes significantly longer than the central task.

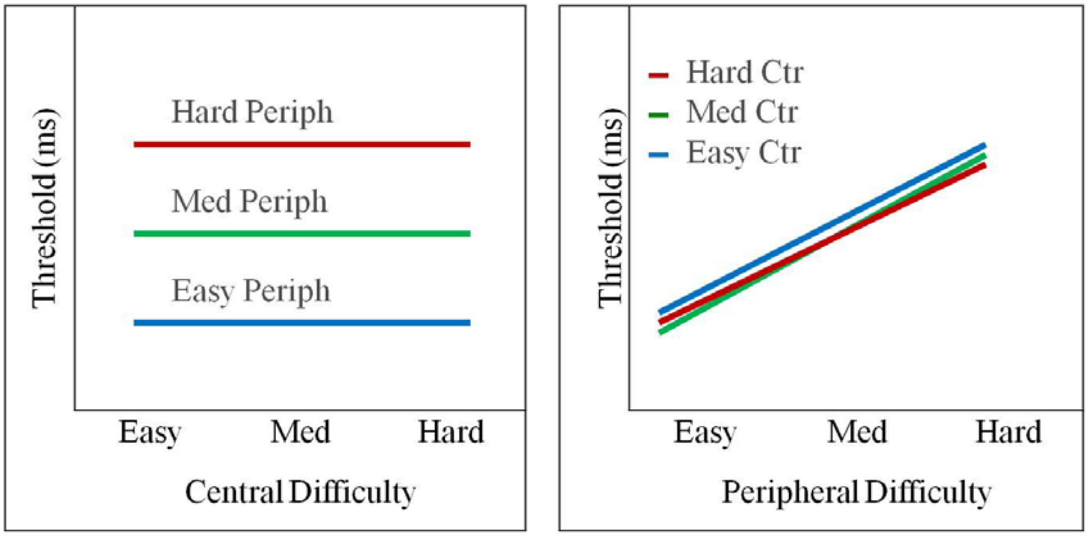


Figure 19. Ideal examples of performance produced by a parallel underlying process. Each line represents a single level (easy = blue, medium = green, hard = red) of one task, while the other task increases from left to right. The left graph holds the peripheral difficulty constant (per line) and varies the central difficulty, while the right holds the central task constant (per line) and varies the peripheral difficulty. Note that only changing the peripheral difficulty has an effect on the performance outcome; that is, a one-way ANOVA performed on data varying the central difficulty (left) show no difference in mean value ($p > .05$), while that performed on data varying the peripheral difficulty (right) show a significant difference between each mean ($p < .05$).

In a completely sequential (serial) dual-task execution, the time to successfully complete the composite task is given by the sum of the times to complete the individual tasks. In that case, the duration at a given execution level would be approximately an additive combination of the times to complete the individual tasks, i.e.,

$$T = T_c + T_p \tag{4}$$

Note that in both cases the times to complete the individual tasks are, in the most general case, functions of both task difficulties, but because of our assumption of independence we can write $T_c(d_c)$ and $T_p(d_p)$. The ideal additive representation is shown in Figure 20.

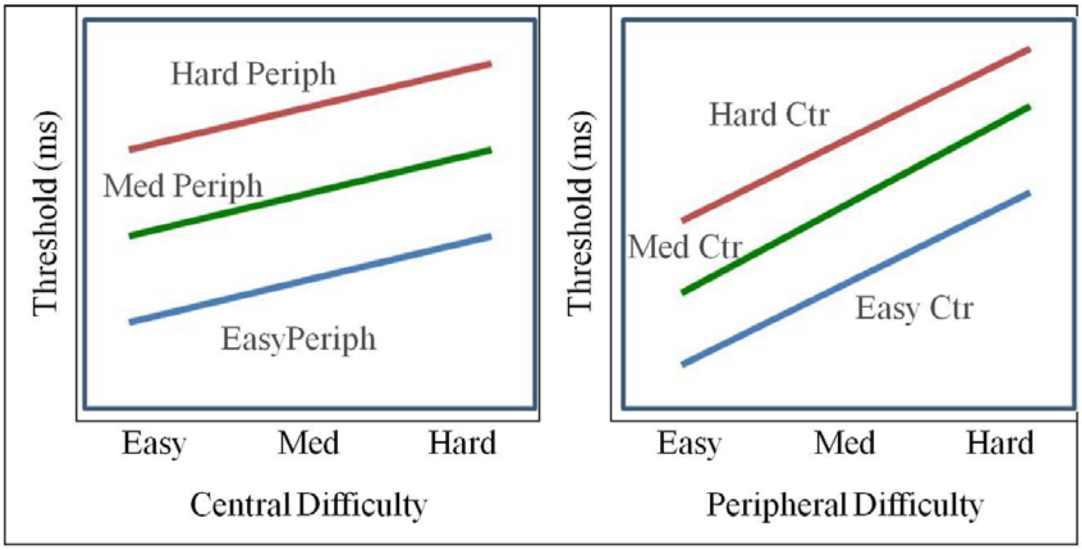


Figure 20. Ideal examples of performance produced by a serial underlying process. Each line represents a single level (easy = blue, medium = green, hard = red) of one task, while the other task increases from left to right. The left graph holds the peripheral difficulty constant (per line) and varies the central difficulty, while the right holds the central task constant (per line) and varies the peripheral difficulty. Note that changing the difficulty of either task affects the observed behavior; a one-way ANOVA performed on either grouping of data will show a significance ($p < .05$) difference between the mean values when either task difficulty is increased.

In order to test these two special cases (parallel execution with non-overlapping distribution, and serial execution), we can use analysis of variance. In the parallel case, we test the hypothesis that there is no main effect of one of the task difficulties. In the serial case, we test the same hypothesis regarding both main effects.

It should be noted that not all subjects need have the same underlying process, and in fact we see examples of both serial and parallel processes in our data. Figures 21 and 22 show examples of each of these types of processes from our data. Figure 21 shows an example of a participant with an underlying serial process during the enhanced UFOV portion of the experiment. Note that the boxes for the easy conditions overlap only minimally, and the 1-way ANOVA performed on these conditions strongly suggests that at least one of the medians is truly different from the others, in both directions (i.e., increasing the difficulty in either task does increase the threshold value). Figure 22 shows an example of a participant with a parallel process, with the peripheral task taking significantly longer than the central task. Note that when the peripheral remains easy (indicating a minimum of interference with the central task), increasing the central task difficulty produces no increase in threshold duration. This is confirmed by one-way ANOVA, which shows no statistically significant difference between the mean durations (p > .05).

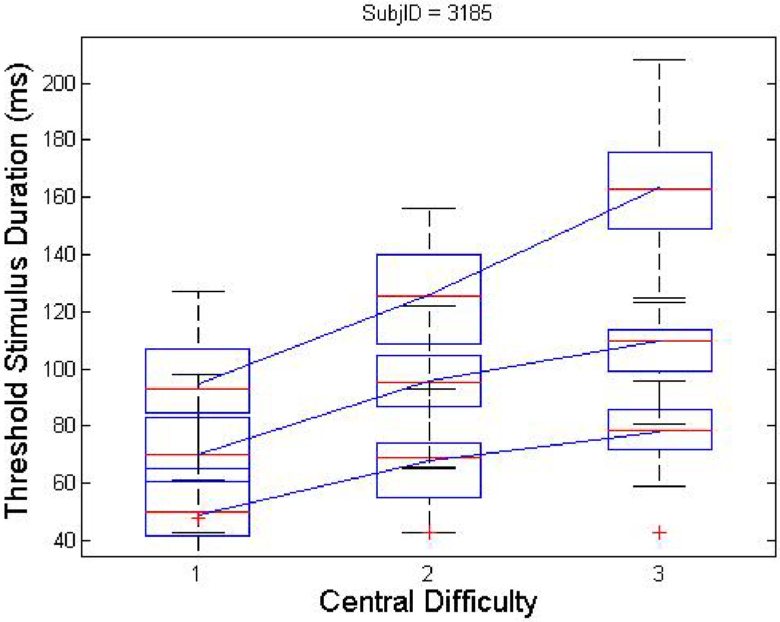


Figure 21. UFOV serial process example. Each point (center of the box) is the 75% accuracy threshold value for an enhanced UFOV trial set with particular conditions: central difficulty is easy on the left and hard on the right, while peripheral difficulty is easy for the bottom and hard for the top. Box plots represent the 25-75 quartiles, with the whiskers showing the full range of stimulus durations used in the ANOVA and + representing outliers. As calculated by one-way ANOVA, none of the means in any one difficulty set are the same (e.g., P value for easy peripheral task = 1.2e^-17, and for easy central task = 8.5e^-9). This indicates that both task difficulties are affecting performance in each condition. In other words, the participant is performing these tasks serially.

It is of note that most participants appear to fit the parallel model on the enhanced UFOV, while the serial model fits most of the performances on the Tally Puzzle better (see Table 8). This may be due to the increased complexity inherent in the Tally Puzzle itself, forcing most participants to perform a serial search in order to pick the best place to play a card. Alternatively, it may indicate that the central task difficulty increase (blur) was insufficient to elicit behavior change when the peripheral task was easy; this could be tested by increasing the delta of the blur difficulty (from five to ten pixels, for instance).

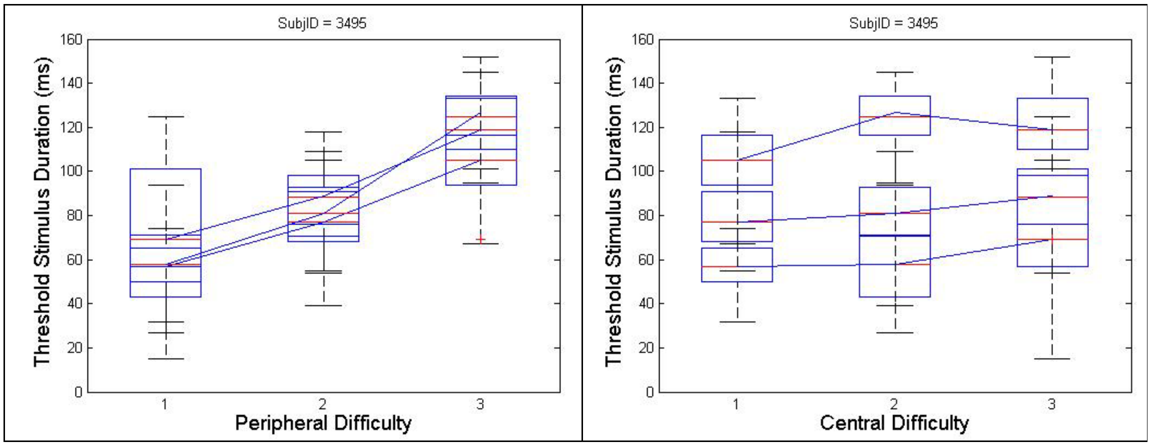


Figure 22. UFOV parallel process example. Each point (center of the box) is the 75% accuracy threshold value for an enhanced UFOV trial set with particular conditions: for each graph, central difficulty is easy on the left and hard on the right, while peripheral difficulty is easy for the bottom and hard for the top. Box plots represent the 25-75 quartiles, with the whiskers showing the full range of stimulus durations used in the ANOVA and + representing outliers. Both graphs are showing the same nine points, but are displaying them differently: the left graph shows the change in performance with the increase in peripheral difficulty, while the right shows the change as central difficulty increases. As calculated by one-way ANOVA, the means when the central task is held at easy (left graph, bottom line) are different (P = .00018), while those when the peripheral task is held at easy (right graph, bottom line) are not significantly different (P = 0.289). This indicates that only the peripheral difficulty is affecting performance when interaction between the tasks is kept to a minimum. In other words, the participant is performing both tasks in parallel, and only the slower task controls the observed threshold duration.

	Serial	Parallel	Significant Interaction	No Significant Interaction
Enhanced UFOV	2	48	45	5
Tally Puzzle	45	5	48	2

Table 8. Process model fitting of observed participant behavior. These are counts of participants whose behavior is best fits the particular model represented by the cell in question. Note that most participants fit a parallel model better for eUFOV, while most fit a serial model for Tally Puzzle. For the majority of participants in both parts of the experiment, an interaction term significantly improved the model fit.

These two models fit the observed behavior well when there is minimal interference between the two tasks (i.e., when one of the tasks has easy difficulty). However, as the two tasks both become more difficult, most people show evidence of some interference of one upon the other – that is, when one task's difficulty is increased, it makes the *other* task take longer as well. This is the cost of having insufficient divided attention ability, and as such it can be used as a proxy measure for divided attention. In terms of the models, however, it means that an additional term is required to represent the interaction between the two difficulties. Thus, the serial model becomes:

$$T = T_c + T_p + (T_c * T_p) \tag{3}$$

wherein the multiplication of the two difficulties is the interaction term. This can likewise be added to the parallel model, as in:

$$T = \max(T_c , T_p) + (T_c * T_p) \tag{4}$$

Table 8 shows that the vast majority of participants have a significant (P < .05) interaction between the two difficulties, in both enhanced UFOV and Tally Puzzle. Figures 23 and 23 show example participants in each section of the experiment who show either a large interaction or no significant interaction. Note that most subjects show a relatively shallow threshold duration increase with the variable task difficulty increase when the invariate task is easy; however, when the invariate task is difficult, most subjects show a much larger increase in threshold duration. This shows that the difficulty of one task is interfering with performance on the other task. Subjects with less interference are better able to divide their attention between the two tasks, and have a more linear/additive difficulty increase, appropriate to the difficulty of doing each of two difficult tasks independently. Subjects with more interference have worse divided attention ability, meaning that the increased difficulty of the second task severely impacts their ability to perform on the first task. This difference is the primary way we calculate divided attention, and the most important difference between the Standard and Enhanced UFOV.

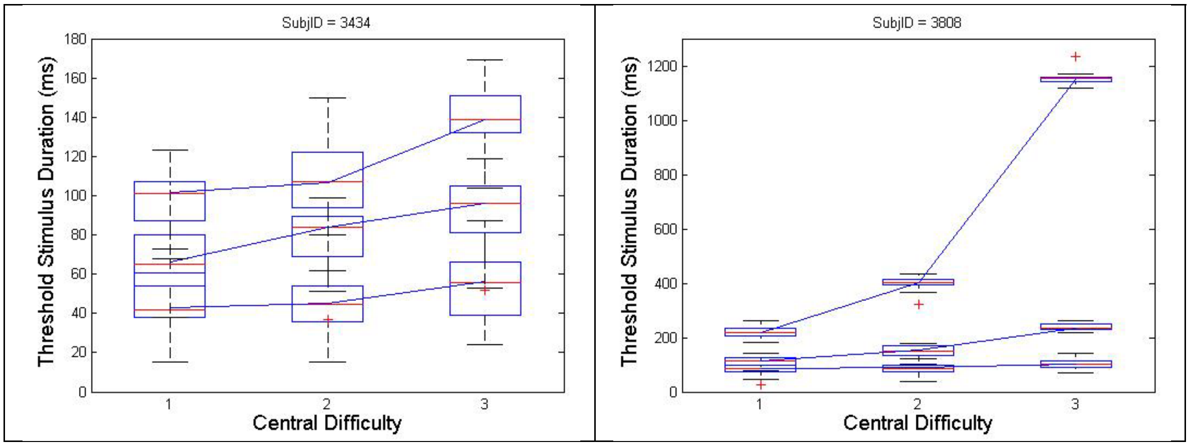


Figure 23. Example UFOV results, showing no interaction (left) and a large interaction (right). Each point (center of the box) is the 75% threshold value for a trial set with particular conditions: central difficulty is easy on the left and hard on the right, while peripheral difficulty is easy for the bottom and hard for the top. Box plots represent the 25-75 quartiles, with the whiskers showing the full range of stimulus durations used in the ANOVA and + representing outliers. P value for significance of interaction term on the left = 0.41, and on the right = 2.8e^-234.

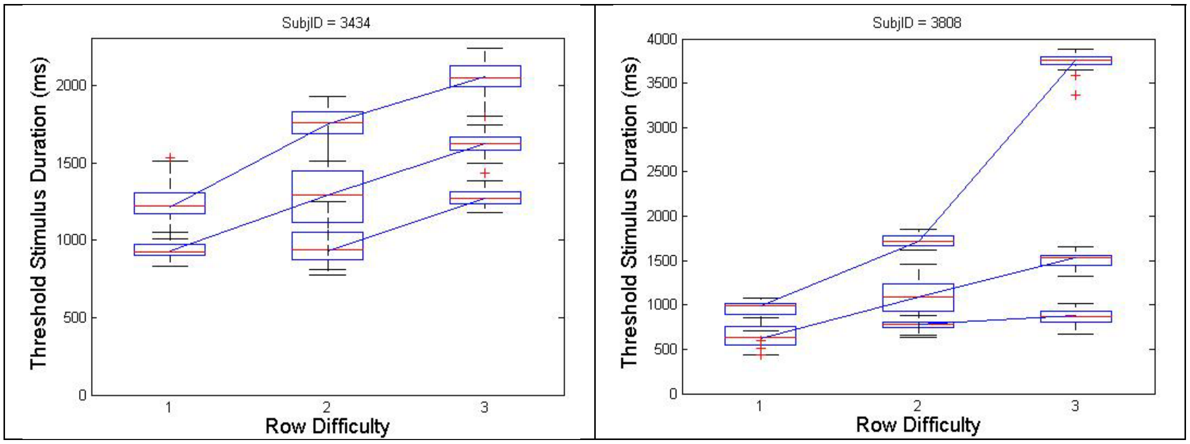


Figure 24. Example Tally results, showing no interaction (left) and a large interaction (right). These are the same subjects as in the Figure 23. Again, each point (center of the box) is the 75% threshold value for a trial set with particular conditions: central difficulty is easy on the left and hard on the right, while peripheral difficulty is easy for the bottom and hard for the top. Note that the [easy, easy] condition for Tally Puzzle (bottom left) was not measured, as it violated the single-correct-answer constraint of our experimental board construction. Box plots represent the 25-75 quartiles, with the whiskers showing the full range of stimulus durations used in the ANOVA and + representing outliers. The p-values for significance of the interaction terms calculated by two-way ANOVA are 0.15 (left) and 5.9e^-132 (right).

Another, very simple way to envision the interaction between the difficulty levels is simply to take the difference between the observed threshold duration at the hardest difficulty level and what we might expect that value to be, given performance on the easy difficulty levels. With three different levels for each of two tasks (in each of two different tests), this gives us two 3x3 matrices of performance values (see Table 9). Each value in this table represents a 75% accuracy duration for the intersecting set of conditions.

Subject		21 Tally Puzzle				eUFOV	
3808	easyCol	medCol	hardCol		easyPrph	medPrph	hardPrph
easyRow	-	766	805	easyCtr	86	99	84
medRow	612	996	1476	medCtr	50	91	107
hardRow	978	1712	3783	hardCtr	211	402	1151

Table 9. Observed values (in milliseconds) of the 75% accuracy duration for the stimulus presented in each of the conditions of the main experiment.

If we were to take these data and map them onto a surface, with the difficulty of each task in the X and Y dimension, and the threshold value in the Z dimension, the easy row and column would form five points (or four, in the case of Tally) that describe a plane. If we then extend that plane out to the difficulty levels of [hard, hard], we can visualize the difference between the observed value and what might be predicted if these difficulties were strictly additive. Figures 25 and 25 show examples of this visualization for participants with large and small differences between observes and predicted values, respectively. Note that the Tally data is on a much larger scale than the UFOV data; nevertheless, some similarities seem to exist between the two.

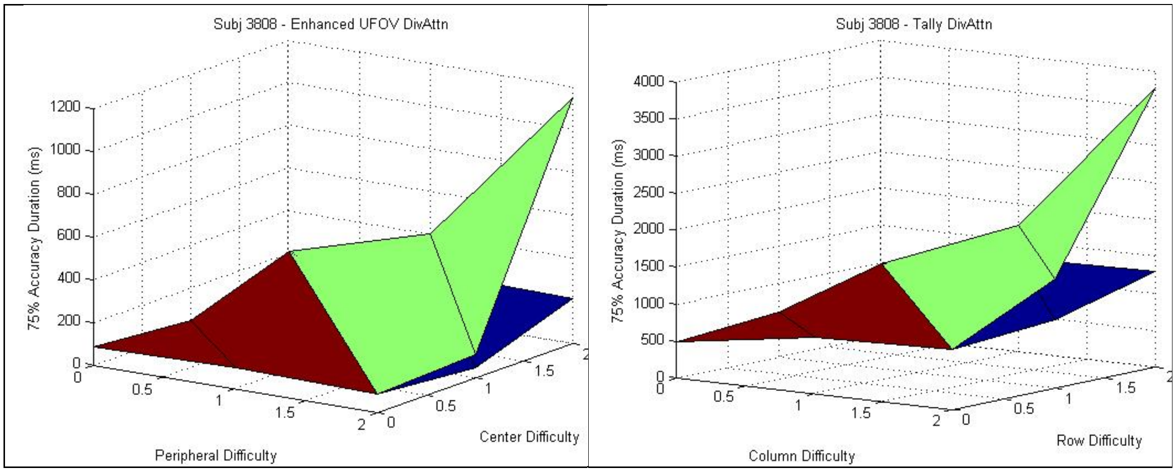


Figure 25. Surface visualization of a large interaction. By extending the easy difficulty threshold durations into a plane, one can visualize the contrast between the observed valued and what an observation with no interaction would look like.

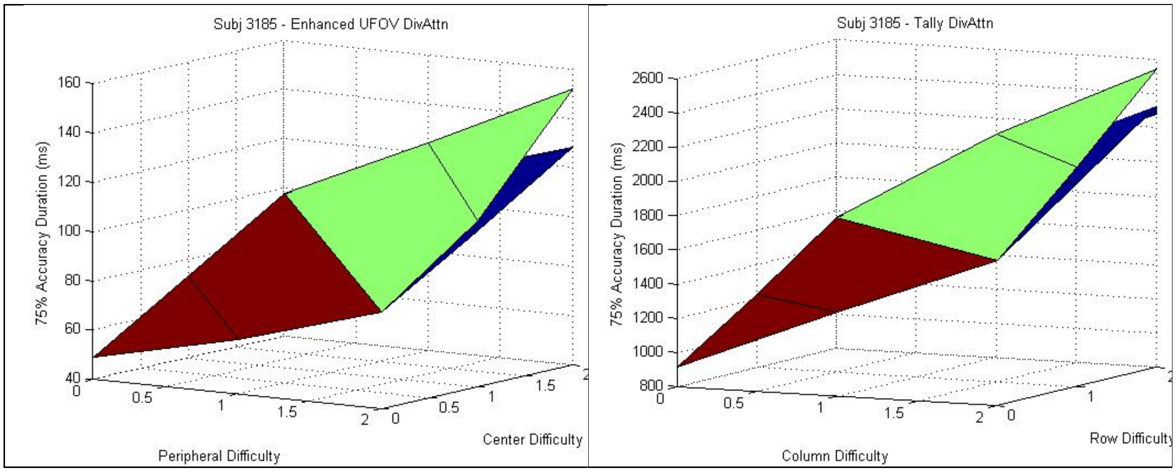


Figure 26. Surface visualization of a small interaction. In the same way, one can visualize a non-significant (P > .05) interaction as being much closer to the theoretical plane generated by threshold values observed during easy conditions.

Taking this simple method of envisioning the interaction between the difficulty levels, we can also show its relationship to the age of the participant (see Figures 27 and 28). As expected, age does not entirely determine the divided attention score, but in general divided attention does decline with age. Perhaps more importantly, the range of divided attention scores increases dramatically with age. That is, those with the best divided attention ability are relatively similar between the youngest and oldest age groups; those with the worst divided attention ability are increasingly different as the age gap increases.

Measurement of divided attention skill from Tally performance

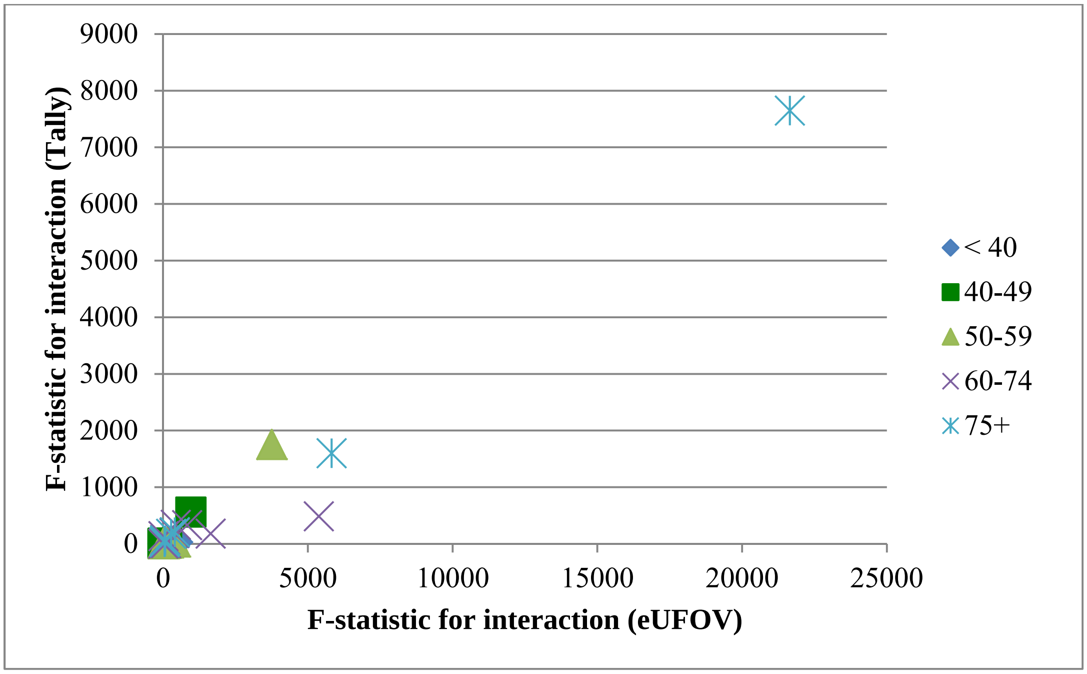


Figure 27. F-statistic of the interaction value as calculated by Tally Puzzle versus enhanced UFOV, split into age groups. As these interaction values were calculated through repeated-measures ANOVA, the F-statistic represents the variability within the data explained by the interaction, or the magnitude of the interaction. Since the magnitude of the interaction is a measure of a participant's inability to successfully deal with both tasks as the difficulty level increases, this is a reasonable proxy for divided attention ability, with a higher F-statistic indicating a lower divided attention. Note that variability in divided attention certainly seems to increase with age, as expected. Though some members of the oldest age group do perform similarly to younger age groups, the range increases dramatically with age.

In order to take advantage of the unobtrusive monitoring opportunities for the elderly presented by already having 21 Tally in place as part of a cognitive health coaching platform, we need to assess whether the computer game measures the same internal processes as the eUFOV. A correlation between the Tally Puzzle and the eUFOV results would provide evidence that these two tasks are measuring the same cognitive skill. Given that the eUFOV was specifically created to highlight the best measurements of divided attention from the standard test, this result supports our hypothesis that the 21 Tally game measures divided attention.

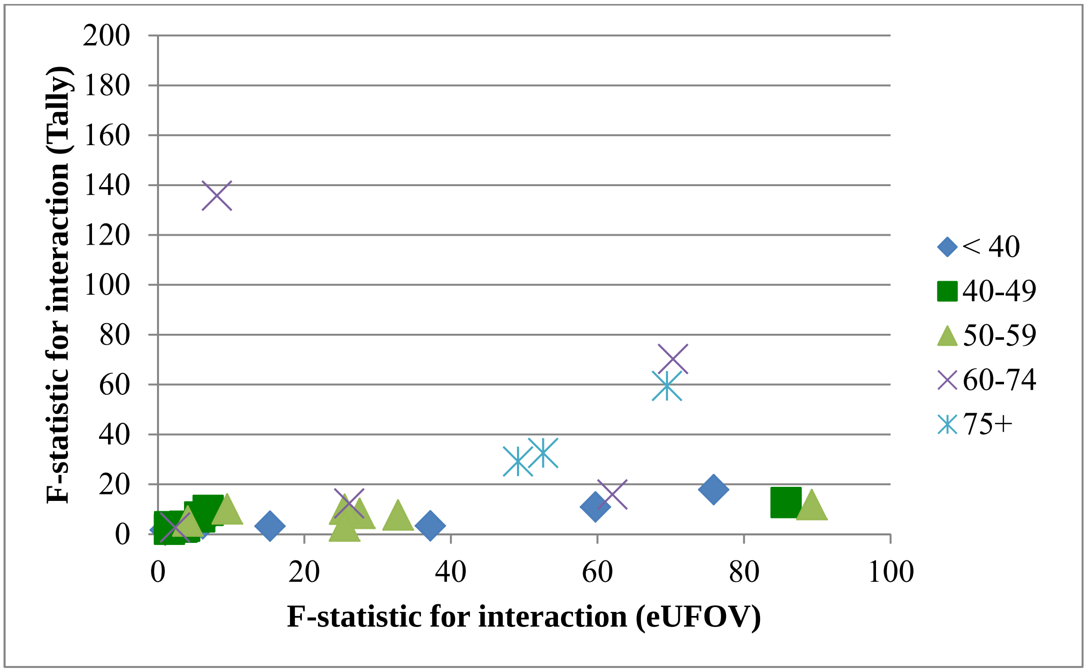


Figure 28. F-statistic of the interaction value as calculated by Tally Puzzle versus enhanced UFOV, split into age groups, detail view. Here we can see that the youngest age groups cluster toward the lowest F-statistic (i.e., the best divided attention). This is as expected given the link between age and declining cognitive ability.

Having a more accurate measure of divided attention has theoretical importance, but without incorporating it into a relatively fun computerized task, it remains unsuitable for unobtrusive monitoring, as it would not fit into a normal daily or weekly routine. Thus, in order to relate the divided attention measurements on eUFOV to those from the Tally Puzzle, we must compare the interaction terms. Figures 23 and 24 show the performance of the same two subjects in eUFOV and Tally. Note that the subject on the left has essentially parallel lines, indicating a lack of interaction between the two tasks, on both eUFOV and Tally. The subject on the right, by contrast, shows an increasing amount of interference as both tasks' difficulties grow, culminating in a sharp increase in the [hard, hard] condition representing the maximum interaction measured during this experiment.

The other subjects tested in this experiment showed a similar correlation between the amount of interaction between difficulties on the eUFOV and on the Tally Puzzle. Specifically, the Pearson's correlation between the p-values of the interaction terms for Tally and eUFOV is 0.82 (see Figure 29). This suggests that while the Tally Puzzle is undoubtedly engaging significantly more complex cognitive processes than the eUFOV (symbol manipulation, basic addition, spatial reasoning, etc.), the threshold value obtained can be utilized in a behavioral model of cognition to estimate divided attention ability.

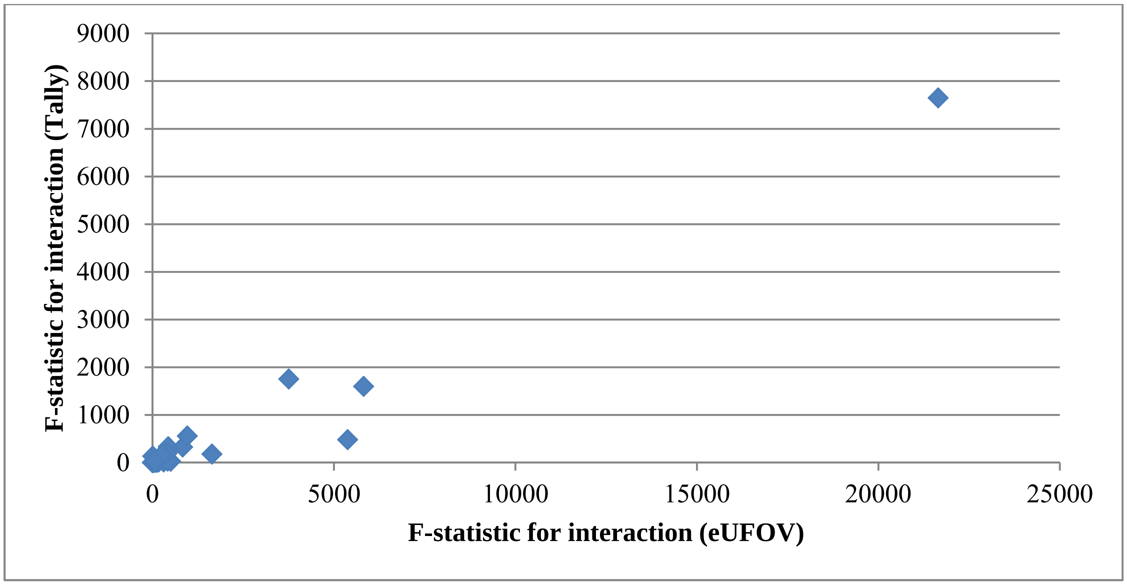


Figure 29. Correlating enhanced UFOV interaction term F-statistics to Tally Puzzle interaction term significance-statistics. This figure shows the F-statistic for the significance of the interaction term; each point is a single participant. Pearson's correlation = 0.98. This correlation is highly influenced by the single point above 20,000 for eUFOV, which may be an outlier; however, without this point, the correlation between the two ways of measuring divided attention (i.e., interaction) retain a correlation value of 0.81, which remains relatively high for a cognitive metric.

Qualitative measures and self-report

Three questionnaires were given during the course of the experiment, containing a total of one numeric, eighteen coded / multiple choice, and five free-text answers. Additionally, notes were taken by the researcher during the experiment regarding the nature and timing of unique behaviors such as questions, restroom and snack breaks, and conversation or exclamations. These data were not fully analyzed and may contain further useful information; however, as a first-pass attempt at understanding how the participants' background affected their performance, several factors were compared to both the standard UFOV score and calculated divided attention score to determine if any correlations existed:

- Number and type of game played regularly by the participant
- Categorization of participant as playing games traditionally thought of as divided-attention intensive (first-person shooters, driving simulators) or not
- Personal rating of / perception of divided attention ability
- Personal rating of / perception of driving ability.

Only one type of game listed in our questionnaires had a significant correlation with standard UFOV threshold value. Figure 30 shows the relationship between people who reported enjoying first-person shooters and the standard UFOV score. As expected, these participants generally had a shorter threshold duration than other participants. However, it should be noted that all of these participants were under 50 years old; therefore, age also plays a role in this contrast. We hypothesize that participants who enjoy driving simulation games would also score well on sUFOV; however, only one participant reported enjoying this type of game, so no reasonable analysis could be undertaken.

No other self-report factors that were investigated were significantly correlated with UFOV score. This is somewhat interesting, in that neither perceived driving ability nor perceived divided attention ability were strongly correlated with UFOV threshold. However, this could be related to the fact that elders have a strong motivation to rate themselves as competent drivers, so as to ensure that facet of their independence. Only two out of the twelve participants over age 75 rated themselves as anything less than a good driver, and those two noted that they no longer drive.

We would have liked to investigate relationships between first language and preferred direction in Tally Puzzle (for example: does English predispose towards rows while Mandarin predisposes toward columns?), but all participants reported English as a first language. Likewise, we were also interested to know whether and correlation existed between UFOV scores for those having taken the test before, which would provide a measure of reliability or long-term learning. Unfortunately, only two participants had taken any variant of the test before.

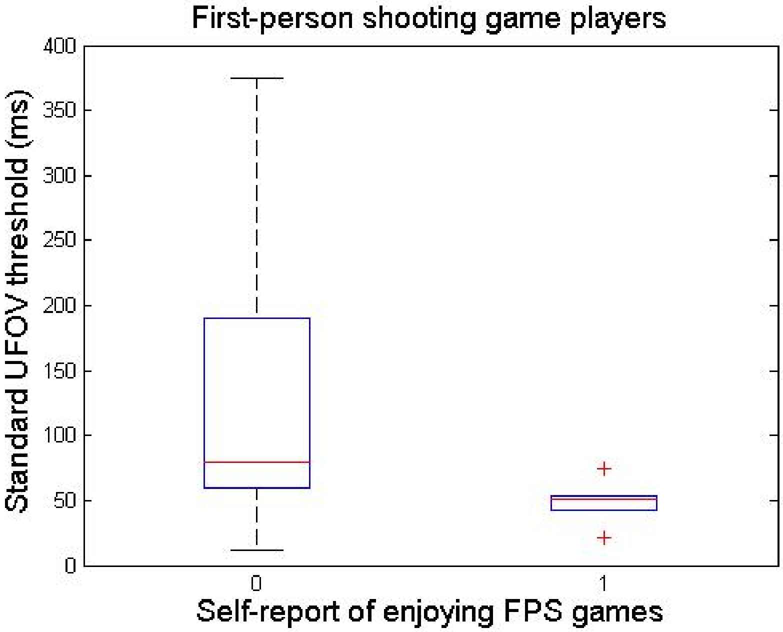


Figure 30. Participants who reported playing and enjoying "first-person shooter" (FPS) games have shorter standard UFOV threshold durations, on average. On the X-axis, 0 indicates reporting "no" when asked about playing FPS games, while 1 indicates "yes." A two-tailed t-test of unequal variance gives a value of .00378, indicating that these two groups are significantly different. This result is consistent with previous work by Bavelier, et al., who suggest that playing such games may in fact increase divided attention ability. Note, however, that only 6 participants reported enjoying FPS games, and the UFOV range of participants reporting "no" encompassed the entire range of those who reported "yes."

One final note of interest is the relationship between age and enjoyment of the game. Participants over age 75 reported that they would play it much more frequently than any other group. We hypothesize that this could be either due to a desire to please the researcher, or due to cognitive health being a greater concern for people of that age group.

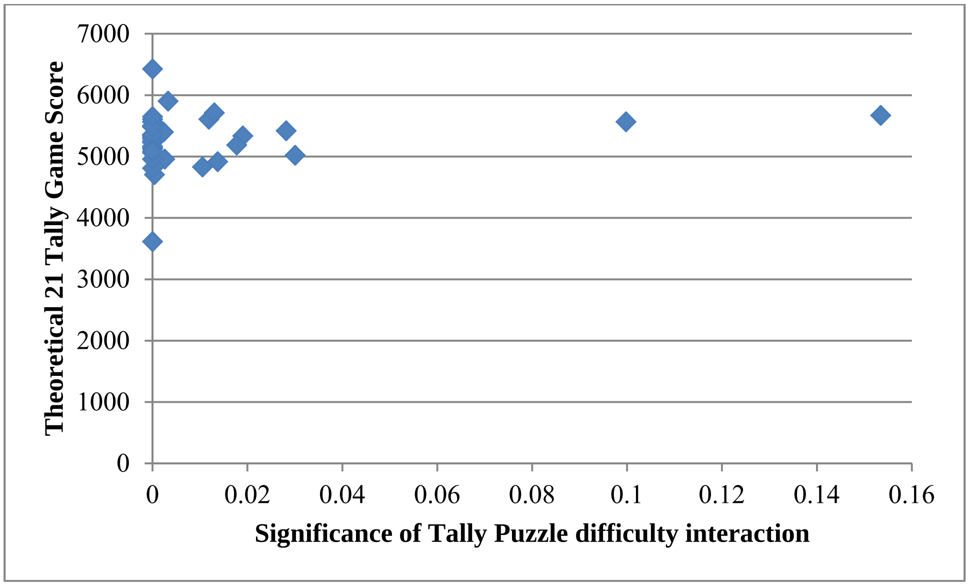


Figure 31. Performing well on the 21 Tally game does not correlate with high divided attention score. Score in 21 Tally is calculated simply as +21 points for getting 21 in either direction, and -21 points for getting a bust in either direction. Note that some participants have an insignificant interaction term (indicating very good divided attention) without having a high score on the game. This is due to the fact that divided attention score is calculated using metrics that compare a player to his or her own baseline performance, rather than to some standard of performance across players. Thus, actual game skill is removed from the diagnostic aspect, which is important in the development of a useful cognitive test.

Finally, Figure 31 shows the relationship between divided attention and the score that participants would have received in the 21 Tally game for their busts and 21s over the course of the experiment. There is no significant correlation between the two, which is important in a game to be used for unobtrusive monitoring. This separation allows the game score to be what it is intended to be: an incentive for the player to continue playing the game, rather than being obviously correlated to a measure of cognitive decline.

Chapter 4: Discussion

Summary of Major Findings

One key finding is that the enhanced UFOV not only relates as well to the "divided attention" subtest of standard UFOV as that test does to itself, but also provides a framework for investigating divided attention ability apart from confounding speed of processing differences. Our hypothesis that divided attention can be measured through analysis of interactions on a dual task seems to be supported by this work, and helps to further our understanding of the way divided attention works in the brain. Another important contribution of this work lies in its relationship to the Tally Puzzle, and through that, the 21 Tally casual computer game. The clinical implications of having a self-reinforcing (i.e., fun) set of tasks to complete on the computer rather than a test battery in a neuropsychologist's office are vast, both for screening and remediation purposes. Current clinical tests are not designed to diagnose cognitive decline at its outset; they gather no longitudinal data, have no way of monitoring cognitive variability, and have scores that are compared to national averages rather than relating to the individual taking them. Unobtrusively received data taken from a game like 21 Tally corrects all of these deficits, allowing patients the opportunity for earlier and more effective interventions, as well as helping them maintain their independence rather than waiting until impairment is severe enough to warrant loss of independence and increased levels of care. Additionally, people utilizing unobtrusive monitoring can serve as their own baselines, both within a single test and over the long term / multiple tests.

Aim 1: Separating divided attention from speed of processing in UFOV

Scatter plots of the 75% accuracy threshold durations for the UFOV tests are shown in Figure 32. The range for any subtest of the standard UFOV is from 16ms to 500ms.(114) As can be seen on the abscissa in Figure 32, our participants span much of that range; this is as expected with our large age range, and means that (according to our best standard test) our participants have measurably different divided attention abilities. Further, the Pearson's correlation between the threshold durations for the first and second standard UFOV trial sets is 0.82. This is similar to previous studies' findings on test-retest reliability for the UFOV divided attention subtest among young, healthy participants.(62) The correlation of the standard UFOV to the easy trial set of the enhanced UFOV is even higher (0.91), which is strong evidence that these two tests are measuring the same cognitive process. As the easy conditions in eUFOV represent the least amount of interference between the two tests, this suggests that the standard UFOV is largely measuring speed of processing, and that the enhanced UFOV distills the parts of the standard test most relevant to divided attention. Additionally, the eUFOV can reasonably approximate the standard UFOV by using the threshold from the easy conditions.

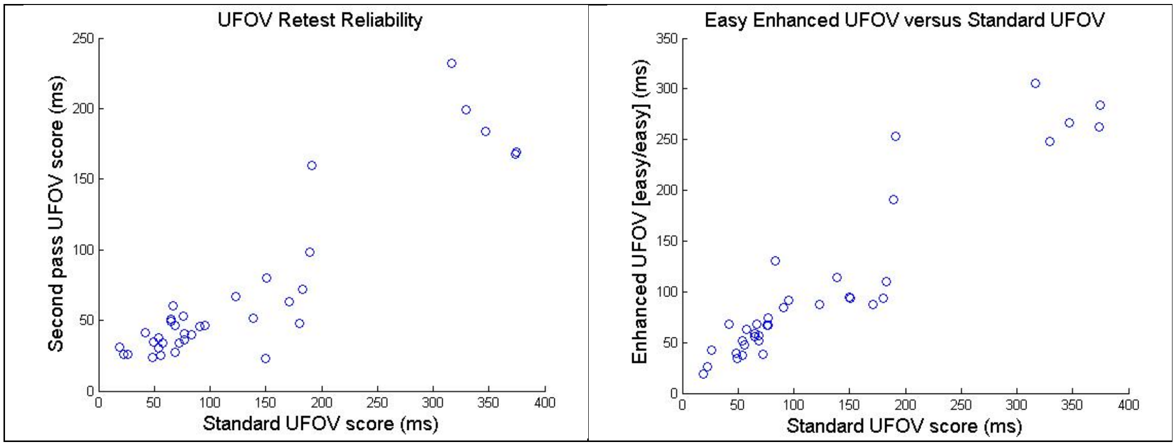


Figure 32. 75% accuracy thresholds on standard UFOV subtest 2 trial sets, compared to a second run of standard UFOVs2 and easy enhanced UFOV trial sets. Each point is a single subject. Left: test-retest reliability run of UFOVs2 (Pearson's r = 0.82). Right: first-run UFOVs2 thresholds compared to easy (no blur in central letters, four peripheral locations) conditions in enhanced UFOV (Pearson's r = 0.93). These comparably high correlation numbers suggest that these tests are measuring the same thing.

Table 11 shows the correlations between standard UFOV threshold values and each of the enhanced UFOV conditions' thresholds. As expected, the correlations suffer as the difficulty in each direction grows (though it is useful to note that there are no real differences in the column of four peripheral icons, suggesting that central difficulty may matter less when peripheral difficulty is easy). This indicates that the easy enhanced UFOV and standard UFOV are measuring something different than the difficult enhanced UFOV. Either the difficult enhanced task has considerably less to do with divided attention, or divided attention is only part of what the standard task is measuring.

	Four peripheral icons	**Eight peripheral icons**	**Twelve peripheral icons**
No central blur	0.9338	0.9190	0.8997
5-pixel central blur	0.9326	0.8829	0.6067
10-pixel central blur	0.9345	0.8262	0.2096

Table 11. Pearson's correlation of standard UFOV threshold duration to threshold durations from each set of enhanced UFOV conditions. Note that the easy conditions correlate extremely well, and that as conditions get harder (particularly peripheral conditions, and particularly with interaction), that correlation suffers. This suggests that standard UFOV is largely measuring speed of processing, rather than divided attention.

This evidence that the standard test conflates divided attention with speed of processing, combined with our modeling of the underlying process of divided attention measurable by analysis of the interaction between the two difficulties, suggest that the enhanced test may be more useful in pinpointing divided attention than the standard test. For subjects whose data show a significant interaction term, the magnitude of that interaction inversely corresponds to their ability to successfully avoid interference from one task upon the other as the difficulty of both increase; in other words, it inversely corresponds to their divided attention, separate from their speed of processing (present in all the trials that make up the data in which the interaction is observed, and thus making no contribution to the significance or magnitude of the interaction).

Finally, we believe there may be useful information contained within the errors of one task relative to the other. We have noted that people with more significant interaction tend to make more errors in both directions at once, rather than favoring one or the other. However, more extensive investigation is required to understand the implications of these data. Possibly we could analyze the errors using contingency tables; unfortunately, the "no error" box would probably overwhelm anything else (makes it so that almost all the "across all conditions for this participant" contingency tables show that the error rates for the two tasks are dependent (i.e., above the critical value for chi-squared with one degree of freedom, meaning we can reject the null hypothesis of independence). However, because of the tiny (sometimes zero) counts in all three error boxes (though never the "no error" box) in the individual conditions, almost all of those *cannot* reject the hypothesis of independence. However, in this data set, the "errors in both directions" box (lower right of the contingency table) is almost always *also* higher than would be expected from independent variables. Thus, it is really both the "no errors" and the "errors in both directions" boxes that cause almost all participants to reject the null hypothesis when a contingency table of errors is made over all test conditions.

This is why it's so interesting (and potentially useful) that the people who show worse divided attention (by our working "distance from ideal" measurement) tend to show much higher numbers in the "both" column than people with good divided attention. Participants don't vary that much, percentagewise, in the single-task-only-error boxes, possibly because as a rule the errors get more frequent as a particular task gets harder. For example, if the central task is very hard but the peripheral task is very easy, almost everyone will make more errors on the central task than the peripheral task. However, the people with poor divided attention will tend to make more errors on the peripheral (easy) task *while* they're making errors on the central (difficult) task. This means they're much more likely to make an error in both directions at once. It also means that their chi-squared values are higher, because if the errors from one direction were truly independent of the other direction, it would be far less likely that someone would make an error on both tasks than that they would make an error on only one task. This is true regardless of the difficulty of the tasks, as long as they are more likely to each task correct than to make an error, which is true due to the nature of the staircase algorithm. Thus, we may be able to correlate our "distance from ideal" divided attention numbers with both chi-squared value on these contingency tables and with errors in both directions at once.

Aim 2: Prediction of divided attention from Tally performance

Figure 33 shows that standard UFOV test results do correlate to the easiest balanced-difficulty-conditions threshold values from Tally Puzzle. Note that this correlation is not quite as strong as the test-retest correlation in these participants (0.72 compared to 0.82), but it is still a strong correlation especially for psychometric data.(120) The differences are likely due to the fact that a true [easy, easy] condition could not be measured for Tally under the current experimental design, and thus there is some amount of interference between the two tasks in this reference condition. Additionally, Tally is a more difficult visual task, requiring most people to approach the two parts of it serially. Nevertheless, the standard UFOV value can largely be predicted from performance on Tally.

Table 12 shows correlations between thresholds for each of the measured Tally conditions and standard UFOV. This likewise shows higher correlation at easier conditions and lower correlation in difficult and highly-interaction conditions. This, along with the high correlation between the magnitude of interaction terms for the enhanced UFOV and Tally Puzzle (see Results), suggests that Tally can utilize the same underlying process models to measure divided attention through examination of the interaction between difficulties.

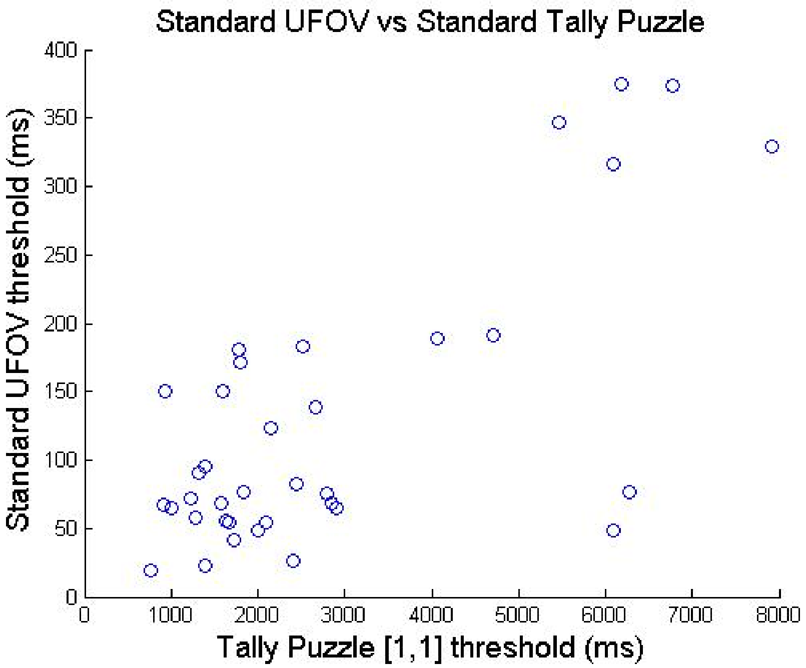


Figure 33. 75% accuracy thresholds on standard UFOV subtest 2 trial sets, compared to reference Tally Puzzle trial sets. Each point is a single subject. These are 75% accuracy duration thresholds from first-run UFOVs2 thresholds compared to reference (one decision in rows and one decision in columns) conditions in Tally Puzzle (Pearson's r = 0.72). As the [easy, easy] condition could not be tested for Tally Puzzle, this is the closest set of conditions to standard UFOV. Note that Tally Puzzle is considerably more visually complex than standard UFOV, even with easy conditions.

Implications

The implications for this work are twofold: first, for the Useful Field of View as a test for divided attention, and second, for 21 Tally as a contribution to the field of cognitive health coaching and unobtrusive continuous monitoring.

To understand the implications for the UFOV, it's important to note that the standard test was developed as a driving test for elderly people, to help determine whether they should continue to be on the road. In that, it is a success; the UFOV score is significantly negatively correlated with crashes, at-fault crashes, and driving errors.(36;50;121) However, it was not developed as a clinical test to separate out speed of processing, divided attention, or any of the other cognitive skills that make up a competent driver. It has been adopted as the de facto standard test of divided attention for two reasons: first, because no other widely validated test measures divided attention as well, and second, because the primary goal of many clinicians in testing divided attention in the elderly is to determine driving ability.

	No column decisions	**One column Decision**	**Two column decisions**
No row decisions	N/A	0.7224	0.6476
One row decision	0.7103	0.7176	0.5113
Two row decisions	0.6329	0.4841	0.3471

Table 12. Pearson's correlation of standard UFOV threshold duration to threshold durations from each set of Tally Puzzle conditions. Similarly to enhanced UFOV, correlation between Tally Puzzle and standard UFOV deteriorates as conditions become more difficult. Correlation is generally worse than eUFOV, as might be expected from a considerably more complex dual task; however, the easy conditions still have > 0.7 correlation coefficients, which is relatively high for psychomotor tasks. This suggests that these easy conditions may be applicable for measurement of speed of processing (though there may be more suitable unobtrusive tests).

The current UFOV subtest 2 claims to measure divided attention. However, in making its final "score" the only element that is varied (or in fact measured) is the stimulus duration. Even though there are two separate visual tasks being performed simultaneously, and the difficulty of each of those tasks is arguably changing due to the changing duration of the stimulus, neither task's difficulty is being varied with respect to the other task. This is akin to measuring a car's gas tank by how far it travels before running out of fuel. If two cars are the same weight, this might be a reasonable measure; however, in real life cars can be far different sizes. In this analogy, a person with poor speed of processing is like a very heavy car; even if it has the same size gas tank (divided attention ability), it won't be able to travel as far (score as well on the UFOV). Much better would be to measure the gas tank directly; in this case, that means comparing the performance on each task with and without the interference of the other task. The ability to deal with that interference is the essence of divided attention, and should be the proper measure of that ability. The UFOVs2, on the other hand, can only measure relative divided attention ability for people who have roughly the same speed of processing. One could use the

UFOVs1 to determine subjects' speed of processing and then only compare their UFOVs2 scores for those within a certain range, but this still would not fundamentally allow us to give someone a divided attention score that would be relevant across contexts and contribute to meaningful diagnoses and remediation strategies.

What we've done with the Enhanced UFOV is to vary the difficulty of each task independently, allowing us to compare each when the other is at either a high difficulty or a low difficulty. This allows us to calculate a divided attention score that is both biologically relevant and comparable across subjects. Additionally, the extremely high correlation between standard score and easy enhanced conditions threshold, indicates that the loss of information when switching to the enhanced UFOV is virtually zero. This, combined with the benefit of more accurate divided attention measurement, suggests that the enhanced UFOV would be a suitable replacement for the UFOV as a standard test for divided attention.

That said, the further implications of this work upon UFOV reflect a potential misunderstanding of the significance of divided attention to driving ability. The UFOV was developed as an off-road driving test, and remains the test most highly correlated with driving ability (crashes, at-fault crashes, and driving errors). As such, if this test is in fact measuring speed of processing more than divided attention, this may indicate that general driving ability is more closely tied to speed of processing, and less to divided attention, than previously believed. If so, we hypothesize that divided attention is more closely tied to driving ability under unusual (i.e., dangerous) conditions, while speed of processing affects driving ability during usual and unusual conditions. If true, this would also help explain the higher correlation between standard UFOV and driving errors (which occur during the course of normal, non-dangerous driving) than between standard UFOV and crashes.

Perhaps equally encouraging is the possibility suggested by this work that the same measurement of both speed of processing and divided attention can be carried out in a cognitive computer game setting. If speed of processing and divided attention can be measured unobtrusively by 21 Tally, this opens up the possibility of longitudinal off-road measurement of driving capability, which may help us to understand what forms of cognition are most vital to safe driving and how those cognitive abilities change over time.

The implications for this work upon cognitive health coaching and unobtrusive monitoring are also important. Without the development of tests that can be incorporated into fun, repeatable tasks (or the development of metrics from those tasks that already exist), the serious health benefits of unobtrusive monitoring can never be realized. Cognitive health coaches need longitudinal in-home monitoring data in order to observe trends in participant behavior, pinpoint problem areas that should be addressed through lifestyle changes or cognitive exercise, and devise individualized strategies for continued cognitive health.

Generalizing from lab tests to real-world environments can be challenging. Many studies have suggested that adhering strictly to experimental designs created to generate particular results in the lab will prevent easy translation to real-world applications.(122) This is why it is vital for new methods of cognitive diagnosis to simultaneously explore both lab-designed experiments and usability-tested (and user-driven) designs in a naturalistic home environment. Smart homes and "living lab" setups help to bridge these gaps; however, only validation both from the lab, that the test measures the expected cognitive or neural processes, and from the home, that people will actually use the software, will allow us to produce unobtrusive continuous measures that will actually allow trend data to be collected over the long term.

The 21 Tally game has already proven its worth as a casual game that participants in the cognitive health coaching studies will play multiple times a week for years.(12) This work allows that wealth of cognitive data to be incorporated into the growing body of cognitive trend information that we have for each of these participants. This, in addition to providing us with better ways to care for these participants' cognitive health, will help us pinpoint indications of the start of decline and hopefully assist future research in preventing cognitive decline or the development of Alzheimer's or dementia.

Limitations

As with all experimental work, this study has limitations. First, we have no correlation with real-life tasks (like driving) because no such testing was performed. We can predict scores on a test which does correlate to driving ability, but that isn't a true indicator of how closely this test predicts real-life behavior. Second, due to time constraints, this study design was cross-sectional rather than longitudinal. Thus, we are mostly only able to compare across subjects rather than being able to compare someone to their own baseline, which is the real strength of continuous unobtrusive cognitive testing.

Further, it's reasonable to suspect that there may be some remediative quality to the Enhanced UFOV or 21 Tally (that is, practicing these tasks actually increases participants' divided attention ability). This has been shown with repeatedly taking the standard UFOV, for instance.(49) We had no way of distinguishing this learning dynamic from our analysis. Though potentially problematic for accurate measurement of divided attention, it is unlikely that such an effect would be large over the short term of the testing done in this study. For future longitudinal tests, this may be a strength rather than a limitation, as it may reflect real improvements in divided attention.

Finally, there are indications of potential reasons for added variability in our data located in the self-report questionnaires and notes taken by the examiner during the study. Theoretically all participants with uncorrected vision problems should have been screened out before taking the study, but some of the elders had different opinions about what "uncorrected" meant, allowing some participants who were not close to 20/20 to participate. Some elders mentioned taking medications that might have a cognitive impact, though this was not investigated. Finally, there were participants who reported that they didn't fully understand either UFOV or Tally in the free-text portion of the post-participation questionnaire.

Testing and Time of Day

One potential confound of this work that was not controlled for is the time of day that tests were administered. Several studies have shown significant time-of-day effects for cognitive testing, both in animal and human models.(123-125) For example, Hasher, Zacks, and Rahhal(126) point out three important factors dealing with age and time-of-day: 1) older adults tend to perform better on cognitive tests in the morning rather than evening, while the opposite is true for younger adults; 2) younger adults score better than older adults, on average, when both are tested at their "peak" time; 3) the difference between scores during "peak" versus "non-peak" times are much larger in older adults than younger adults (i.e., time makes a bigger difference as age increases). Testing times in this experiment were determined by convenience (participant and researcher schedule). The majority (7 out of 11) of our oldest (75+) age group self-selected for morning testing times, while the vast majority (26 out of 30) of our youngest age groups (25 – 59) opted for late afternoon or evening testing times; presumably this was largely due to work schedules. No attempt was made to account for testing time when analyzing our results.

Cognitive versus Perceptual Load

The experimental model of the enhanced UFOV is designed to alter the difficulty of each task by changing the cognitive load – that is, the amount of cognitive processing needed in order to successfully complete the recognition (central) or location (peripheral) task. We believe we were successful in this goal in the central task and somewhat in the peripheral task. However, it should be noted that by changing the number of distractors concurrently displayed on the screen for the peripheral task we also change the perceptual load (i.e., the amount of processing needed by the visual system to simply take in the information displayed before it is sorted into target/distractor categories). Lavie's load theory(127) suggests that the processing of distractors changes based on the amount of perceptual load present in a test; specifically, increasing perceptual load decreases processing of distractors. Further work in this field has shown that increasing cognitive load, by contrast, actually increases processing of distractors.(128;129)

This is a potential confound, as the perceptual and cognitive loads both increase when we “increase difficulty” for the peripheral task in the enhanced UFOV. It is possible that the perceptual load increase contributed to the final threshold value for high-peripheral-difficulty conditions, which would influence our measurement of the interaction between difficulties (and thus, divided attention). Further, the 21 Tally game has an extremely large perceptual load relative to most lab-based cognitive tests – its “game” background is full of irrelevant details that must be ignored in a short-stimulus-duration task. This constantly-high perceptual load (relative to the changing perceptual load of the eUFOV) is a potential confound, and may account for the relatively small range of interaction values measured in the Tally portion of the experiment.

Future Directions

This work provides a strong foundation for a potential wealth of new investigation, both to correct for study limitations and to further analyze collected data. Broadly, we would like to incorporate the Tally puzzle mode into a large-scale cognitive health coaching framework, and continue analysis of the data we have already collected. In terms of insertion into a cognitive health coaching context, we would like to update the games already being played with the new metrics, perform a longitudinal study to determine our ability to assess trends over time, test our correlation to driving ability, and investigate the ability of 21 Tally to improve divided attention. In terms of analyzing already-collected data, we would like to more comprehensively analyze our qualitative reports, including researcher notes and free-text responses, analyze the correlation between our divided attention testing and other neuropsychological testing that our elder participants have undergone in other studies, analyze the EEG data from the pilot study, and further investigate the relationship between errors on each of the two tasks in both eUFOV and Tally.

21 Tally is already being played by many elders in several assisted living facilities around the Portland area as part of the cognitive health coaching project.(130) These elders have played well over a hundred thousand games of Tally in total. Thus, translation of these findings on the Tally “puzzle” mode (taking only the most divided-attention-rich boards, as was done in this dissertation work) to normal Tally performance would allow us to have much more understanding about current and past players’ divided attention ability. This would let us make better use of the data we already have, as well as providing an impetus for updating the game to incorporate this “puzzle” mode into the games suite that they’re already playing, which will enrich the incoming data considerably.

Second, we would like to perform a longitudinal study involving the Tally Puzzle and enhanced (and standard) UFOV testing at regular intervals. This would allow us to study the utility of having a more continuous behavioral record from an unobtrusive monitoring source as someone moves from normal cognition through MCI to AD.

We would also like to use road testing or driving simulation testing with the enhanced UFOV, to determine how much our divided attention score actually maps to driving ability. If this is the primary factor in the correlation (i.e., speed of processing isn’t as important), then eUFOV might be more helpful than standard UFOV in determining potential driving ability for elders. If not, it would be interesting to know that divided attention is really most useful during stressful/unusual driving situations, not under normal/expected conditions. We could then further test the differences between normal and unusual driving conditions (likely in a simulator rather than on the road, for safety), with an eye to the differences between speed of processing and divided attention.

We intend to test the remediative properties of 21 Tally and the Tally Puzzle. Since standard UFOV has been shown to improve elders’ divided attention ability, it is reasonable to suppose that Tally might have some of the same properties. This would in no way diminish its utility as a diagnostic tool. Much as we can take the number of pushups someone can do as a measure of their arm strength even as the exercise increases that strength, Tally could continue to be used to measure divided attention ability even as it improves that ability.

In order to fully understand the participants' experience during the experiment (which reflects heavily upon their likely enjoyment of any game which might be built out of this project), the transcribed questions and exclamations of each participant should be coded into positive or negative categories, such as "frustration" or "excitement." Then the various parts of the experiment can be examined with an eye to improving negative and promoting positive items. Further, systematic analyses of free-text portions of the questionnaires, or of researcher notes during the experiment, may prove beneficial to understanding participant experience.

Many of our elder participants also participate in other studies at OHSU, and have undergone extensive batteries of neuropsychological testing both in the past and at similar times to the collection of our data. These tests include the Trail-Making Test and the Symbol-Digit Modalities Test, both of which test set switching and might well be correlated with our tests of divided attention. Tests of general cognitive ability like the Mini-Mental State Exam and Clinical Dementia Rating might also reasonably be expected to correlate to our findings. More extensive demographic information was also collected in these studies, which might help us to identify broader demographic trends (though with only twelve elder participants, all from the same city, that is somewhat unlikely).

We would like to analyze the EEG data taken from our pilot study, which was largely neglected due to a change in research personnel during the course of the study. We believe these data may shed light on cognitive load during dual task experiments, even if the specific tasks performed were not the ones we eventually used in the main study.

Finally, our investigation of the errors of each subtask relative to the significance and magnitude of the interaction term (see Discussion: Aim 1) needs further needs further analysis. People who show worse divided attention may be making more errors on both subtasks at once than people with better divided attention; if so, it is likely that this investigation will require division of our sample population into subgroups based on error behavior. These subgroups may not be large enough to show statistically significant differences without further data collection.

Conclusion

Fundamental to the cognitive health coaching movement is the idea of unobtrusively monitoring participants in a variety of noisy, variable, information-rich daily activities, and then combining all of these data streams into a single composite model of the participant as a constantly-updating whole. We are still in the beginning stages of this effort, and the identification and investigation of the way particular tasks relate to cognition is ongoing. Divided attention being an increasingly important cognitive skill in our increasingly complex environment, a way to unobtrusively monitor this ability is an important step forward in our attempt to maintain people's independence as they age.

In this study, we have shown that at least in this population both the enhanced UFOV and the 21 Tally Puzzle are better measures of divided attention than the standard UFOV. Further, the Tally Puzzle can be used to assess an individual's ability to divide attention in the lab, and by extension, the 21 Tally casual game can be used to unobtrusively monitor divided attention in the home. This breakthrough will allow participants in the cognitive health coaching program to observe trends, variability, and current divided attention ability relative to an individualized baseline. This, in turn, could shed light on the origins of cognitive decline and allow our cognitive health coaches to recommend remediative strategies targeting divided attention for those identified as at risk for further decline.

References

(1) United States Census Bureau. Population. http://www.census.gov/population/age/data/2011.html . 2011. 10-27-2012.

(2) Fries JF. Reducing cumulative lifetime disability: the compression of morbidity. Br J Sports Med 1998; 32(3):193.

(3) Mahoney DF, Purtilo RB, Webbe FM, Alwan M, Bharucha AJ, Adlam TD et al. In-home monitoring of persons with dementia: ethical guidelines for technology research and development. Alzheimers Dement 2007; 3(3):217-226.

(4) Boise L, Camicioli R, Morgan DL, Rose JH, Congleton L. Diagnosing dementia: perspectives of primary care physicians. Gerontologist 1999; 39(4):457-464.

(5) Howieson DB, Dame A, Camicioli R, Sexton G, Payami H, Kaye JA. Cognitive markers preceding Alzheimer's dementia in the healthy oldest old. J Am Geriatr Soc 1997; 45(5):584-589.

(6) Bonnefond A, Rohmer O, Hoeft A, Muzet A, Tassi P. Interaction of age with time of day and mental load in different cognitive tasks. Percept Mot Skills 2003; 96(3 Pt 2):1223-1236.

(7) Cohen-Mansfield J, Thein K, Dakheel-Ali M, Marx MS. Engaging nursing home residents with dementia in activities: the effects of modeling, presentation order, time of day, and setting characteristics. Aging Ment Health 2010; 14(4):471-480.

(8) KAPLAN OJ, RUMBAUGH DM, MITCHELL DC, THOMAS ED. Effects of level of surviving abilities, time of day, and test-retest upon psychological performance in seniles. J Gerontol 1963; 18:55-59.

(9) Levin HS. A guide to clinical neuropsychological testing. Arch Neurol 1994; 51(9):854-859.

(10) Austin D, Leen T, Hayes TL, Kaye J, Jimison H, Mattek N et al. Model-based inference of cognitive processes from unobtrusive gait velocity measurements. Conf Proc IEEE Eng Med Biol Soc 2010; 1:5230-5233.

(11) Hatt WJ, Vanbaak EA, Jimison HB, Hagler S, Hayes TL, Pavel M et al. The exploration & forensic analysis of computer usage data in the elderly. Conf Proc IEEE Eng Med Biol Soc 2009; 2009:1216-1219.

(12) Jimison HB, Pavel M, Bissell P, McKanna J. A framework for cognitive monitoring using computer game interactions. Stud Health Technol Inform 2007; 129(Pt 2):1073-1077.

(13) Jimison HB, Pavel M, Pavel J, McKanna J. Home monitoring of computer interactions for the early detection of dementia. Conf Proc IEEE Eng Med Biol Soc 2004; 6:4533-4536.

(14) Kelly VE, Schrager MA, Price R, Ferrucci L, Shumway-Cook A. Age-associated effects of a concurrent cognitive task on gait speed and stability during narrow-base walking. J Gerontol A Biol Sci Med Sci 2008; 63(12):1329-1334.

(15) Broadbent D, Broadbent MH. Human attention: the exclusion of distracting information as a function of real and apparent separation of relevant and irrelevant events. Proc Biol Sci 1990; 242(1303):11-16.

(16) Ball LJ, Birge SJ. Prevention of brain aging and dementia. Clin Geriatr Med 2002; 18(3):485-503.

(17) Broadbent DE. Failures of attention in selective listening. J Exp Psychol 1952; 44(6):428-433.

(18) Broadbent DE. The role of auditory localization in attention and memory span. J Exp Psychol 1954; 47(3):191-196.

(19) James W. Attention. The Principles of Psychology. 1890.

(20) Broadbent DE, Gregory M. DIVISION OF ATTENTION AND THE DECISION THEORY OF SIGNAL DETECTION. Proc R Soc Lond B Biol Sci 1963; 158:222-231.

(21) Broadbent DE. A mechanical model for human attention and immediate memory. Psychol Rev 1957; 64(3):205-215.

(22) Broadbent DE. Performance and its measurement. Br J Clin Pharmacol 1984; 18 Suppl 1:5S-11S.

(23) Pavel M. A working definition of attention. Personal Communication. 2013. 4-19-2010.

(24) Posner M. Orienting of Attention. The Quarterly Journal of Experimental Psychology 1980; 32(1).

(25) Sharpe CR. The visibility and fading of thin lines visualized by their controlled movement across the retina. J Physiol 1972; 222(1):113-134.

(26) Brouwer WH, Waterink W, Van Wolffelaar PC, Rothengatter T. Divided attention in experienced young and older drivers: lane tracking and visual analysis in a dynamic driving simulator. Hum Factors 1991; 33(5):573-582.

(27) Scalf PE, Colcombe SJ, McCarley JS, Erickson KI, Alvarado M, Kim JS et al. The neural correlates of an expanded functional field of view. J Gerontol B Psychol Sci Soc Sci 2007; 62 Spec No 1:32-44.

(28) Verghese J, Buschke H, Viola L, Katz M, Hall C, Kuslansky G et al. Validity of divided attention tasks in predicting falls in older individuals: a preliminary study. J Am Geriatr Soc 2002; 50(9):1572-1576.

(29) Parasuraman R, Nestor PG. Attention and driving skills in aging and Alzheimer's disease. Hum Factors 1991; 33(5):539-557.

(30) Sarter M, Turchi J. Age- and dementia-associated impairments in divided attention: psychological constructs, animal models, and underlying neuronal mechanisms. Dement Geriatr Cogn Disord 2002; 13(1):46-58.

(31) Jimison H, Pavel M, McKanna J, Pavel J. Unobtrusive monitoring of computer interactions to detect cognitive status in elders. IEEE Trans Inf Technol Biomed 2004; 8(3):248-252.

(32) Jimison H, Gorman P, Woods S, Nygren P, Walker M, Norris S et al. Barriers and drivers of health information technology use for the elderly, chronically ill, and underserved. Evid Rep Technol Assess (Full Rep) 2008;(175):1-1422.

(33) Jimison HB, McKanna J, Ambert K, Hagler S, Hatt WJ, Pavel M. Models of cognitive performance based on home monitoring data. Conf Proc IEEE Eng Med Biol Soc 2010; 1:5234-5237.

(34) Schwamm LH, Van Dyke C, Kiernan RJ, Merrin EL, Mueller J. The Neurobehavioral Cognitive Status Examination: comparison with the Cognitive Capacity Screening Examination and the Mini-Mental State Examination in a neurosurgical population. Ann Intern Med 1987; 107(4):486-491.

(35) Chen P, Ratcliff G, Belle SH, Cauley JA, DeKosky ST, Ganguli M. Cognitive tests that best discriminate between presymptomatic AD and those who remain nondemented. Neurology 2000; 55(12):1847-1853.

(36) Ball KK, Roenker DL, Wadley VG, Edwards JD, Roth DL, McGwin G, Jr. et al. Can high-risk older drivers be identified through performance-based measures in a Department of Motor Vehicles setting? J Am Geriatr Soc 2006; 54(1):77-84.

(37) Howieson D, Wild K. Neuropsychological tests used for divided attention. Electronic Communication. 4-4-2012.

(38) Reitan R. Trail Making Test Manual. Adjutant General's Office, War Department, US Army, 1944.

(39) Mathias JL, Lucas LK. Cognitive predictors of unsafe driving in older drivers: a meta-analysis. Int Psychogeriatr 2009; 21(4):637-653.

(40) Innes CR, Jones RD, Anderson TJ, Hollobon SG, Dalrymple-Alford JC. Performance in normal subjects on a novel battery of driving-related sensory-motor and cognitive tests. Behav Res Methods 2009; 41(2):284-294.

(41) Bidet-Caulet A, Fischer C, Besle J, Aguera PE, Giard MH, Bertrand O. Effects of selective attention on the electrophysiological representation of concurrent sounds in the human auditory cortex. J Neurosci 2007; 27(35):9252-9261.

(42) Drake CL, Roehrs TA, Royer H, Koshorek G, Turner RB, Roth T. Effects of an experimentally induced rhinovirus cold on sleep, performance, and daytime alertness. Physiol Behav 2000; 71(1-2):75-81.

(43) Pashler H. Do response modality effects support multiprocessor models of divided attention? J Exp Psychol Hum Percept Perform 1990; 16(4):826-842.

(44) Posit Science Corporation. AAA Clubs Help Older Drivers Keep Driving Safely. http://www.positscience.com/news/aaa-clubs-help-older-drivers-keep-driving-safely . 4-28-2012. PSC Press Release. 4-16-2013.

(45) Clay OJ, Wadley VG, Edwards JD, Roth DL, Roenker DL, Ball KK. Cumulative meta-analysis of the relationship between useful field of view and driving performance in older adults: current and future implications. Optom Vis Sci 2005; 82(8):724-731.

(46) Jobe JB, Smith DM, Ball K, Tennstedt SL, Marsiske M, Willis SL et al. ACTIVE: a cognitive intervention trial to promote independence in older adults. Control Clin Trials 2001; 22(4):453-479.

(47) Ball K. Enhancing Mobility in the Elderly: Attentional Interventions for Driving. Assessment and Intervention Issues Across the Lifespan. Lawrence Erlbaum, 1997: 267-292.

(48) Ball K, Owsley C, Sloane ME, Roenker DL, Bruni JR. Visual attention problems as a predictor of vehicle crashes in older drivers. Invest Ophthalmol Vis Sci 1993; 34(11):3110-3123.

(49) Ball K, Berch DB, Helmers KF, Jobe JB, Leveck MD, Marsiske M et al. Effects of cognitive training interventions with older adults: a randomized controlled trial. JAMA 2002; 288(18):2271-2281.

(50) Edwards JD, Ross LA, Wadley VG, Clay OJ, Crowe M, Roenker DL et al. The useful field of view test: normative data for older adults. Arch Clin Neuropsychol 2006; 21(4):275-286.

(51) Oxley J, Whelan M. It cannot be all about safety: the benefits of prolonged mobility. Traffic Inj Prev 2008; 9(4):367-378.

(52) Ross LA, Clay OJ, Edwards JD, Ball KK, Wadley VG, Vance DE et al. Do older drivers at-risk for crashes modify their driving over time? J Gerontol B Psychol Sci Soc Sci 2009; 64(2):163-170.

(53) Horswill MS, Kemala CN, Wetton M, Scialfa CT, Pachana NA. Improving older drivers' hazard perception ability. Psychol Aging 2010; 25(2):464-469.

(54) Horswill MS, Pachana NA, Wood J, Marrington SA, McWilliam J, McCullough CM. A comparison of the hazard perception ability of matched groups of healthy drivers aged 35 to 55, 65 to 74, and 75 to 84 years. J Int Neuropsychol Soc 2009; 15(5):799-802.

(55) Petersen RC, Stevens JC, Ganguli M, Tangalos EG, Cummings JL, DeKosky ST. Practice parameter: early detection of dementia: mild cognitive impairment (an evidence-based review). Report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology 2001; 56(9):1133-1142.

(56) Lyman S, Ferguson SA, Braver ER, Williams AF. Older driver involvements in police reported crashes and fatal crashes: trends and projections. Inj Prev 2002; 8(2):116-120.

(57) Croston J, Meuser TM, Berg-Weger M, Grant EA, Carr DB. Driving Retirement in Older Adults with Dementia. Top Geriatr Rehabil 2009; 25(2):154-162.

(58) Perkinson MA, Berg-Weger ML, Carr DB, Meuser TM, Palmer JL, Buckles VD et al. Driving and dementia of the Alzheimer type: beliefs and cessation strategies among stakeholders. Gerontologist 2005; 45(5):676-685.

(59) Blanchard RA, Myers AM, Porter MM. Correspondence between self-reported and objective measures of driving exposure and patterns in older drivers. Accid Anal Prev 2010; 42(2):523-529.

(60) Staplin L, Stutts J, Martell C. Identifying Behaviors and Situations Associated With Increased Crash Risk for Older Drivers. 6-1-2009. National Highway Traffic Safety Administration.

(61) 2009 Overview of Assisted Living. 2009. American Association of Homes and Services for the Aging, American Seniors Housing Association, Assisted Living Federation of American, National Center for Assisted Living, and National Investment Center for the Seniors Housing & Care Industry.

(62) Richards E, Bennett PJ, Sekuler AB. Age related differences in learning with the useful field of view. Vision Res 2006; 46(25):4217-4231.

(63) Callahan CM, Hendrie HC, Tierney WM. Documentation and evaluation of cognitive impairment in elderly primary care patients. Ann Intern Med 1995; 122(6):422-429.

(64) Wolinsky FD, Unverzagt FW, Smith DM, Jones R, Stoddard A, Tennstedt SL. The ACTIVE cognitive training trial and health-related quality of life: protection that lasts for 5 years. J Gerontol A Biol Sci Med Sci 2006; 61(12):1324-1329.

(65) Wolinsky FD, Unverzagt FW, Smith DM, Jones R, Wright E, Tennstedt SL. The effects of the ACTIVE cognitive training trial on clinically relevant declines in health-related quality of life. J Gerontol B Psychol Sci Soc Sci 2006; 61(5):S281-S287.

(66) Boise L, Neal MB, Kaye J. Dementia assessment in primary care: results from a study in three managed care systems. J Gerontol A Biol Sci Med Sci 2004; 59(6):M621-M626.

(67) Kaye JA, Maxwell SA, Mattek N, Hayes TL, Dodge H, Pavel M et al. Intelligent Systems For Assessing Aging Changes: home-based, unobtrusive, and continuous assessment of aging. J Gerontol B Psychol Sci Soc Sci 2011; 66 Suppl 1:i180-i190.

(68) Clark JE, Lanphear AK, Riddick CC. The effects of videogame playing on the response selection processing of elderly adults. J Gerontol 1987; 42(1):82-85.

(69) Drew D, Waters J. Video games: Utilization of a novel strategy to improve perceptual motor skills and cognitive functioning in the non-institutionalized elderly. Cognitive Rehabilitation 1986; 4:26-31.

(70) Nintendo Brain Age:game overview. http://www.nintendo.com/games/detail/Y9QLGBWxkmRRzsQEQtvqGqZ63_CjS_9F . 1-10-2012. 1-10-2012.

(71) Kawashima R, Okita K, Yamazaki R, Tajima N, Yoshida H, Taira M et al. Reading aloud and arithmetic calculation improve frontal function of people with dementia. J Gerontol A Biol Sci Med Sci 2005; 60(3):380-384.

(72) Mahncke HW, Connor BB, Appelman J, Ahsanuddin ON, Hardy JL, Wood RA et al. Memory enhancement in healthy older adults using a brain plasticity-based training

program: a randomized, controlled study. Proc Natl Acad Sci U S A 2006; 103(33):12523-12528.

(73) Greenfield P, deWinstanley P, Kilpatrick H, Kaye D. Action video games and informal education: effects on strategies for dividing visual attention. Journal of Applied Developmental Psychology 1994; 15:105-123.

(74) Koepp MJ, Gunn RN, Lawrence AD, Cunningham VJ, Dagher A, Jones T et al. Evidence for striatal dopamine release during a video game. Nature 1998; 393(6682):266-268.

(75) Castel AD, Pratt J, Drummond E. The effects of action video game experience on the time course of inhibition of return and the efficiency of visual search. Acta Psychol (Amst) 2005; 119(2):217-230.

(76) Green CS, Bavelier D. Action video game modifies visual selective attention. Nature 2003; 423(6939):534-537.

(77) Belchior P. COGNITIVE TRAINING WITH VIDEO GAMES TO IMPROVE DRIVING SKILLS AND DRIVING SAFETY AMONG OLDER ADULTS. University of Florida, 2007.

(78) Green CS, Bavelier D. Action-video-game experience alters the spatial resolution of vision. Psychol Sci 2007; 18(1):88-94.

(79) Campbell C. Easy to Play, Hard to Master. http://www.businessweek.com/stories/2005-11-28/easy-to-play-hard-to-master . 11-28-2005. BusinessWeek. 3-17-2013.

(80) Paavilainen J, Kuittinen J, Kultima A, Niemelä J. Casual games discussion. Proceedings of the 2007 Conference on Future Play 2007;105-112.

(81) Dobson J. Survey: Popcap releases casual game findings. http://www.gamasutra.com/php-bin/news_index.php?story=10861 . 2013. 1-9-2013.

(82) Gander C. Who plays casual games. http://3rdsense.com/blog/01062010-0203/who-plays-casual-games . 2013. 1-9-2013.

(83) What makes things fun to learn? heuristics for designing instructional computer games. 80; 1980.

(84) Dunn-Rankin P, Leton DA, Shelton VF. Congruency factors related to visual confusion of English letters. Percept Mot Skills 1968; 26(2):659-666.

(85) Gervais MJ, Harvey LO, Jr., Roberts JO. Identification confusions among letters of the alphabet. J Exp Psychol Hum Percept Perform 1984; 10(5):655-666.

(86) Reich LN, Bedell HE. Relative legibility and confusions of letter acuity targets in the peripheral and central retina. Optom Vis Sci 2000; 77(5):270-275.

(87) Thorson G. An alternative for judging confusability of visual letters. Percept Mot Skills 1976; 42(1):116-118.

(88) Treisman AM, Gelade G. A feature-integration theory of attention. Cogn Psychol 1980; 12(1):97-136.

(89) Wolfe JM, O'Neill P, Bennett SC. Why are there eccentricity effects in visual search? Visual and attentional hypotheses. Percept Psychophys 1998; 60(1):140-156.

(90) Wolfe JM, Cave KR, Franzel SL. Guided search: an alternative to the feature integration model for visual search. J Exp Psychol Hum Percept Perform 1989; 15(3):419-433.

(91) Duncan J, Humphreys GW. Visual search and stimulus similarity. Psychol Rev 1989; 96(3):433-458.

(92) Palmer J, Verghese P, Pavel M. The psychophysics of visual search. Vision Res 2000; 40(10-12):1227-1268.

(93) Sireteanu R, Rettenbach R. Perceptual learning in visual search: fast, enduring, but non-specific. Vision Res 1995; 35(14):2037-2043.

(94) Shiffrin RM, Dumais ST. The development of automatism. In: Anderson J, editor. Cognitive Skills and Their Acquisition. Hillsdale, NJ: Erlbaum, 1981: 111-140.

(95) Sireteanu R, Rettenbach R. Perceptual learning in visual search generalizes over tasks, locations, and eyes. Vision Res 2000; 40(21):2925-2949.

(96) Fahle M, Morgan M. No transfer of perceptual learning between similar stimuli in the same retinal position. Curr Biol 1996; 6(3):292-297.

(97) Leonards U, Rettenbach R, Nase G, Sireteanu R. Perceptual learning of highly demanding visual search tasks. Vision Res 2002; 42(18):2193-2204.

(98) Ruthruff E, Johnston JC, Van Selst M. Why practice reduces dual-task interference. J Exp Psychol Hum Percept Perform 2001; 27(1):3-21.

(99) Gilbert DK, Rogers WA. Age-related differences in perceptual learning. Hum Factors 1996; 38(3):417-424.

(100) Wichmann FA, Hill NJ. The psychometric function: II. Bootstrap-based confidence intervals and sampling. Percept Psychophys 2001; 63(8):1314-1329.

(101) Wichmann FA, Hill NJ. The psychometric function: I. Fitting, sampling, and goodness of fit. Percept Psychophys 2001; 63(8):1293-1313.

(102) McKanna JA, Jimison H, Pavel M. Divided attention in computer game play: analysis utilizing unobtrusive health monitoring. Conf Proc IEEE Eng Med Biol Soc 2009; 2009:6247-6250.

(103) Cohen A. Simulation for power calculations. Personal Communication. 2013. 6-21-2011.

(104) Pavlovic G, Tekalp AM. Maximum likelihood parametric blur identification based on a continuous spatial domain model. IEEE Trans Image Process 1992; 1(4):496-504.

(105) Pashler H. Dual-task interference in simple tasks: data and theory. Psychol Bull 1994; 116(2):220-244.

(106) Pashler HE, Johnston JC. Attentional limitations in dual-task performance. In: Pashler HE, editor. Attention. Hove (United Kingdom): Psychology Press, 1998: 155-189.

(107) Pashler H, Johnston JC. Cronometric Evidence for Central Postponement in Temporally Overlapping Tasks. Q J Exp Psychol 1989; 41A(1):19-45.

(108) Smith MC. The psychological refractory period as a function of performance of a first response. Q J Exp Psychol 1967; 19(4):350-352.

(109) Smith MC. Theories of the psychological refractory period. Psychol Bull 1967; 67(3):202-213.

(110) Ruthruff E, Pashler HE, Klaassen A. Processing bottlenecks in dual-task performance: structural limitation or strategic postponement? Psychon Bull Rev 2001; 8(1):73-80.

(111) Allport D, Styles E, Hsieh S. Shifting intentional set: Exploring the dynamic control of tasks. In: Umilta C, Moscovitch M, editors. Attention and performance XV. Cambridge, MA: MIT Press, 1994: 421-452.

(112) Hsieh S, Allport A. Shifting attention in a rapid visual search paradigm. Percept Mot Skills 1994; 79(1 Pt 1):315-335.

(113) Pentland A. Maximum likelihood estimation: the best PEST. Percept Psychophys 1980; 28(4):377-379.

(114) Roenker D. UFOV User's Guide. http://crag.uab.edu/VAI/PDF%20Pubs/UFOV_Manual_V6.1.4.pdf , 20. 6-1-2009. Visual Awareness Research Group, Inc. 3-8-2013.

(115) Bichot NP, Desimone R. Finding a face in the crowd: parallel and serial neural mechanisms of visual selection. Prog Brain Res 2006; 155:147-156.

(116) Howe PD, Cohen MA, Pinto Y, Horowitz TS. Distinguishing between parallel and serial accounts of multiple object tracking. J Vis 2010; 10(8):11.

(117) Luck SJ, Hillyard SA. Electrophysiological evidence for parallel and serial processing during visual search. Percept Psychophys 1990; 48(6):603-617.

(118) Oken BS, Kishiyama SS, Kaye JA. Age-related differences in visual search task performance: relative stability of parallel but not serial search. J Geriatr Psychiatry Neurol 1994; 7(3):163-168.

(119) Thornton TL, Gilden DL. Parallel and serial processes in visual search. Psychol Rev 2007; 114(1):71-103.

(120) Hobart J, Cano S, Posner H, Selnes O, Stern Y, Thomas R et al. Putting the Alzheimer's cognitive test to the test I: traditional psychometric methods. Alzheimers Dement 2013; 9(1 Suppl):S4-S9.

(121) Classen S, McCarthy DP, Shechtman O, Awadzi KD, Lanford DN, Okun MS et al. Useful field of view as a reliable screening measure of driving performance in people with Parkinson's disease: results of a pilot study. Traffic Inj Prev 2009; 10(6):593-598.

(122) Kingstone A, Smilek D, Eastwood JD. Cognitive Ethology: a new approach for studying human cognition. Br J Psychol 2008; 99(Pt 3):317-340.

(123) Bugg JM, DeLosh EL, Clegg BA. Physical activity moderates time-of-day differences in older adults' working memory performance. Exp Aging Res 2006; 32(4):431-446.

(124) Winocur G, Hasher L. Age and time-of-day effects on learning and memory in a non-matching-to-sample test. Neurobiol Aging 2004; 25(8):1107-1115.

(125) Winocur G, Hasher L. Aging and time-of-day effects on cognition in rats. Behav Neurosci 1999; 113(5):991-997.

(126) Hasher L, Zacks RT, Rahhal TA. Timing, instructions, and inhibitory control: some missing factors in the age and memory debate. Gerontology 1999; 45(6):355-357.

(127) Lavie N. Perceptual load as a necessary condition for selective attention. J Exp Psychol Hum Percept Perform 1995; 21(3):451-468.

(128) Bahrami B, Lavie N, Rees G. Attentional load modulates responses of human primary visual cortex to invisible stimuli. Curr Biol 2007; 17(6):509-513.

(129) Lavie N. Distracted and confused?: selective attention under load. Trends Cogn Sci 2005; 9(2):75-82.

(130) Parasuraman R, Mutter SA, Molloy R. Sustained attention following mild closed-head injury. J Clin Exp Neuropsychol 1991; 13(5):789-811.

Appendices

Appendix I: Surveys

Surveys were given either 1) as a word file on the computer, in which case participants typed their answers or indicated a choice by highlighting the text, or 2) printed out on paper, in which case the participants wrote answers with a pen or indicated a choice by circling or checking next to the choice.

Survey 1: Pilot Study Pre-Experiment Survey

"

Pre-participation questionnaire

The following questions will assist us in determining your eligibility to participate in this study. If you are not eligible to participate, this information will not be stored or shared with anyone.

Whether or not you are eligible to participate will not affect your health care or your relationship with OHSU in any way.

Please just mark the answer or answers which best describe your situation or preferences. You can skip any question you don't feel comfortable answering, though of course we hope you will answer them all. If you feel unsure about what a question is asking, please ask the researcher who gave you this questionnaire to clarify.

For questions in which answers are already provided, please **bold** or highlight your choice.

Age: [ ]
Race: [ ]
Gender: [ ]

Do you enjoy playing computer or video games? [Yes] [No]

If yes:

- What sorts of games do you play? Mark all that apply: [RPG] [Casual Internet] [MMO] [Gambling] [Console] [FPS] [Card] [Teaching/Training] [Puzzle] [Platformer] [Other (please specify):]
    - Please list up to three games that you have played recently: [ ]
- What systems (type of computer, type of console, etc) do you play on? [ ]
- How often do you play, on average? [< 1 hour per week] [1-3 hours per week] [3-7 hours per week] [1-2 hours per day] [> 2 hours per day]

Do you have any vision problems? [Yes] [No]

If yes:

- Please specify the nature of the problem: [ ]
- How is this currently being corrected? [ ]
- Does this correction bring you back to approximately normal vision? [Yes] [No]
  - If not, please describe your current vision: [ ]
- When was your last eye exam? [> 3 years ago] [2-3 years ago] [1-2 years ago] [< 1 year ago]

Was English your first language? [Yes] [No]

If not:

- What was your first language? [ ]
- Are you currently fluent in English? [Yes] [No]

What was the last grade you completed in school? [< 8] [some high school] [high school or equivalent] [some college] [Bachelor's or equivalent] [some grad school] [Master's, PhD, J.D., or other advanced degree]

"

Survey 2: Pilot Study Post-UFOV Survey

"

Post-participation questionnaire

If you don't understand a question or need clarification on any point, please ask the researcher who gave you this questionnaire. Also, feel free to be as candid as possible; we never judge our participants (after all, to the people receiving this you're just an anonymous number), and negative responses help us learn just as much as positive ones.

For questions in which answers are already provided, please **bold** or highlight your choice.

Useful Field Of View:

Have you ever taken any version of the UFOV before? [Yes] [No]

If so, when was the last time you took it? [ ]

Have you ever taken any other tests of divided attention before? [Yes] [No]

If so:

- What test did you take? [ ]
- When did you take it? [ ]

Did you feel that you understood the test sufficiently after completing the training? [Yes] [No]

If not, what should have been added to help your understanding of the test? [ ]

How good do you feel you are at dividing your attention (paying attention to multiple things at once in order to achieve some goal)? [Excellent] [Better than average] [Average] [Below average] [Terrible]

Do you think you did well on the UFOV? [Yes] [No] [About average] [I have no idea]

Do you consider yourself a good driver? [Yes] [No] [About average]

When was the last time you had a traffic accident or got a ticket for a moving violation? [< 1 year ago] [1-2 years ago] [3-4 years ago] [> 4 years ago] [I've never had one]

When was the last time you had a traffic accident that was determined to be your fault? [< 1 year ago] [1-2 years ago] [3-4 years ago] [> 4 years ago] [I've never had one]

"

Survey 3: Pilot Study Post-Tally Survey

"

<u>Post-participation questionnaire</u>

If you don't understand a question or need clarification on any point, please ask the researcher who gave you this questionnaire. Also, feel free to be as candid as possible; we never judge our participants (after all, to the people receiving this you're just an anonymous number), and negative responses help us learn just as much as positive ones.

For questions in which answers are already provided, please **bold** or highlight your choice.

<u>21 Tally:</u>

Had you already played 21 Tally or a game like it prior to being in the study? [Yes] [No]

Did you feel that you understood the game sufficiently after completing the training? [Yes] [No]

If not, what should have been added to help your understanding of the game? [ ]

How many training boards would you have preferred? [more boards] [fewer boards] [the number was about right]

Was the overall time required to train you to play appropriate? [I would have liked more time] [I wish it had taken less time] [it was about right]

Did you employ a particular strategy during the game? [Yes] [No]

If so, please briefly describe your strategy: [ ]

Did you enjoy playing this game, purely as a game? [Yes] [No]

If so:

- How often would you play it if it were available? [A few times a month or less] [Maybe once a week] [A few times a week] [Once a day or more]

If not:

- Please briefly describe the differences between this game and the games you play most often: [ ]

What makes the games you play fun? If you ever play games more than once, what keeps you coming back to play them again? [ ]

Purely from playing, would you know that this game was also a cognitive test? [Yes] [No]

If so, does that make it more or less motivating to play, in real life outside the lab? [More motivating] [Less motivating] [No real change]

If this were part of a health solution, which could help keep you cognitively healthy, would that change the amount you would use it? [Yes] [No]

If so, please specify how it would change your potential use: [ ]

Would you like a copy of this game to play at home? [Yes] [No]

If so, would you mind if we continued to monitor your performance for changes in cognitive state (Note: this is for future study information / planning purposes only; we aren't yet set up for remote monitoring)? [I don't want to be monitored, I just like to play the game] [I don't mind being monitored, as long as I get a copy of whatever information you collect] [I don't mind being monitored, but please don't send me the information you're collecting]

"

Survey 4: Pilot Study General post-participation questions

"

Thank you for participating in our study! If it weren't for people like you, we could never get anything done around here. If you wouldn't mind taking just a few more minutes to tell us about how this might be improved in the future, we would really appreciate it. As always, if you don't understand a question or need clarification on any point, please ask the researcher who gave you this questionnaire. Also, feel free to be as candid as possible; we never judge our participants (after all, to the people receiving this you're just an anonymous number), and negative responses help us learn just as much as positive ones.

For questions in which answers are already provided, please **bold** or highlight your choice.

General questions:

Overall, did you feel like the amount of time you spent here today was: [Too long] [Reasonable]

Would you like to participate in more studies similar to this one? [Yes] [No]

If so:

- May we contact you to participate in our next study? [Yes] [No]

If not:

- What part(s) of the study should be improved to make it better for other participants? [ ]

Did anything happen that you weren't expecting during the study, or was there anything you found particularly interesting?

[ ]

Is there anything else you'd like to make sure our researchers know before you leave? [ ]

That's it – you're done.
And once again, thank you so much!

"

Survey 5: Main Study Pre-Experiment Survey

Pre-participation questionnaire

The following questions will assist us in determining your eligibility to participate in this study. If you are not eligible to participate, this information will not be stored or shared with anyone.

Whether or not you are eligible to participate will not affect your health care or your relationship with OHSU in any way.

Please just mark the answer or answers which best describe your situation or preferences. You can skip any question you don't feel comfortable answering, though of course we hope you will answer them all. If you feel unsure about what a question is asking, please ask the researcher who gave you this questionnaire to clarify.

For questions in which answers are already provided, please **bold** or highlight your choice.

Age: [ ]
Race: [ ]
Gender: [ ]

Do you enjoy playing computer or video games? [Yes] [No]
If yes:

- What sorts of games do you play? Mark all that apply: [RPG] [Casual Internet] [MMO] [Gambling] [Console] [FPS] [Card] [Teaching/Training] [Puzzle] [Platformer] [Other (please specify):]
  - Please list up to three games that you have played recently: [ ]
- What systems (type of computer, type of console, etc) do you play on? [ ]
- How often do you play, on average? [< 1 hour per week] [1-3 hours per week] [3-7 hours per week] [1-2 hours per day] [> 2 hours per day]

Do you have any vision problems? [Yes] [No]
If yes:

- Please specify the nature of the problem: [ ]
- How is this currently being corrected? [ ]
- Does this correction bring you back to approximately normal vision? [Yes] [No]
  - If not, please describe your current vision: [ ]
- When was your last eye exam? [> 3 years ago] [2-3 years ago] [1-2 years ago] [< 1 year ago]

Was English your first language? [Yes] [No]
If not:

- What was your first language? [ ]
- Are you currently fluent in English? [Yes] [No]

What was the last grade you completed in school? [< 8] [some high school] [high school or equivalent] [some college] [Bachelor's or equivalent] [some grad school] [Master's, PhD, J.D., or other advanced degree]

Survey 6: Main Study Post-UFOV Survey

Post-participation questionnaire

Thank you for participating in our study! If it weren't for people like you, we could never get anything done around here. If you wouldn't mind taking just a few more minutes to tell us about how this might be improved in the future, we would really appreciate it. As always, if you don't understand a question or need clarification on any point, please ask the researcher who gave you this questionnaire. Also, feel free to be as candid as possible; we never judge our participants (after all, to the people receiving this you're just an anonymous number), and negative responses help us learn just as much as positive ones.

Useful Field Of View:

Have you ever taken any version of the UFOV before?

● Yes ○ No

If so, when was the last time you took it? [ ]

Have you ever taken any other tests of divided attention before?

● Yes ○ No

If so:

- What test did you take? [ ]
- When did you take it? [ ]

Did you feel that you understood the test sufficiently after completing the training?

● Yes ○ No

If not, what should have been added to help your understanding of the test?
[ ]

How good do you feel you are at dividing your attention (paying attention to multiple things at once in order to achieve some goal)?

● Excellent

○ Better than average

○ Average

○ Below average

○ Terrible

Do you think you did well on the UFOV? [Yes] [No] [About average] [I have no idea]

● Yes

○ No

○ About average

○ I have no idea

Do you consider yourself a good driver?

● Yes

○ No

○ About average

When was the last time you had a traffic accident or got a ticket for a moving violation?

● < 1 year ago

○ 1-2 years ago

○ 3-4 years ago

○ > 4 years ago

○ I've never had one

When was the last time you had a traffic accident that was determined to be your fault?

● < 1 year ago

○ 1-2 years ago

○ 3-4 years ago

○ > 4 years ago

○ I've never had one

Any other comments or anything you'd like the researchers to know about this part of the test?

[ ]

Survey 7: Main Study Post-Tally Survey and general questions

Post-participation questionnaire

Thank you for participating in our study! If it weren't for people like you, we could never get anything done around here. If you wouldn't mind taking just a few more minutes to tell us about how this might be improved in the future, we would really appreciate it. As always, if you don't understand a question or need clarification on any point, please ask the researcher who gave you this questionnaire. Also, feel free to be as candid as possible; we never judge our participants (after all, to the people receiving this you're just an anonymous number), and negative responses help us learn just as much as positive ones.

For questions in which answers are already provided, please **bold** or highlight your choice.

21 Tally:

Had you already played 21 Tally or a game like it prior to being in the study? [Yes] [No]

Did you feel that you understood the game sufficiently after completing the training? [Yes] [No]

If not, what should have been added to help your understanding of the game?
[ ]

How many training boards would you have preferred? [more boards] [fewer boards] [the number was about right]

Was the overall time required to train you to play appropriate? [I would have liked more time] [I wish it had taken less time] [it was about right]

Did you employ a particular strategy during the game? [Yes] [No]

If so, please briefly describe your strategy: [ ]

Did you enjoy playing this game, purely as a game? [Yes] [No]

If so:

- How often would you play it if it were available? [A few times a month or less] [Maybe once a week] [A few times a week] [Once a day or more]

If not:

- Please briefly describe the differences between this game and the games you play most often: [ ]

What makes the games you play fun? If you ever play games more than once, what keeps you coming back to play them again? [ ]

Purely from playing, would you know that this game was also a cognitive test? [Yes] [No]

If so, does that make it more or less motivating to play, in real life outside the lab? [More motivating] [Less motivating] [No real change]

If this were part of a health solution, which could help keep you cognitively healthy, would that change the amount you would use it? [Yes] [No]

If so, please specify how it would change your potential use: [ ]

Would you like a copy of this game to play at home? [Yes] [No]

If so, would you mind if we continued to monitor your performance for changes in cognitive state (Note: this is for future study information / planning purposes only; we aren't yet set up for remote monitoring)? [I don't want to be monitored, I just like to play the game] [I don't mind being monitored, as long as I get a copy of whatever information you collect] [I don't mind being monitored, but please don't send me the information you're collecting]

General questions:

Overall, did you feel like the amount of time you spent here today was: [Too long] [Reasonable]

Would you like to participate in more studies similar to this one? [Yes] [No]

If so:

- May we contact you to participate in our next study? [Yes] [No]

If not:

- What part(s) of the study should be improved to make it better for other participants? [ ]

Did anything happen that you weren't expecting during the study, or was there anything you found particularly interesting?

[ ]

Is there anything else you'd like to make sure our researchers know before you leave?

[ ]

That's it – you're done.

And once again, thank you so much!

Appendix II: Instructional Text

A dashed line ("------") indicates a break between screens, generally waiting for user input.

Pilot Study UFOV Instructions

Welcome! This tutorial will teach you how to take this test. If at any time you have questions, please call the researcher and ask! Click Start to begin.

During this experiment, please try to keep your head in the same place relative to the screen. Ideally your eyes should be close enough that the top and bottom of the screen are at the edges of your vision, but it is more important that you be a distance away that you can maintain throughout the experiment.

Before each individual trial in this experiment, you will be presented with a small cross in the middle of the screen, like this. Use it as a place to focus your eyes before the trial begins. To keep the experiment time to a minimum, this cross will only show up for a little while; remember to focus on it whenever it appears!

Your first step in this experiment is to remember what image you saw in the center. For example, in this trial, you will be shown either a car or a truck. Then, after some static, you'll be shown both of these images and asked to click on the one that matches the first vehicle you saw. Click Start to begin.

[one demo trial with only one vehicle, after which the user could choose to try an identical trial again, or continue]

After a correct response: Well done! If you'd like to try another example at this speed, press Try Again. If you're ready for a faster example, press Faster.

After an incorrect response: You made an error. Please press Try Again to attempt this type of test one more time.

Over the course of this experiment, some trials will be faster, and others will be slower. For each set of conditions, we'll try to start you out at a reasonable speed and then change it up from there. Press Start for a slightly faster version of the last trial.

[two more demo trials with only one vehicle, after each of which the user could choose to try an identical trial again, or continue]

After a correct response: Excellent work! I think you're ready to move on. If you'd like to try another example at this speed, press Try Again. If you're ready to continue the tutorial, press Continue.

After an incorrect response: You made an error. Please press Try Again to attempt this type of test one more time.

Now, in addition to the vehicle in the center of the screen, another vehicle will appear on the periphery of the screen. Your job is to keep track of both. Note what vehicle appears in the center, and where the vehicle appears on the periphery.

Remember to focus on the cross when it appears. Press Start when you're ready to begin.

[three demo trials with two vehicles, as in standard UFOV]

After a correct response: Excellent work! I think you’re ready to move on. If you’d like to try another example at this speed, press Try Again. If you’re ready to continue the tutorial, press Continue.

After an incorrect response: You made an error. Please press Try Again to attempt this type of test one more time.

Sometimes, over the course of this experiment, you will have other shapes to identify. For example, the peripheral spaces might be filled with circles, and you would need to identify the one with a cross in it.

Let's try a few examples of this type. Press Start, and then focus on the cross, when you're ready to begin.

[three demo trials with a vehicle in the center and eight peripheral images (one target, seven distractors)]

After a correct response: Excellent work! I think you’re ready to move on. If you’d like to try another example at this speed, press Try Again. If you’re ready to continue the tutorial, press Continue.

After an incorrect response: You made an error. Please press Try Again to attempt this type of test one more time.

The last thing we should mention is that in some trials, the center image will be replaced by a letter. They're the same as the trials with a car or truck in the middle in all other respects.

Let's try a few examples with letters. Remember to focus on the cross when it appears! Press Start when you're ready to begin.

[three demo trials with a letter, chosen from the first 10 in the alphabet, in the center, and eight peripheral images (one target, seven distractors)]

After a correct response: Excellent work! I think you’re ready to move on. If you’d like to try another example at this speed, press Try Again. If you’re ready to continue the tutorial, press Continue.

After an incorrect response: You made an error. Please press Try Again to attempt this type of test one more time.

Good job! You made it through all the demos with flying colors. There are just a few differences between the demos and the actual trial...

------

1. During the real experiment, you will not receive feedback about whether your responses are correct or incorrect.

------

2. During most trials, you will go straight from one to the next without having to press Start in between. Only when changing trial conditions will you have to push the Next Trial button.

------

3. During some trials, there may be different numbers of peripheral locations to choose from (more or less than the eight you saw here). In each case, just choose the button closest to the peripheral image you saw.

Finally, remember to keep your head in a similar position throughout the experiment. If you need a break, wait until one of the Start Next Trial buttons appears, and then you can stretch, get up, or whatever you need to do. When you're ready to resume, just put your head back in approximately the same place.

That concludes this tutorial. If you have any further questions, please call over the researcher now and ask. Otherwise, press End, and then Trial Set on the next screen.

Main Study UFOV Instructions

Welcome! During this experiment, you will be asked to perform a series of tasks which involve viewing and identifying images and letters. This tutorial will show you how, and let you get some practice. If at any time you have questions, please call the researcher and ask! Click Start to begin.

When viewing these trial sets, it is important that your head remain in the same position relative to the screen. Please try to keep your eyes 24-30 inches from the center of the screen unless you are taking a break from the experiment.

Before each individual trial in this experiment, you will be presented with a small cross in the middle of the screen, like this. Use it as a place to focus your eyes before the trial begins. To keep the experiment time to a minimum, this cross will only show up for a little while; remember to focus on it whenever it appears!

Your first step in this experiment is to remember what image you saw in the center. For example, in this trial, you will be shown either a car or a truck. Then, after some static, you'll be shown both of these images and asked to click on the one that matches the first vehicle you saw. Click Start to begin.

[one demo trial with only one vehicle]

After a correct response: Well done! Let's try another example of this type.

After an incorrect response: You made an error. Please press Try Again to attempt this type of test one more time.

Over the course of this experiment, some trials will be faster, and others will be slower. For each set of conditions, we'll try to start you out at a reasonable speed and then change it up from there. Press Start for a slightly faster version of the last trial.

[two more demo trials with only one vehicle]

After a correct response: Well done! Let's try another example of this type.

After an incorrect response: You made an error. Please press Try Again to attempt this type of test one more time.

Now, in addition to the vehicle in the center of the screen, another vehicle will appear on the periphery of the screen. Your job is to keep track of both. Note what vehicle appears in the center, and where the vehicle appears on the periphery.

Remember to focus on the cross when it appears. Press Start when you're ready to begin.

[three demo trials with two vehicles, as in standard UFOV]

After a correct response: Well done! Let's try another example of this type.

After an incorrect response: You made an error. Please press Try Again to attempt this type of test one more time.

Sometimes, over the course of this experiment, you will have other shapes to identify. For example, the peripheral spaces might be filled with circles with Xs in them, and you would need to identify the one with the + in it.

Let's try a few examples of this type. Press Start, and then focus on the cross, when you're ready to begin.

[three demo trials with a vehicle in the center and eight peripheral images (one target, seven distractors)]

After a correct response: Well done! Let's try another example of this type.

After an incorrect response: You made an error. Please press Try Again to attempt this type of test one more time.

The last thing we should mention is that in some trials, the center image will be replaced by a letter, and sometimes this letter may be blurred. This blurring is intentional, to try to make the task more challenging. They're the same as the trials with a car or truck in the middle in all other respects.

Let's try a few examples with letters and blurred letters. Remember to focus on the cross when it appears! Press Start when you're ready to begin.

[three demo trials with a 10-pixel-blurred letter in the center and eight peripheral images (one target, seven distractors)]

After a correct response: Well done! Let’s try another example of this type.

After an incorrect response: You made an error. Please press Try Again to attempt this type of test one more time.

Good job! You made it through all the demos with flying colors. There are just a few differences between the demos and the actual trial...

------

1. During the real experiment, you will not receive feedback about whether your responses are correct or incorrect except between trial sets.

------

2. During most trials, you will go straight from one to the next without having to press Start in between. Only when you see the feedback between sets of trial conditions will you have to push the 'Start X/Y' button.

------

3. During some trials, there may be different numbers of peripheral locations to choose from (more or less than the eight you saw here). In each case, just choose the button closest to the peripheral image you saw.

Finally, remember to keep your head in a similar position throughout the experiment. If you need a break, wait until the feedback appears, and then you can stretch, use the restroom, or have something to eat or drink. Just remember to return to the same position before beginning the next trial!

That concludes this tutorial. If you have any further questions, please call over the researcher now and ask. Otherwise, press End, and then Trial Set on the next screen.

Tally Puzzle Instructions

Welcome! This tutorial will teach you how to take this test. If at any time you have questions, please call the researcher and ask! \r\n\r\nClick Start to begin.

[blank board is shown]

This is the two-dimensional blackjack puzzle board. On it there are sixteen places for cards, laid out in a grid of four rows and four columns. There is also a space for the card to be played in the lower left. Press continue to see an example of a puzzle that would have to be solved.

[cards are added to board]

In this puzzle, the goal is to find the best place to play the five of spades in order to make rows or columns summing to 21 (blackjack) and avoid rows or columns summing over 21 (bust). Note that each row and column has its sum total in the black box next to it; aces are worth 1, number cards are worth their number value, and face cards are worth 10.

[worst places are marked in red]

The card to be played is located in the bottom left (in this case, the five of spades). It can be played in any of the open spaces on the board.

In this case, playing the card in the upper left (marked in red) would be the worst possible move, giving you two busts. This would give you -42 points (2 times -21).

[bad places are marked in orange]

Playing it in either of the places marked here in orange wouldn't be much better. The top one would be a bust in the rows, and the bottom one would be a bust in the columns. Each of these would give you -21 points.

[neutral places are marked in blue and okay places in light green]

Playing the card in the spaces marked in dark blue will cause both a bust and a blackjack; this is worth zero net points. The one in light blue is neither a bust nor a blackjack; this is worth zero points as well.\r\nThe space marked in green, by contrast, is a reasonably good move - with one blackjack, you would score 21 points.

[best place is marked in dark green]

Placing it in the dark green space marked here, however, is the best move: this is two blackjacks, worth 42 points. This would be the correct selection for this particular puzzle. When presented with each puzzle, it is important to find the best possible move for that set of cards.

Let's try a few examples now. Press Start when you're ready to begin.

[three demo trials, best play is always a double 21, cards never turn over]

After a correct response: Perfect! Well done. Press Continue to try another example.

After an incorrect response: [identifies error and specifies it, as in] That's a blackjack in the column, but not in the row. There is a better place to play. Let's try another board.

Note that there will not always be a place to play that will score a double blackjack. In these cases, you need to simply identify the best place that can be played. Sometimes this will be a single blackjack, sometimes even just avoiding a bust. There will always be a better place to play than a double bust. Press Start to try some examples.

[three demo trials, best play is not a double 21, cards never turn over]

After a correct response: Perfect! Well done. Press Continue to try another example.

After an incorrect response: You have made an error. Please press Try Again to attempt this type of test one more time.

Excellent work. During the real test, however, you will not have very long to study the board. The cards will be shown for a certain amount of time and then turned face-down while you make your selection. This is considerably more challenging. NOTE: You will be unable to play until the cards have turned face-down! Press Start to try an example.

[three demo trials, cards eventually turn over]

After a correct response: Perfect! Well done. Press Continue to try another example.

After an incorrect response: You have made an error. Please press Try Again to attempt this type of test one more time.

Well done. You have completed all of the demos. Remember that during the real trial, there will be no feedback, and the boards will appear immediately after you've made your selection without waiting for you to press a button.

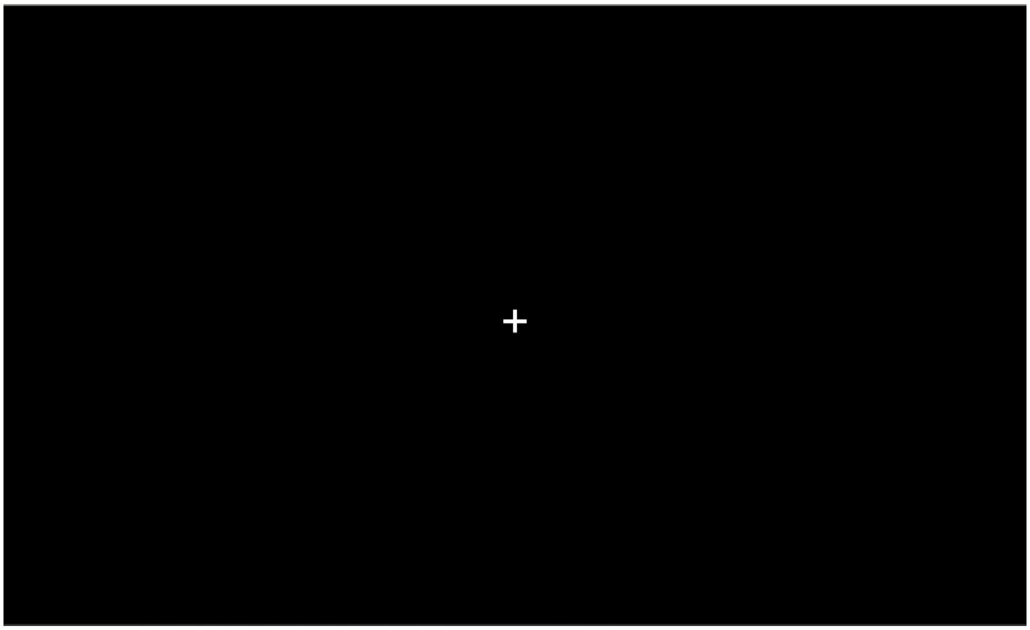
For most trials, you will be asked to select the best overall place to play. If you're not instructed to pay attention to only rows or only columns, assume that you should select the best overall place.

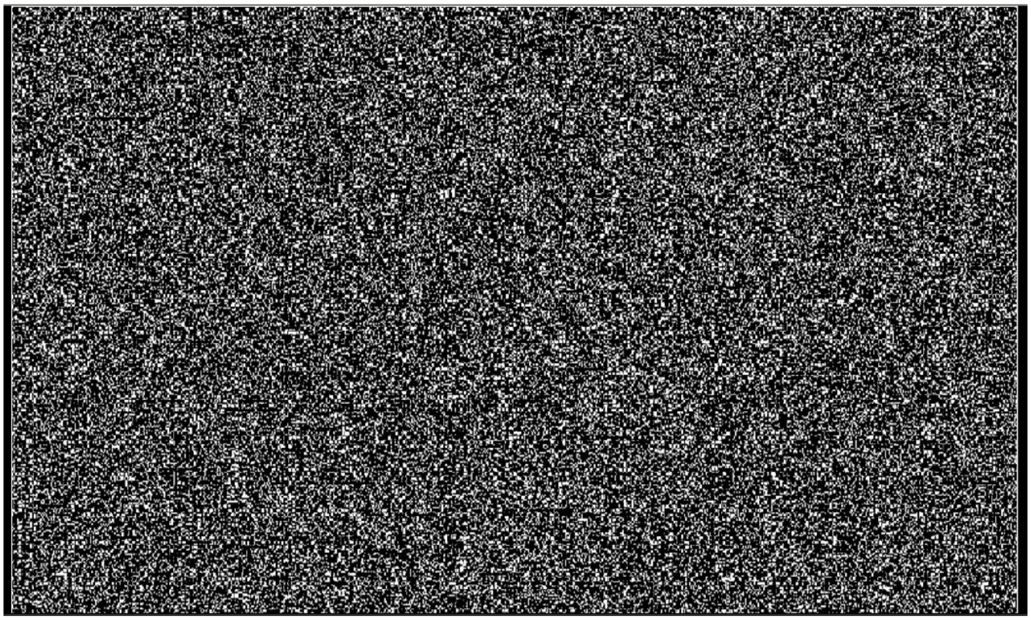
That concludes this tutorial. If you have any further questions, please call over the researcher now and ask. Otherwise, press End, and then Trial Set on the next screen.

Appendix III: Screenshots

All Enhanced UFOV Focal Point and Luminance Equalization

These occurred for a random amount of time between 500 and 1000ms each prior to every trial.

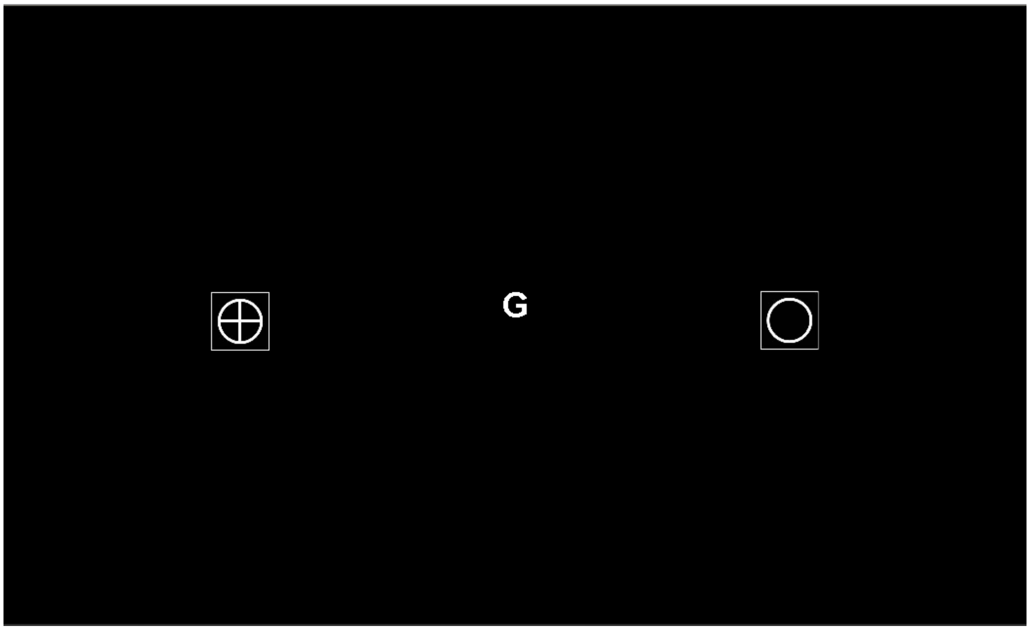


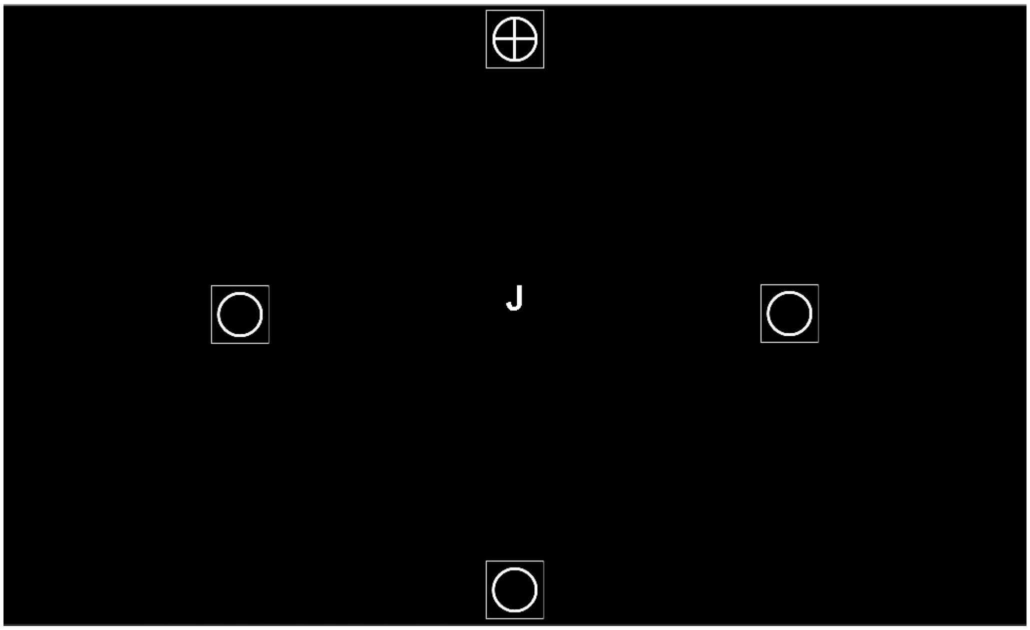


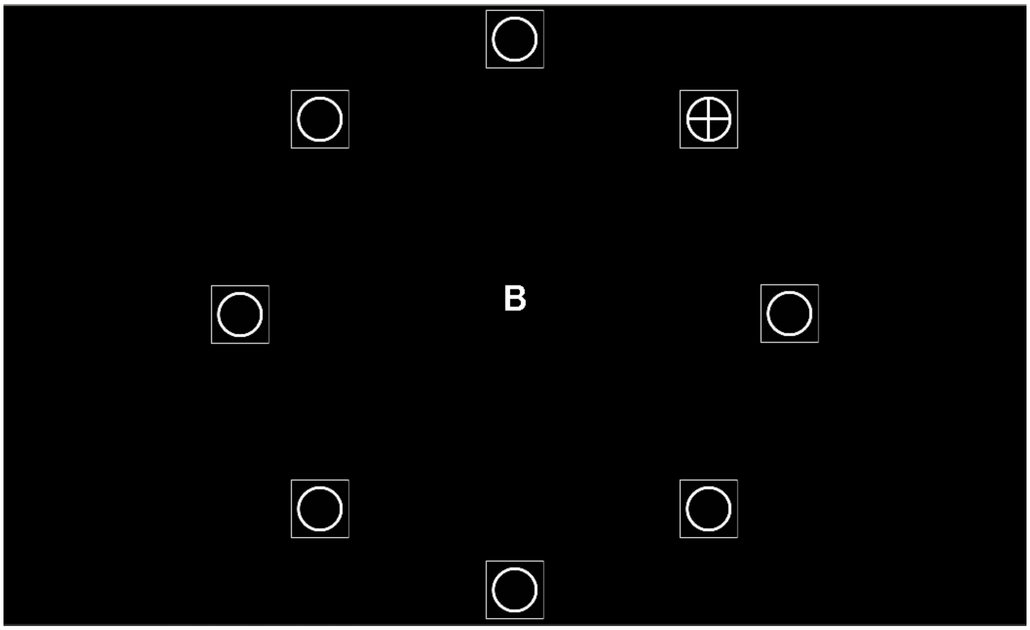
Pilot Study UFOV Screenshots

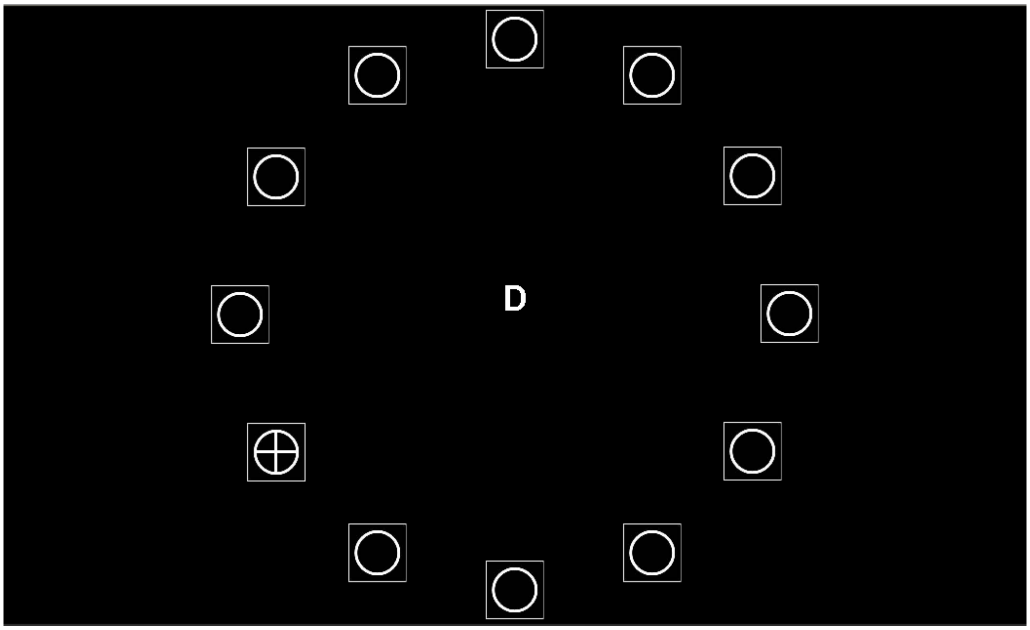
Central letters were picked from the first letters in the alphabet, in groups of 2, 8, and 16. Peripheral icons had 2, 4, 8, 12, or 16 locations.

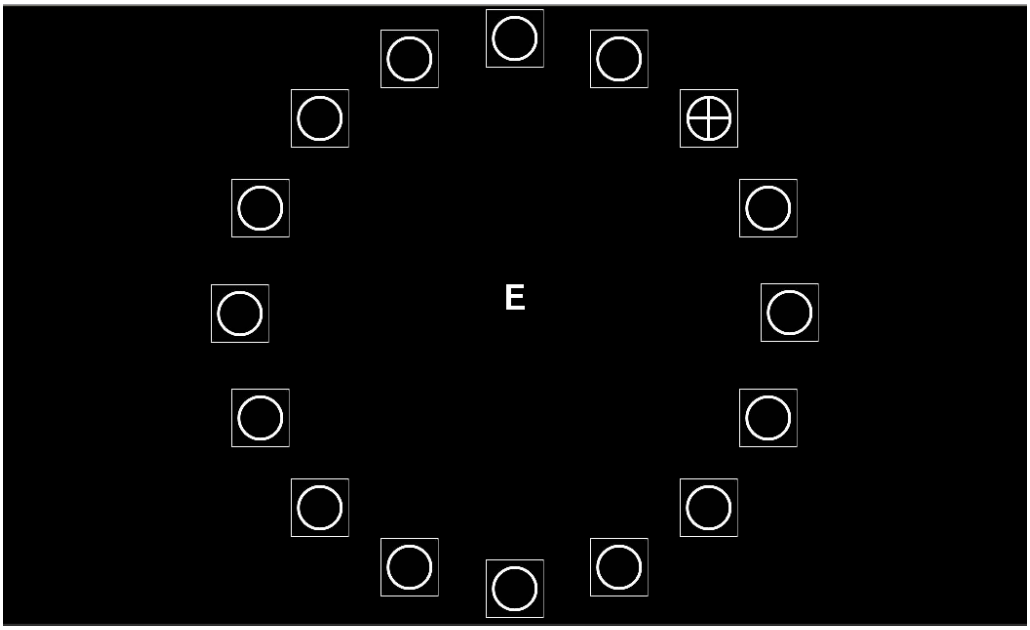
Stimulus example screenshots:



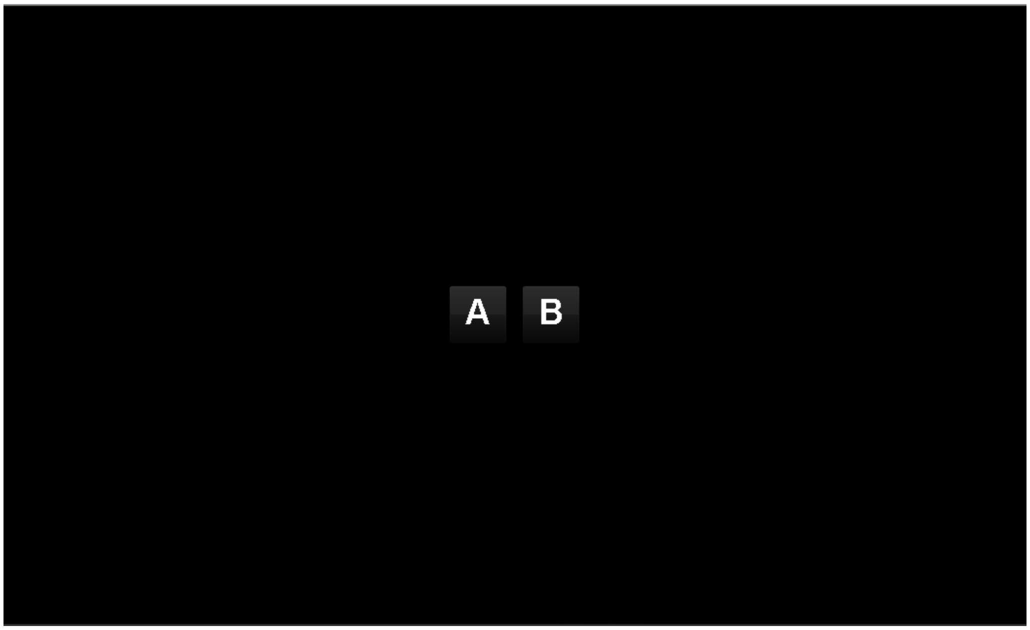

J

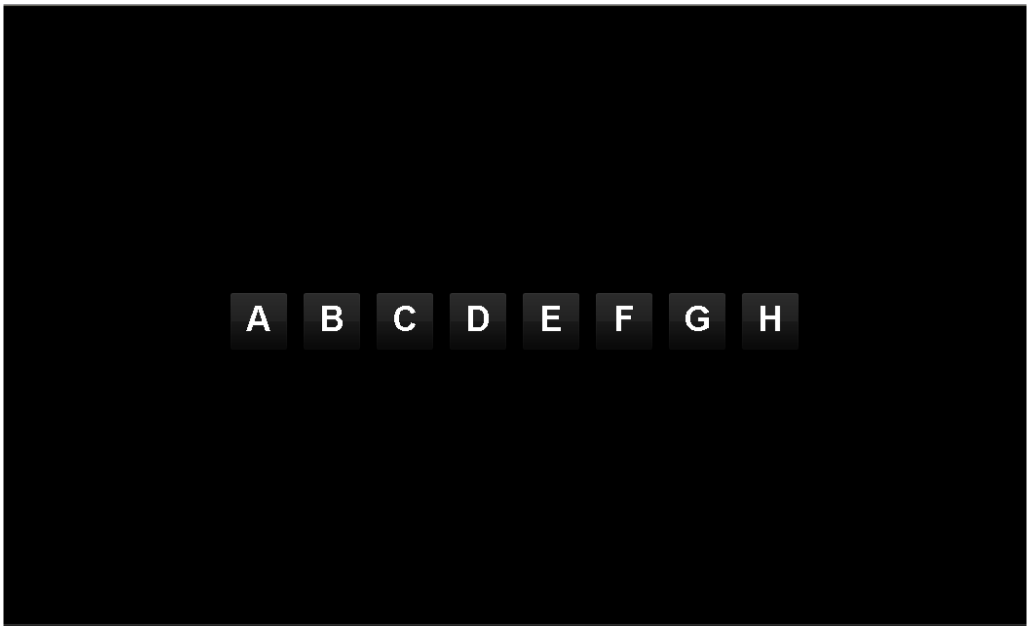

B



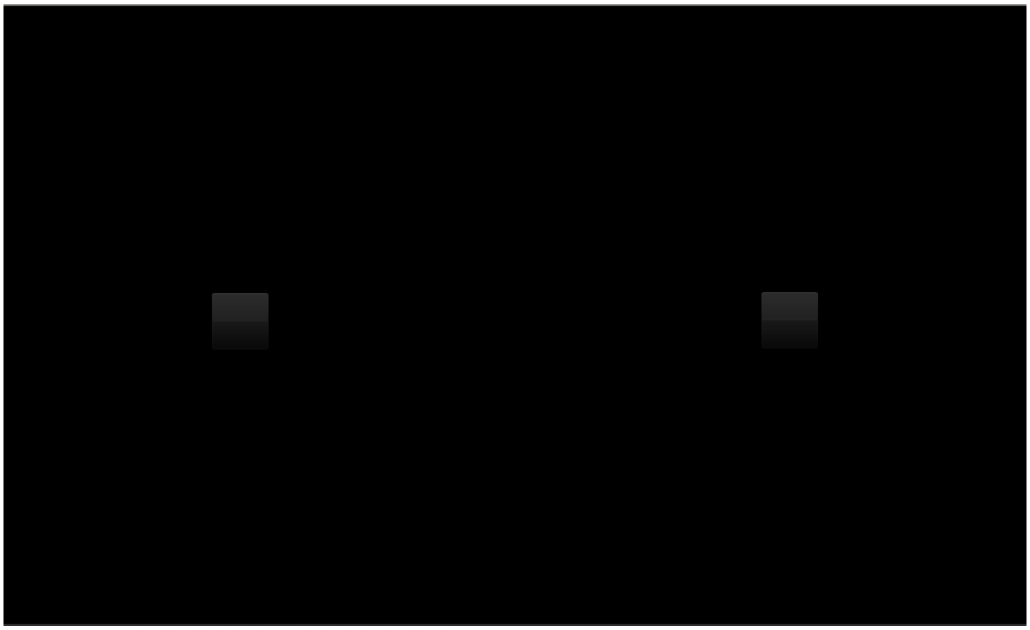


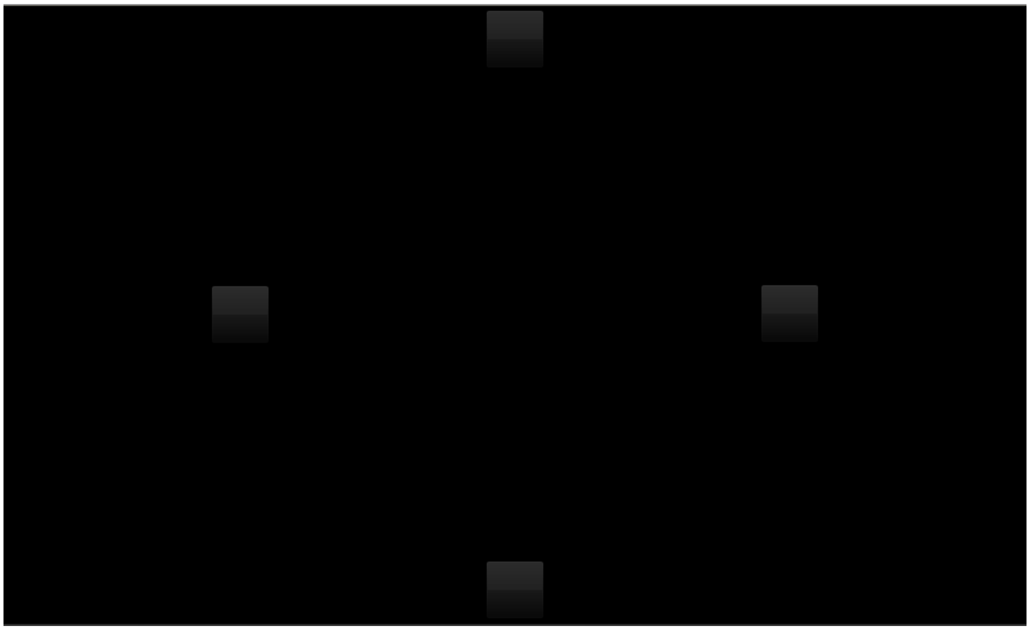
Central response screenshots:

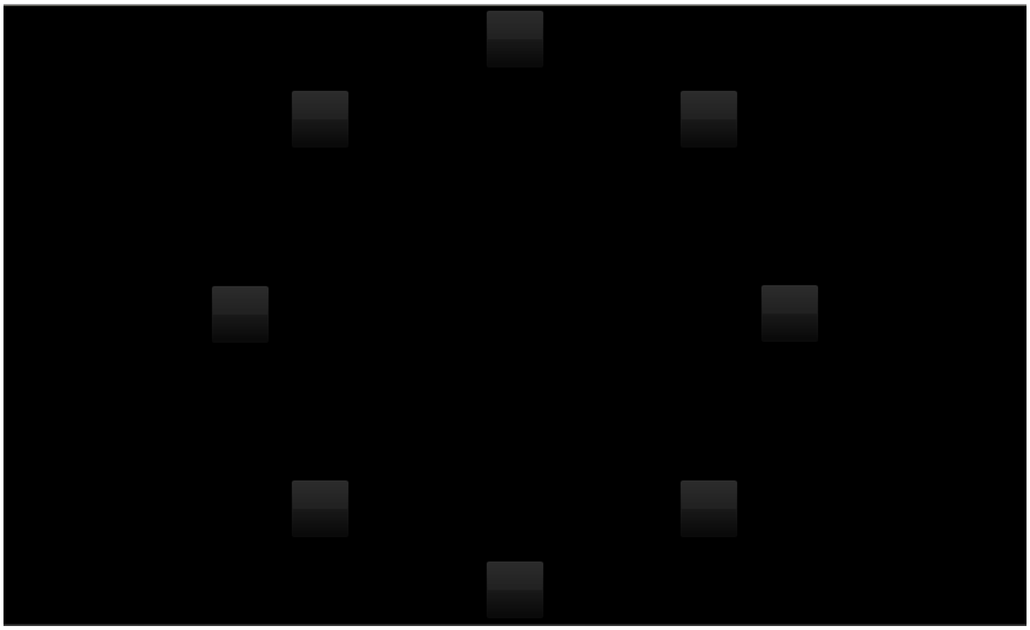

A
B

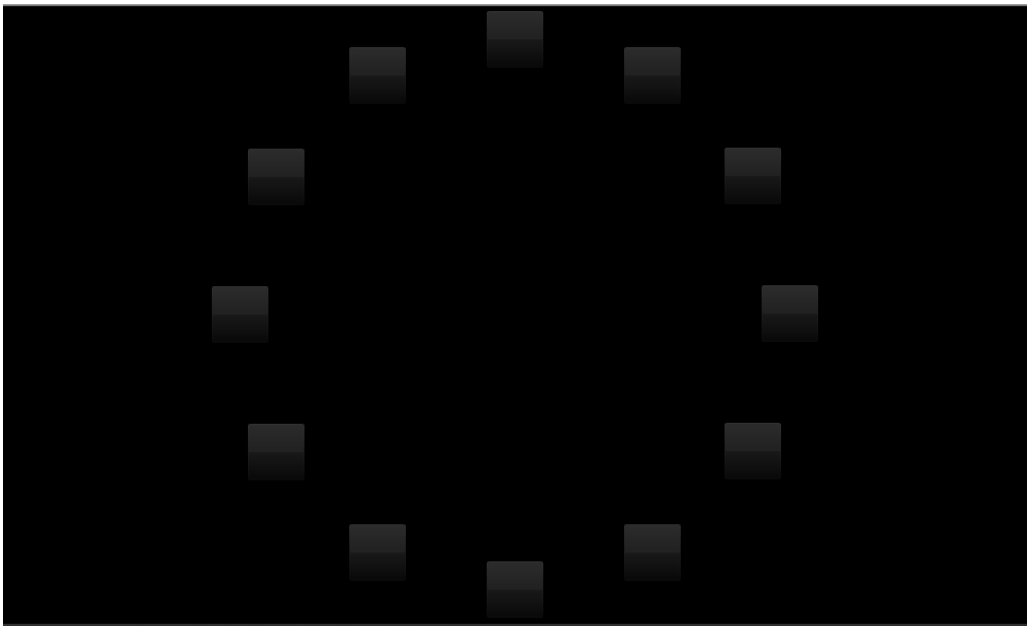

A
B
C
D
E
F
G
H

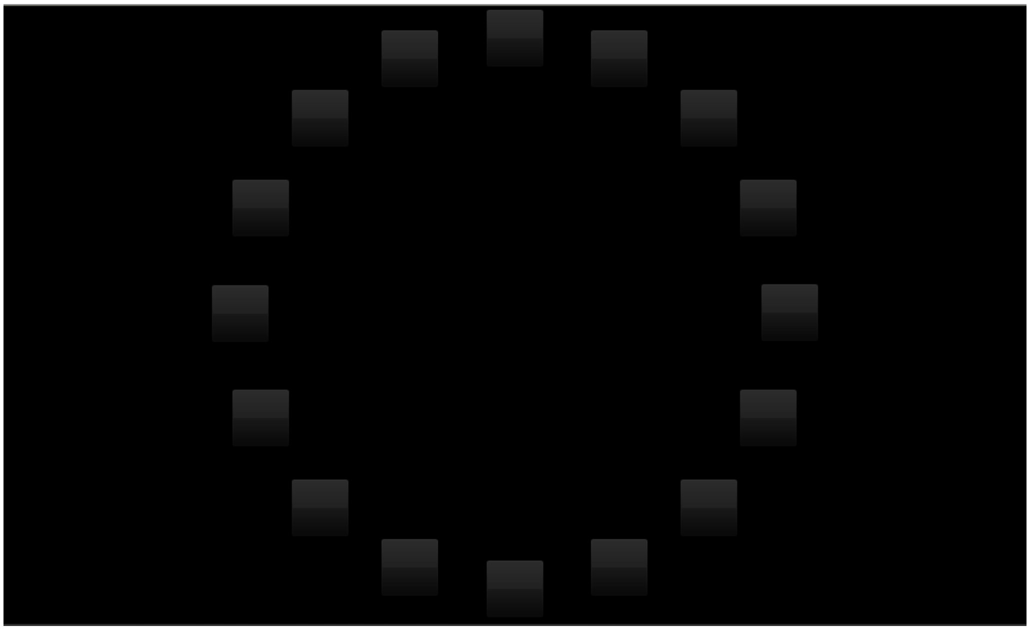
Peripheral response screenshots:





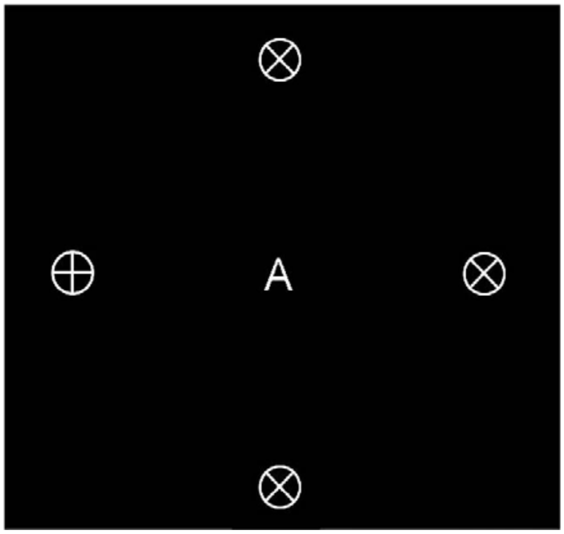


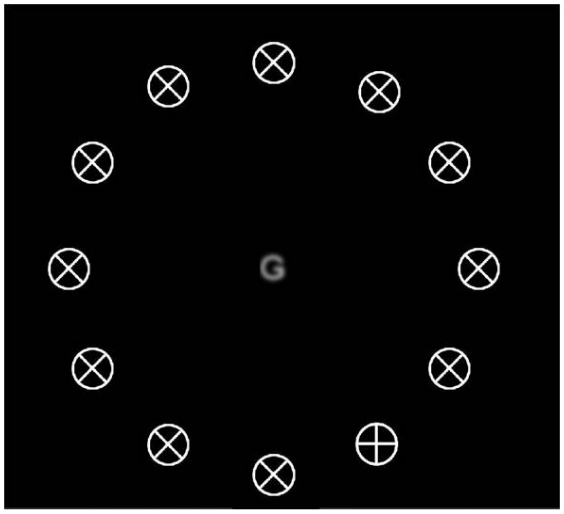




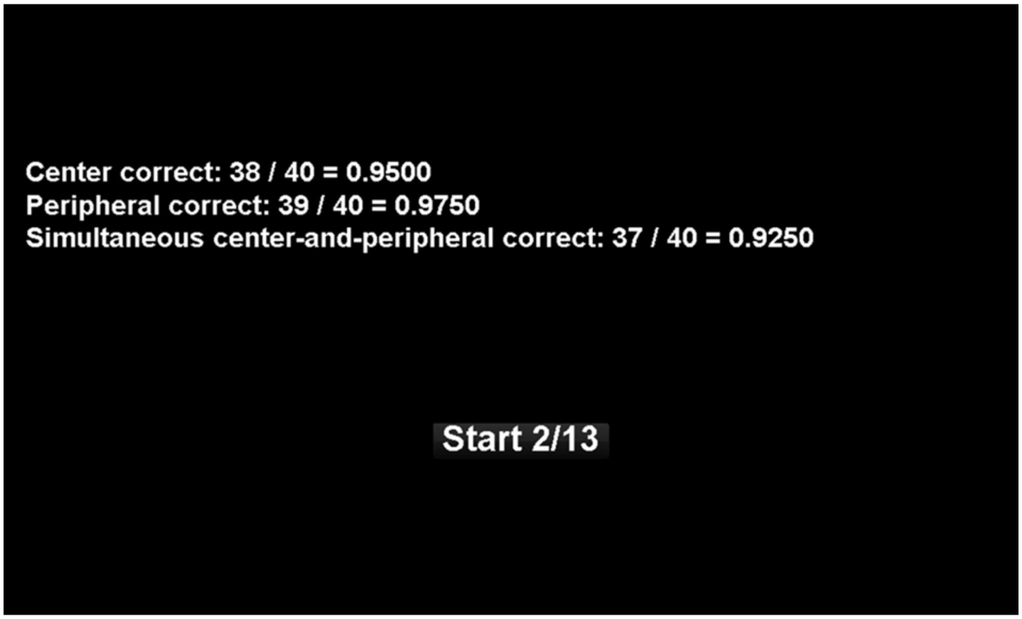
Learning Curve and Main Study UFOV Screenshots

Examples of the two extremes in peripheral and central difficulty:





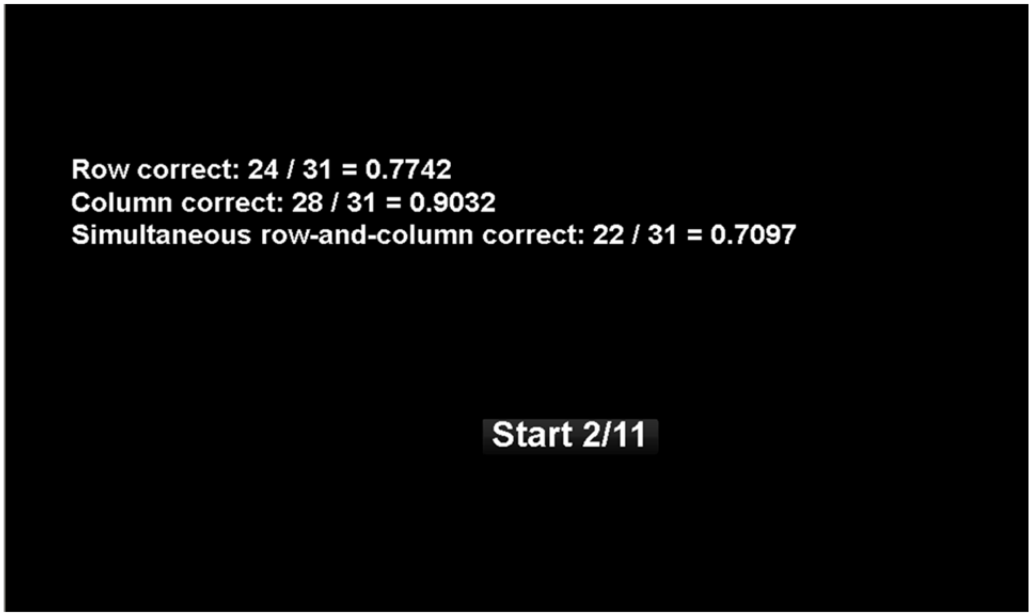
Screenshot of the feedback screen between trials:



Center correct: 38 / 40 = 0.9500
Peripheral correct: 39 / 40 = 0.9750
Simultaneous center-and-peripheral correct: 37 / 40 = 0.9250

Example Tally Screenshots for All Studies

Screenshot of the feedback screen between trials (only showed on Main study):



Row correct: 24 / 31 = 0.7742
Column correct: 28 / 31 = 0.9032
Simultaneous row-and-column correct: 22 / 31 = 0.7097

Appendix IV: recruitment emails

These are two emails used to recruit participants into the study.

For older (65+) participants who had already volunteered to be part of another OHSU study in the “living lab,” or model apartment at OHSU:

Dear <name>,

We would like to extend an invitation for you to participate in a new study testing attention through computer games. You’ve been selected to receive this email because of your association with the OHSU Living Lab studies and your previous interest in our work on coaching for brain health.

Participation in this study will involve coming in to OHSU Center for Health and Healing, being tested for certain aspects of cognition, and then playing a few rounds of certain attention-related computer games. This study will not affect your healthcare or your relationship with OHSU in any way, whether you choose to participate or not. We expect that participation will take no more than two hours of your time. As a thank-you for participating, you will receive a ten-dollar gift certificate to favorite nearby coffee shop.

If you think you might like to be a part of this study, please just reply to this email and one of our researchers will contact you to find out what day and time would work best for you. If not, thanks very much for taking the time to read this email.

Sincerely,

Holly Jimison (Primary Investigator)

James McKanna (Researcher)

For younger (20-65) participants who may not have participated in any research at OHSU previously:

Dear <name>,

We would like to extend an invitation for you to participate in a new study testing cognition through computer games. We're specifically looking for cognitively healthy volunteers between the ages of 18 and 65, who will participate in pre-tests for a study that will later involve older adults. If you have friends or colleagues who might also be interested in this work, please feel free to pass this invitation along.

Participation in this study will involve coming in to OHSU Center for Health and Healing, being tested for certain aspects of cognition, and then playing a few rounds of certain attention-related computer games. This study will not affect your healthcare or your relationship with OHSU in any way, whether you choose to participate or not. We expect that participation will take no more than two hours of your time. As a thank-you for participating, you will receive a five-dollar gift certificate to favorite nearby coffee shop.

If you think you might like to be a part of this study, please just reply to this email and one of our researchers will contact you to find out what day and time would work best for you. If not, thanks very much for taking the time to read this email.

Sincerely,

Holly Jimison (Primary Investigator)

James McKanna (Researcher)

Appendix V: glossary of terms

These are the working definitions of each of these terms as used in this document.

21 Tally: this is a casual computer game designed to help measure divided attention unobtrusively. It involves playing blackjack in two dimensions simultaneously. In this experiment, this term is interchangeable with the Tally Puzzle, though in reality 21 Tally is considerably more game-like and less test-like than the puzzle.

75% accuracy duration: the duration, in milliseconds, that a set of stimuli should be presented in order to give the participant a 75% likelihood of successfully responding to a given trial. This value changes with different conditions and different participants, and is the means by which the UFOV rates divided attention ability.

Enhanced UFOV (eUFOV): this includes all versions of the UFOV that involve nonstandard conditions intended to allow researchers to independently control the difficulty of each task and thereby separate out divided attention ability from speed of processing. Specifically in this experiment, all conditions that involved letters (blurred or otherwise) in the center, or crosses (with or without distractors) on the periphery is considered eUFOV.

Tally Puzzle: also known as Tally, this is the experimental version of the 21 Tally game. In this version, rather than having continuous boards which can be built by the player move-by-move, each board (and thus, each move) is completely independent. This allows us to only provide boards which have the desired number of decisions in either dimension. In this experiment, the Tally Puzzle is presented in trial sets much like the UFOV, and produces similar 75% accuracy durations.

Trial: a single presentation of stimuli and the accompanying responses from the participant.

Trial set: a series of trials presented using a single set of conditions, continuing via a staircase algorithm until the termination condition was met. In this experiment, as in the UFOV, each trial set produced a 75% accuracy duration.

UFOV: Useful Field of View test, a standard test designed to measure driving ability in elders. It consists of four subtests, each of which is a single trial set with different conditions. The final score for the UFOV is the sum of each trial set score. See the Background section on UFOV for more information.

UFOV subtest 2 (UFOVs2): The Useful Field of View, subtest 2. This subtest of the UFOV is specifically intended to measure divided attention. It is a single trial set in which the conditions consist of images of automobiles. On each trial, the participant must identify which image appeared in the center and where the image appeared on the periphery.