THE TEST-TAKING PUPIL: EFFECTS OF DEPLETION, DIFFICULTY, AND THREAT ON PUPIL RESPONSIVITY

by

B. Brian Kuhlman

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Educational Psychology

The University of Utah

August 2014

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **B. Brian Kuhlman**

has been approved by the following supervisory committee members:

Michael Gardner	, Chair	**April 11, 2014** Date Approved
Jonathan Butner	, Member	**April 11, 2014** Date Approved
Kirsten Butcher	, Member	**April 11, 2014** Date Approved
Zac Imel	, Member	**April 11, 2014** Date Approved
Robert Zheng	, Member	**April 11, 2014** Date Approved

and by **Anne Cook** , Chair of

the Department of **Educational Psychology**

and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

Pupil dilation measures provide a useful index of test-taking processes. Prior research has established a simple positive relationship between pupil dilation magnitude and (i) threat levels, (ii) task difficulty levels, and (iii) working memory capacity. Surprisingly few studies have investigated the interaction of these three pupil response drivers. Do they add in a linear fashion, like separate weights on a single scale (as the “load” metaphor suggests), or is their relationship more complicated? To test of this question, I used a 2 X (2 X 3) mixed experimental design with random assignment to working memory resource depletion and nondepletion groups. These groups completed two versions of the same task, where response inhibition is required repeatedly in the depleting but is not required in the nondepleting version. Next, all subjects completed a test (90 factor-multiple judgment items) that employed two levels of difficulty (easy and difficult) and three levels of threat (safe, partially cued threat, and fully cued threat). Test-taking pupil data were collected at 60 Hz using a Tobii eye-tracker. Results indicated that levels of threat and task difficulty independently contribute to pupil response magnitude and they do not moderate one another. Apparently, the effects of difficulty and threat are not moderated by resource depletion; however, this study lacked power to detect anything less than a strong depletion effect. Results indicate that test-taking pupil responses are sensitive to testing conditions (e.g., threat and difficulty), but it remains unclear whether these responses are also sensitive to priming conditions (e.g., resource depletion).

Every active intellectual process, every psychical effort, every exertion of attention, every active mental image, regardless of content, particularly every affect just as truly produces pupil enlargement as does every sensory stimulus.

Oswald Bumke (1911; translated in Hess, 1975, pp. 23-24)

Any sensory occurrence - whether tactile, auditory, gustatory, olfactory, or noxious - evokes a pupillary reflex dilation. Exceptions to this are light stimuli and accommodations to near visual stimuli, both of which produce pupil constrictions. However, one should not assume that pupillary reflex dilations occur only to external sensory events, because emotions, mental processes, increases in intentional efforts, and motor outputs also produce systematic changes in pupillary diameter.

Beatty and Lucero-Wagoner (2000, pp. 145-146)

CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

ACKNOWLEDGMENTS

The dissertation is done and it feels like a page is turning. As I reflect on this life chapter, a number of important people come to mind.

First, I thank my professors. Throughout this project, Drs. Mike Gardner, Kirsten Butcher, Jon Butner, Zac Imel, and Robert Zheng challenged my thinking and guided my work with care and brilliance. Similarly, courses taught by Drs. Anne Cook, Doug Hacker, and Dan Woltz were formative experiences that will not be forgotten.

I am grateful to my ever-supportive family – Mom, Dad, Casey, Leigh, Cameron, Ryan, Uncle Don, Lance, Beth, Don Fabian, Dorita, Maria, Nano, and Frodo – you all feel like immediate family (and some of you are). I admire each of you for too many reasons to list, and I treasure our times together.

I am so happy that my wife Lucy took her seat next to mine in this wild ride. Her charm and companionship are my greatest delights. Just as my dissertation is wrapping up, her dissertation is about to get going. While I doubt hers will be as concise as mine, I hope it will be as rewarding.

I dedicate this dissertation to Dr. John Kircher, whose support and influence have been paramount. John taught me to think like a scientist, but also to laugh in the midst of frustration and persevere through failures. I hope all my upcoming life chapters will make him proud.

CHAPTER 1

INTRODUCTION

Across a century of pupillometry research, it has proven difficult to disentangle the many sources of pupil responsivity, yet that is the purpose of the current study. This chapter begins with a discussion of what others have called “cognitive pupillometry”: studies of differential pupil responsivity to manipulations of cognitive load (e.g., reasoning tasks). The discussion then focuses on what I am calling “affective pupillometry”: studies of differential pupil responsivity to manipulations of affective load (e.g., fear conditioning). The discussion will emphasize the consistent finding that pupil responses increase with levels of threat, task difficulty, and working memory resources. Finally, the focus will turn to an under-studied question in the field: How do these drivers of pupil responsivity (i.e., threat, task difficulty, and working memory resources) interact?

A terminological note is called-for at the onset of this discussion. The acronym EPR will be used extensively throughout this document. It stands for “evoked pupil response” which is the preferred way to reference changes in pupil diameter across time due to cognitive and/or emotional factors.

Cognitive Pupillometry

Schiff (1875; Schiff & Foa, 1874) pioneered this field of research by demonstrating that pupil dilations could be evoked by a variety of nonvisual stimuli. Soon thereafter, Heinrich (1896) measured pupillary dilations evoked by mental multiplication tasks. This fostered scientific attention, yet, as Beatty and Lucero-Wagner (2000) point out, interest in task-evoked pupil responses (task-EPRs) failed for nearly a hundred years to spread beyond the German neurological community. This changed when Hess and Polt (1964) extended Heinrich's (1896) finding that mental arithmetic items evoke pupillary dilations. Of particular interest was their finding that item difficulty moderated pupil response magnitude, such that easier items (e.g., 7 x 8) evoked smaller pupil dilations than harder items (e.g., 13 x 14). Hess and Polt's (1964) publication in the journal *Science* popularized cognitive pupillometry as a field of research.

Across the next few decades, Hess and Polt's findings were extended. For example, Kahneman and Beatty (1966) published evidence of 0.1 mm pupil responses to easier and 0.5 mm pupil responses to harder digit transformation and digit span tasks. Figure 1 shows the summary graph from Beatty's (1982) *Psychological Bulletin* meta-analysis of the relation between task difficulty and EPR magnitude across memory, perception, reasoning, and language task domains. Later, in one of their many reading cognition studies, Just and Carpenter (1993) found that EPRs are sensitive even to levels of syntactic complexity; sentences that are easier to read evoked smaller pupil responses than syntactically complex (difficult to read) sentences. More recently, in a lucid set of studies using modern gaze tracking pupillometers, Klingner (2011) replicated a number of seminal findings from cognitive pupillometry. These included Kahneman and Beatty's (1966, see above) digit span task findings, as well as Hess and Polt's (1964) mental multiplication findings (see Figure 2).

While task difficulty is perhaps the best-researched driver of evoked pupil responsivity, a smaller number of studies have confirmed the related contribution of working memory capacity (WMC), an individual difference factor that measures the maximum amount of information one can use at a given time on a given task. Goldinger,

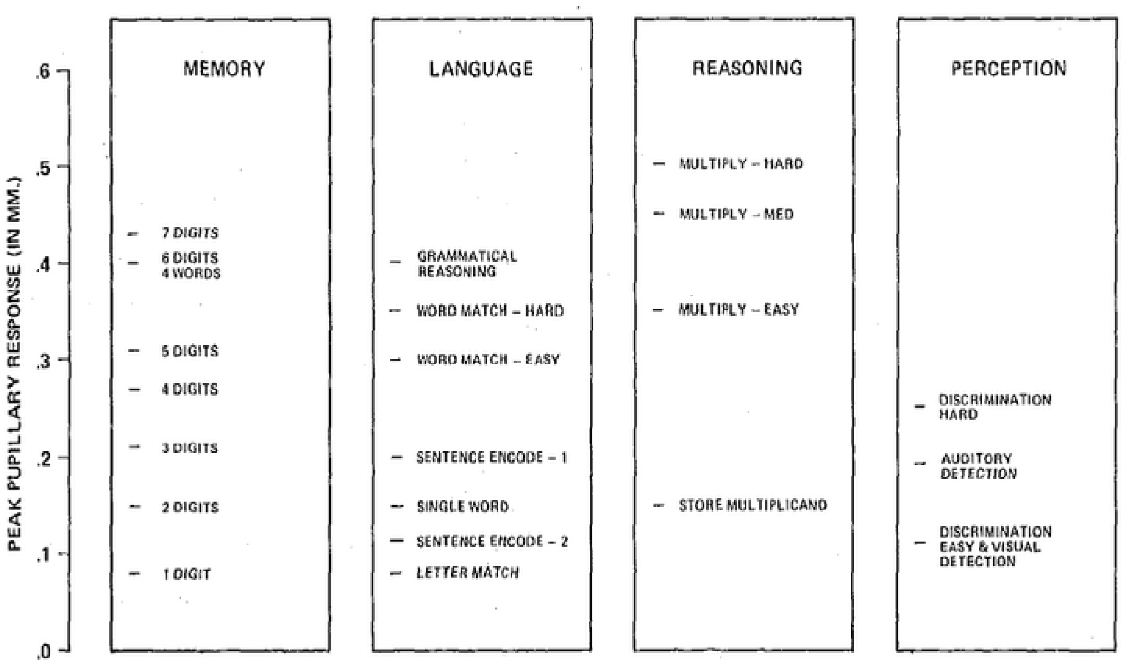

Figure 1. **Beatty's (1982) Meta-Analytic Summary of Task Difficulty X EPR Magnitude Findings Across a Variety of Task Domains.** Reprinted with permission. Taken from page 285 of original work.

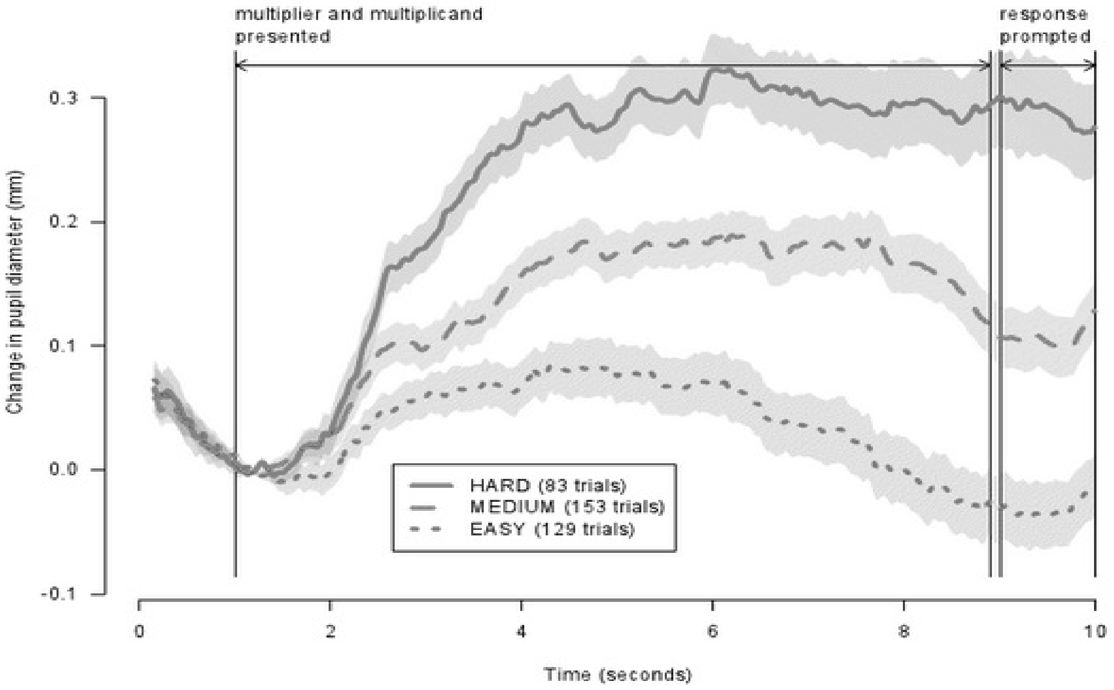

Figure 2. **Klingner's (2011) Replication of Earlier Task Difficulty x EPR Magnitude Findings Using Mental Multiplication Items.** Reprinted with permission. Taken from page 54 of original work. Lighter-colored grey areas indicate standard errors of response averages.

He, and Papesh (2009) found, for example, that participants who were better at encoding and remembering faces that are of a different race than their own were also higher in WMC, presumably because this task requires the simultaneous use of many details about the encoded faces. At the beginning of this experiment, an eye tracker measured participants' pupil responses while they memorized faces that appeared on the computer screen. Both Asian and Caucasian participants exhibited bigger task-EPRs when encoding cross-race faces than when encoding same-race faces, an effect they explained by the greater difficulty of encoding cross-race faces. Importantly, they found that participants' performance on a subsequent test of facial memory was predicted by the average magnitude of their EPRs during facial encoding; bigger EPRs preceded better memory performance. The researchers explained this finding by suggesting that EPR magnitudes are an index of WMC.

As an individual difference variable, working memory capacity has been of great interest to cognitive scientists in large part because of its strong positive relationship with general (Conway, Kane, & Engle, 2003) and fluid (Kane, Hambrick, & Conway, 2005) intelligence. Meer and colleagues (2010) used geometric analogy tasks to demonstrate the independence of task difficulty and fluid IQ as drivers of EPR magnitude. Geometric analogies are a measure of fluid IQ; the analogy tasks vary from simple to very complex and fluid IQ is measured as a function of accuracy at different levels of task complexity. In the study, all participants were given the same analogy tasks to complete. Meer's team found that task difficulty is a driver of EPR magnitude: all participants had bigger task-EPRs during difficult analogy tasks. They also found that stronger performers had bigger task-EPRs. The researchers explained this finding by suggesting that EPR magnitudes are an index of fluid IQ. Whether labeled fluid IQ or WMC, it appears that one's cognitive processing ability acts as a type of "ceiling" on task-EPR magnitude.

Granholm, Asarnow, Sarkin, and Dykes (1996) published the exception that proves the task-EPR rule. Three levels of digit span difficulty (5-span for moderate load, 9-span for high load, 13-span for excessive load) yielded task-EPRs that increased until working memory hit its typical capacity (7 plus or minus 2 digits) and then decreased when the task demands exceeded WMC. Task-EPRs were greater to the 9-span than to the 5-span or 13-span tasks. This suggests that task-EPRs (i) increase with below-limit processing demands [from 1 to 5 digits], (ii) change little during processing at or near the limit [7 plus or minus 2 digits], and (iii) decline after processing demands exceed the limit of working memory resources [from 9 to 13 digits]. Thus, pupil responses appear to index *decreases* as well as increases in cognitive effort.

Broadly speaking, the great discovery of cognitive pupillometry is that task-EPRs provide a sensitive index of cognitive load. Manipulations of task difficulty and individual differences in working memory capacity -- two key variables in cognitive research -- both display a strong positive relationship with task-EPR magnitude. In addition to indexing cognitive load, pupil responses are sensitive to the emotional value of tasks and conditions. The effect of threat on evoked pupil responses is the subject of the next section.

Affective Pupillometry

Prior to their seminal 1964 publication in *Science*, Hess and Polt (1960) published evidence of differential pupil responsivity to manipulations of stimulus valence. Their discussion posited a “bidirectional” relationship between emotional valence and pupil response, where rewarding (e.g., sexually-arousing) imagery evoked pupil *dilations* and aversive (e.g., threatening) imagery evoked pupil *constrictions*.

Later researchers concluded that Hess and Polt were wrong about aversive stimuli; like rewarding stimuli, aversive stimuli dilate the pupil more than neutral stimuli.

Across a number of studies, threatening conditions have simultaneously increased baseline pupil diameter and suppressed pupil constrictions to a flash of light (see, e.g., Hourdaki et al., 2003). Nunnally, Knott, Duchnowski, and Parker (1967) informed subjects that the digits 1 to 5 would be presented in sequence, and that a loud gunshot (acoustic startle) would fire during the digit 3. Pupil size increased during fearful anticipation (digits 1-3) and decreased following the acoustic startle (digits 4-5). Similar data on fearful anticipation were published throughout the 1960s and 1970s (see review by Stanners, Coulter, Sweet, and Murphy, 1979). Pain-induced arousal also intensifies pupil dilations. Chapman, Oka, Bradshaw, Jacobson, and Donaldson (1999) used four intensities of fingertip electrical stimulation (ranging from very faint to barely tolerable). As the electrical stimulation increased, so did the participants' pupil responses (from 0.25 mm to 0.37 mm, respectively).

The field of deception detection – especially Dr. John Kircher's lab – also informs how aversive stimuli increase pupil responses. For example, Webb's (2008) dissertation indicated that Guilt (of stealing $20 cash in an experimental task) and Question Type factors interacted in driving pupil response magnitude. Her guilty subjects' responses (see Figure 3) suggest that guilty test-takers found the relevant crime (cash theft) items more aversive than the irrelevant crime (exam theft) or the general world knowledge (neutral) test items.

Question Type was the most salient driver (eta^2 = 0.467) in Webb's study. Why? There is no reason to believe that Webb's crime-related questions were more difficult to answer than her general world knowledge (i.e., Neutral) questions, but the question types certainly differ in emotional value. Webb's crime items may represent an interesting type of threat. Cash items such as "It is true that I stole the $20 from the secretary's purse. T or F" imply a threat to the subject's standing as a good citizen, whereas Neutral items such as "Moses took two of each kind of animal into the ark. T or

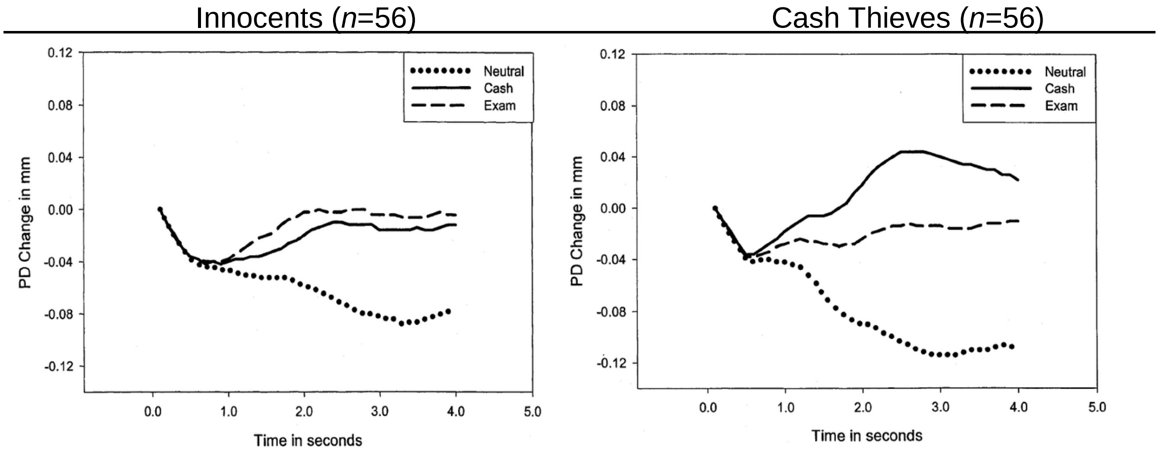

Figure 3. **Webb's (2008) EPR Findings from Guilty and Innocent Subjects During a Computerized Deception Test.** Reprinted with the author's expressed permission. Taken from pages 46 and 47 of the original work. Lines represent EPRs during three question types: "Neutral" = general world knowledge; "Cash" = involvement in a relevant crime (cash theft); "Exam" = involvement in an irrelevant crime (exam theft). "PD" = pupil diameter. *Note:* Due to a misprint, the Y-axes on these diagrams are incorrect; they are one order of magnitude too small (the decimal point should be moved one position to the right).

F" do not imply any threat. Subjects are aware of this threat without the items themselves containing any direct threat cues. The threat is partially cued or implied by the test items. In sum, Webb's (2008) large effect of Question Type appears to reflect a fear-evoked pupil response (fear-EPR), where the threat is partially cued.

No study to-date has compared the effects of partially- and fully cued threats on EPRs. As discussed above, the fearful-anticipation study by Nunnally and colleagues (1967) is a good example of fully cued threat conditions; subjects knew to anticipate the acoustic startle when the count reached three. Webb's (2008) relevant crime test items are a good example of partially cued threat conditions; subjects were not explicitly told that their good social standing was at risk, but the use of phrases like "It is true that I stole…" implied the threat. A systematic comparison of these types of threat-cues could lead to deeper understanding within the field of affective pupillometry.

Broadly speaking, affective scientists can use fear-EPRs to index affective load just as cognitive scientists use task-EPRs to index cognitive load. By extension, pupil responses should provide a useful index of interactions between cognitive and affective load, as discussed in the next section.

Combinations of Affective and Cognitive Load

The widely used "load" metaphor (as in cognitive or affective load) suggests that responses to difficulty and threat should add in a linear fashion, as if they were two independent weights placed onto a single scale. But perhaps the situation is more complicated than this metaphor suggests. To test this, threat and difficulty conditions must be experimentally mixed.

Relatively few studies have used threat-difficulty combinations. However, the work to-date suggests two main findings. First, engagement in a difficult task can prevent fear conditioning. Exemplars of this finding include studies by Carter, Hofstotter, Tsuchiya, and Koch (2003) and by Straube and colleagues (2011), which tested for fear conditioning in the context of working memory testing. Indeed, their subjects failed to respond to threat cues while engaged in the working memory task. It is noteworthy, however, that these studies paired threat cues (conditioned stimuli) with aversive (unconditioned) stimuli in a manner that was *unrelated* to the working memory task; moreover, successful performance on the task *required* the subject to ignore threat cues and aversive stimuli. While it is interesting that working memory engagement prevents conditioning to task-irrelevant fear cues, this finding does not suggest that fearful responding is incompatible with thoughtful responding.

Second, under the right conditions, the effects of task difficulty can independently add to (instead of prevent or reduce) the effects of threat. This appears to happen when task demands are relevant to the threat contingency. For example, Polt (1970; see

Figure 4) presented two groups (7 subjects each) with two tests (three multiplication problems each). All six items matched the medium and high difficulty levels used in other mental multiplication studies (see Figures 1 and 2). During the first test, the groups were under identical (no threat) conditions and had similar task-EPRs. In the second test, which was equal in difficulty to the first test, the experimental group was threatened with electric shock for an incorrect answer but the control group was not. During this test, the experimental group's average dilations (0.75 mm) were nearly twice as large as nonthreatened controls (0.42 mm). This response level is higher than the typical range of fear-EPR and task-EPR magnitudes discussed above. Polt's findings suggest that threat and difficulty independently contribute to EPR magnitudes.

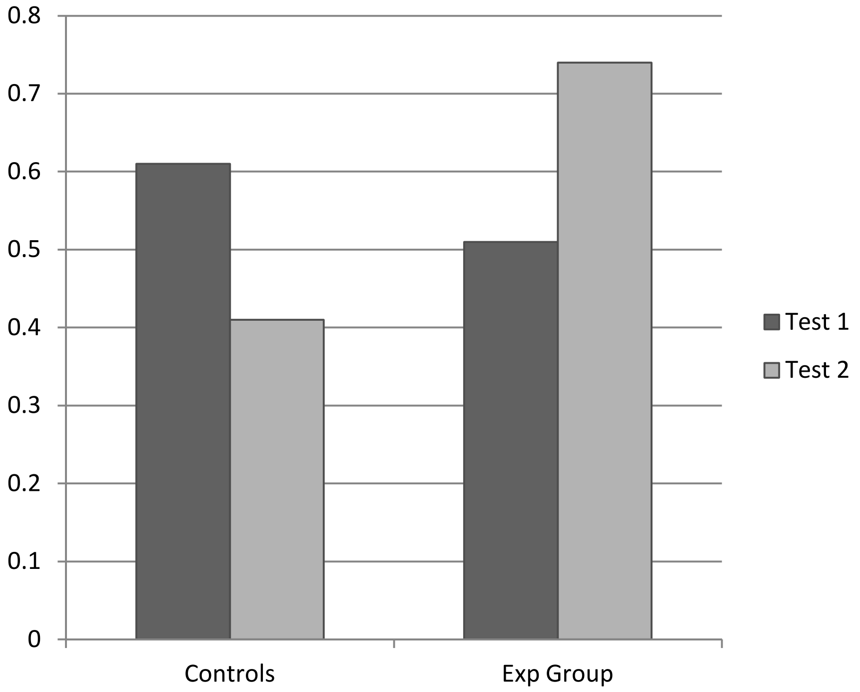

Figure 4. **Polt's (1970) Comparison of Task-EPRs with a Fearful Task-EPR.** These data were aggregated from Table 1 on page 590 of the original work, but this figure did not appear in the paper. The far-right bar represents the fearful task-EPR condition. All three other bars reflect task-EPRs.

Resource Depletion, Difficulty, and Threat

We have already discussed the fact that EPRs increase with task demands up to the subject's capacity for managing working memory load – for example, digit spans of 7 plus or minus 2 – and then decrease after task demands exceed that capacity limit (Granholm, Asarnow, Sarkin, & Dykes, 1996). In theory, working memory resource capacity can be decreased using the dual task "resource depletion" approach pioneered by Baumeister and colleagues (1994). In this approach, the experimental group completes a depleting task while the control group completes a nondepleting task, and then both groups are given the same Challenge Task (e.g., a difficult arithmetic test, a cold-presser test). A recent meta-analysis (Hagger, Wood, Stiff, & Chatzisarantis, 2010) found that dual task resource depletion studies typically yield moderate-to-large between-group mean differences (average d = 0.62) across a variety of dependent variables. During the Challenge Task, depleted subjects are typically less persistent and less accurate than nondepleted controls (see e.g., Johns et al., 2008; Stewart et al., 2009; Vohs et al., 2005; Vohs et al., 2008).

While main effects of depletion on persistence and accuracy are robust across Challenge Task conditions, depletion effects on physiological reactivity are sensitive to Challenge Task conditions. Cardiovascular (CV; e.g., blood pressure and heart rate) responses to the Challenge Task, for example, appear to be sensitive to threat-by-resource interactions. For example, Wright and colleagues (2007, Experiment 1) required participants to perform either an easy (nondepleting) or a difficult (depleting) counting task and then presented them with an arithmetic Challenge Task. All participants were informed that they could avoid an aversive noise if they attained a modest (50th percentile) performance standard on the Challenge Task. The depletion group had stronger cardiovascular responses during the Challenge Task than the nondepletion group (Wright et al., 2007). Taken alone this finding reflects a main effect of depletion on CV responding, but the finding failed to replicate across variations in Challenge Task conditions. According to two of Wright's collaborative multi-experiment papers (Wright et al., 2003; Wright et al., 2007), depleted-subject cardiovascular responsivity *increases* under easy, high reward expectancy and threat conditions but *decreases* under difficult and low reward expectancy conditions. Apparently, the main effects of depletion on CV responsivity are modulated by Challenge Task conditions. While it is too early to conduct such a study, I suspect that future meta-analyses will find that depletion effects on autonomic responses are fully mediated by Challenge Task conditions.

Among the resource depletion manipulations reviewed in their meta-analysis, Hagger and colleagues (2010) found that aversive social interactions (e.g., the social stress test) yield some of the strongest main effects on task persistence and accuracy. While none of these meta-analyzed studies used a mock crime as the resource depletion manipulation, mock crime tasks typically involve aversive social interactions. For example, Webb (2008) required her guilty-condition subjects to have a conversation with a female secretary-confederate prior to stealing a $20 bill from her purse. A number of subjects dropped out of Webb's study – forfeiting the monetary participation-reward – when they were assigned to the guilty condition. Because her mock crime involves an aversive social interaction, it is interesting that Webb found no main effect of Guilt on pupil responses during her Challenge Task (a computerized deception test). Across the Challenge Task as a whole, average guilty and innocent pupil responses were equivalent. This finding suggests that pupil responsivity may be robust to main effects of resource depletion. Yet Webb's detection of a strong Guilt by Question Type interaction (see Figure 3) suggests that resource depletion sensitizes participants to relevant threats. Perhaps resource depletion does not directly affect biomarkers of mental activity (e.g., CV responses and EPRs) but rather alters one's *sensitivity* to different environmental conditions.

Recall Polt's (1970) findings, which suggest that pupil circuitry treats difficulty and threat as if they were two weights on a single scale. It is unknown whether this pattern holds under depleted conditions. Webb's (2008) mock crime experiment suggests that the commission of a mock crime fosters threat-sensitivity in pupil responses. However, it is unknown whether the commission of a mock crime is a proxy for resource depletion. Also, Webb's Challenge Task items were not very difficult. In comparisons of depleted and nondepleted subjects (Wright et al., 2003; Wright et al., 2007), the CV reactivity of depleted subjects was higher under easy and threatening conditions but lower under difficult conditions; do pupil responses act like CV responses? How might resource depletion moderate pupil responsivity to different threat-difficulty combinations? The following section outlines a few hypotheses.

Theoretical Models and Hypotheses

In this study, EPRs to a variety of threat-difficulty combinations will be observed within a dual task resource depletion context. In terms of a path diagram, the experiment is designed to test the effects modeled in Figure 5.

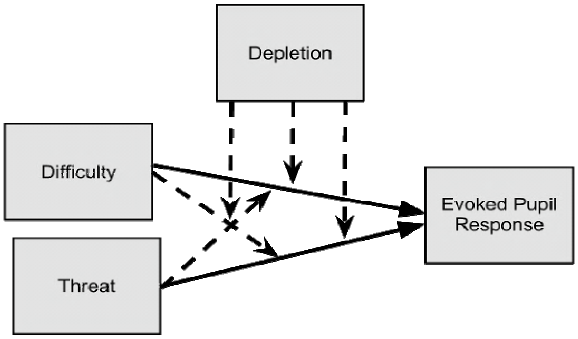

Figure 5. **Path Diagram of Effects to Be Tested in This Study.** Solid-lines indicate main effects; dashed-lines indicate moderators.

While Figure 5 is agnostic to theoretical models, Table 1 lists three hypotheses (and expected interactions) derived from the literature summarized above. Hypothesis 1 follows from Polt's (1970) demonstration that Difficulty and Threat are factors that interact linearly as codrivers of EPR magnitude; it regards what happens under "normal" test-taking situations in which subjects' working memory resources are not depleted prior to the test. Hypotheses 2 and 3 follow from the dual task resource depletion studies summarized in Table 1 (i.e., Johns et al., 2008; Stewart et al., 2009; Vohs et al., 2005; Vohs et al., 2008; Wright et al., 2003; Wright et al., 2007). Some of those studies demonstrated that subjects' task persistence decreases when their resources are depleted, which leads to Hypothesis 2. Others of those studies showed that depleted subjects' cardiovascular responses were greater than nondepleted subjects under threatening and easy but not difficult tasks; these results lead to Hypothesis 3.

Because no prior studies have compared partially cued with fully cued threat contingencies, I made no predictions as to how these conditions would compare. Each hypothesis is graphically illustrated in Figures 6-8. It is noteworthy that Hypotheses 2 and 3 are mutually exclusive.

Table 1. **List of Hypotheses and Expected Interactions.**

#1: For nondepleted subjects, Difficulty and Threat will both drive EPR magnitude and will not moderate one another.

#2: Resource Depletion will create a "ceiling" effect on EPRs. Neither Difficulty nor Threat will have any effect on depleted subjects.

#3: Resource Depletion will increase EPRs to threatening and easy -- but decrease EPRs to difficult -- items.

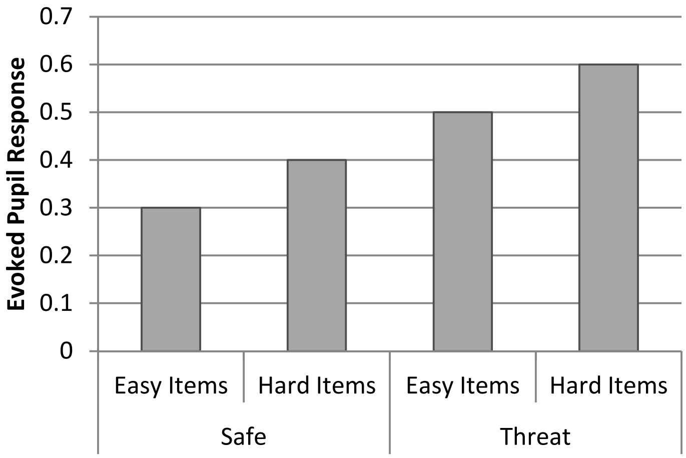

***Figure 6*. Main (But Not Interaction) Effects of Difficulty and Threat.** This graph corresponds to Hypothesis 1.

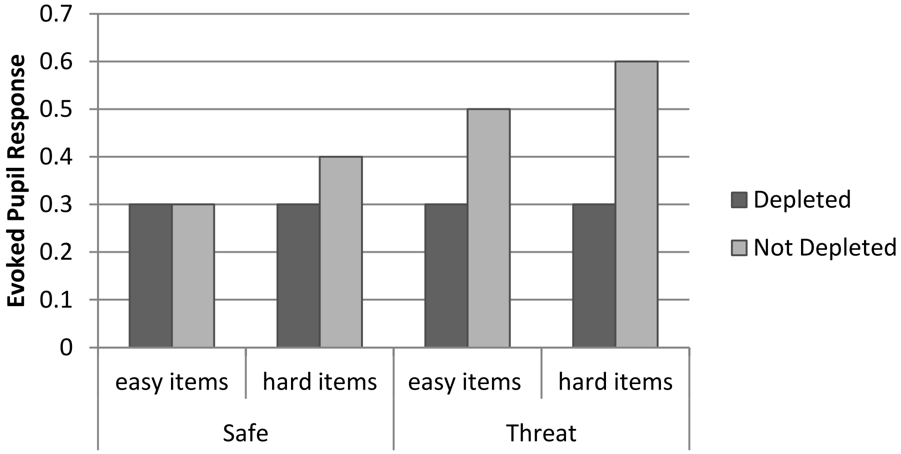

***Figure 7*. Depletion Creates an EPR Ceiling.** This graph corresponds to Hypothesis 2.

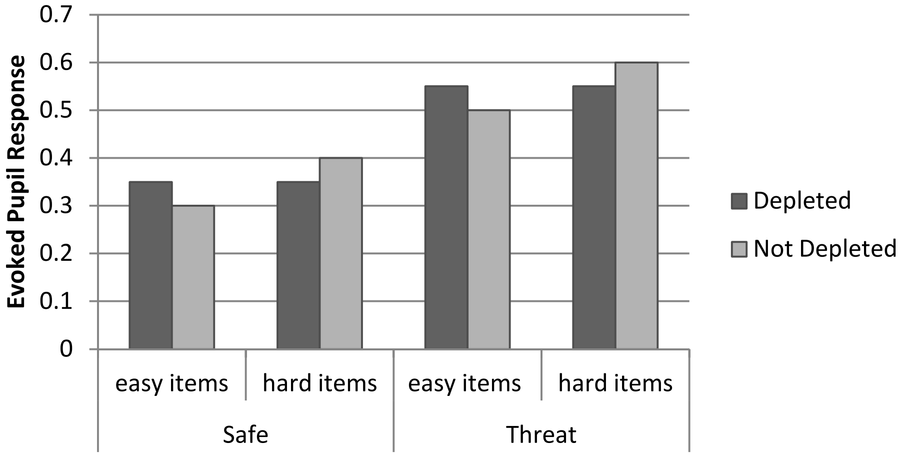

***Figure 8.* Depletion Sensitizes Threat and Desensitizes Difficulty.** This graph corresponds to Hypothesis 3.

Given the approaches taken in previous research, what is new about the current study? The methodological contributions are three-fold. First, working memory resources were manipulated as a between-subjects factor, where (prior to test-taking) one group completed a depleting task and the other group completed a nondepleting task. This approach is new within pupillometry research. Second, factor-multiple judgment test items – a previously unexplored type of Challenge Task – were created for this study. Third, in an effort to compare partially cued and fully cued threats, this study employed a novel approach to fear conditioning. The factor-multiple judgment items provide the necessary conditions for comparing partially cued and fully cued threats. Details of these experimental conditions are the topic of our next chapter.

CHAPTER 2

PROCEDURES

Participants

Eighty subjects were recruited from undergraduate courses in the College of Education at the University of Utah; they participated in return for course credit. Since the study is focused on responsivity to test-taking processes, college undergraduates are a suitable population. All subjects were (a) 18 years of age or older, (b) fluent in English, and (c) able to use a standard keyboard. Most (73%) of the subjects were female. Subjects were not allowed to repeat the experiment.

Apparatus

Test-taking pupil data were collected at 60 Hz (samples-per-second) using a Tobii T-120 eye-tracker. Stimuli were presented on the Tobii T-120 17-inch LCD monitor. Subjects' head movements were constrained using a chin rest and the monitor was positioned approximately 51 cm from the chin rest.

Unconditioned Stimulus

The unconditioned stimulus (US) used for fear conditioning was a 500-millisecond presentation (via headphones) of white noise at 90 decibels. In the fear conditioning literature, no studies have paired pupil dilation (as the conditioned response [CS]) with white noise (as the US), but a few studies have successfully found fear conditioning in pupil constrictions to electrical shock (a tactile US; see e.g, Borrego & Gardner, 1986). In a meta-analysis of fear conditioning studies that used brain activation as the CR (Sehlmeyer et al., 2009), white noise was found to be just as effective as tactile, olfactory, and visual USs. More specifically, the meta-analysis was a review of 46 prior studies (9 of which used auditory [loud white noise] as the unconditioned stimulus) and concluded the following:

> … activation of the fear network was observed to be independent of US-modality. In spite of different USs, activations of the amygdala, ACC and insula were reported for every stimulus type. (p.10)

Because the acoustic startle is known to activate the brain's fear network, it is an appropriate US for fear conditioning.

Measures

Kit of Factor-Referenced Cognitive Test [Number Facility – Division Test]

Also named the "French Kit" after one of its creators, John W. French, this battery of cognitive / intelligence tests includes a basic paper-pencil test of basic math ability. The test used in this study – called, simply, the Division Test – involves two pages of basic division problems. Each item involves a two- or three-digit number divided by a single digit factor of that number (e.g., 95 ÷ 5 = ?) – with 60 items on each page. Each subject is instructed to complete as many items as possible on the first page; after 2 minutes, the subject is instructed to turn the page and work on the second page. After 2 more minutes, the subject is instructed to stop the test. The Division Test items were normed on 119 ninth-grade males (Ekstrom, French, Harman, & Dermen, 1976, p.13) and the scale was found to have strong internal reliability (*alpha* = 0.94). This norming happened in a time before calculators were in common use and so it is not surprising that the norming sample performed much better on the test than the participants in my study (norming sample: *mean* = 34.1, *s.e.* = 1.1; my sample: *mean* =

21.1, *s.e.* = 2.6). In this study, the test provided a check for differences in math ability between experimental groups (depleted and nondepleted).

Crossing out *e*'s

In the dual task resource depletion literature, one of the most commonly used depleting tasks is "Crossing Out Letters." The nondepletion condition involves crossing out all instances of the letter 'e' in a printed text. In the task's original use (Baumeister, Bratlavsky, Muraven, & Tice, 1998), the depletion condition was described as follows:

> We told people not to cross out the letter *e* if any of several other criteria were met, such as if there was another vowel adjacent to the *e* or one letter removed. These people would presumably then scan for each *e* but would have to override the response of crossing it out whenever any of those criteria were met. Their responses thus had to be regulated according to multiple rules, unlike the others who could simply respond every time they found an *e*. (p.1260)

In the current study, depleted subjects were given two rules to follow: do not cross out *e*'s if (i) the *e* is the first letter of the word, and/or (ii) there are more than two *e*'s in a word. Nondepleted subjects were instructed to simply cross out all the *e*'s.

This task yielded moderate between-group effects across 20 prior studies (see meta-analysis: Hagger, Wood, Stiff, & Chatzisarantis, 2010). For example, in the task's original use (Baumeister, Bratlavsky, Muraven, & Tice, 1998; Experiment #4) depleted subjects had greater increases – measured before and after the Depletion Task – in self-reported tiredness than nondepleted subjects ($t[83] = 2.79$, $p < 0.01$). They also found that depleted subjects reported higher levels of concentration than nondepleted subjects ($t[63] = 2.30$, $p < .025$) during the crossing out *e*'s task.

In the current study, three details of the original task conditions were not replicated. First, I printed rather than photocopied my materials. In the original task (Baumeister, Bratlavsky, Muraven, & Tice, 1998; Experiment #4), subjects were given a single page copied from an advanced statistics book that used a highly technical writing style. Baumeister and colleagues did not report the font-size, but they did report the following:

> (For participants in the depletion condition), the photocopy of the stimulus page had been lightened, making it relatively difficult to read and thus further requiring close attention. In contrast, participants in the no-depletion condition were given an easily legible photocopy with good contrast and resolution… (p. 1260).

In contrast, I did not have access to a modern copy machine and so I printed equally legible materials for all subjects in this study. Second, I wanted to minimize the overall difficulty of the depletion task and so my materials were three pages of *Lorem Ipsum* (in Times New Roman 12-point font) instead of one page of statistics jargon. Third, I stopped subjects after 10 minutes of this activity whereas Baumeister and colleagues (apparently) did not impose a time limit on their subjects. Because it was a single page instead of three pages, the original task likely did not take as long as mine. However, this cannot be surmised because the authors did not publish a word count or a font size. It is unknown whether these deviations from the original task design represent a validity threat, but they are discussed further in Chapters 3 and 4.

Factor-Multiple Judgment Test

During the Learning and Challenge Tasks (see Procedure section below), subjects completed a computerized test that I created for the purposes of this study. In this test, items are presented one-at-a-time in black print on a single line in the center of a grey computer screen. Each test item requires the test-taker to make a judgment about the veracity of statements like “Five is a factor of twenty-two” and “Negative thirty is a multiple of six”.

Test items were designed to reflect two levels of Difficulty (*easy* = items with all positive numbers; *difficult* = items with a negative number) and three levels of Threat (*safe* = items without an explicit or invisible/implied six; *partially cued threat* = items with an invisible/implied six; *fully cued threat* = items with “six” explicitly written into the stimulus). For this experiment, one block of 25 test items was presented during the Learning Task and two blocks of 45 test items were presented during the Challenge Task. Each block had an equal number of items from both levels of Difficulty (easy vs. hard). Additionally, each level of Threat (safe, partially cued threat, and fully cued safety) was crossed with Difficulty. Other cross-balanced item variations included the following: verbal negation (whether the statement contained “not”), type of judgment (either “multiple” or “factor” in the statement), and item veracity (correct answer is True vs. False).[1] All items in this experiment were presented once; none of the Learning Task items were used during the Challenge Task and no Challenge Task items were repeated.

Figure 9 illustrates the basic item presentation sequence. Every item was followed by 1 second of performance feedback (i.e., “Correct” or “Incorrect” was presented on an otherwise blank screen), which in turn was followed by a 2-second interstimulus interval (blank grey-colored screen) before the next item onset. Other than the stipulation that items paired with the acoustic startle must be followed by safe items, item-types were presented in a random sequence.

Procedure / Experimental Tasks

When a subject arrived for the study, he/she received an Informed Consent Form from the principal investigator (PI; B. Brian Kuhlman). Then he/she was given the paper-pencil Division Test discussed above. Next, I instructed the subject to put on the headphones, place his/her fingers on the keyboard, and position his/her chin comfortably into the chin-rest. When the subject reported he/she was ready, I initialized the computerized Learning Task.

---

[1] While these three factors were controlled /cross-balanced, they were not factors of theoretical interest.

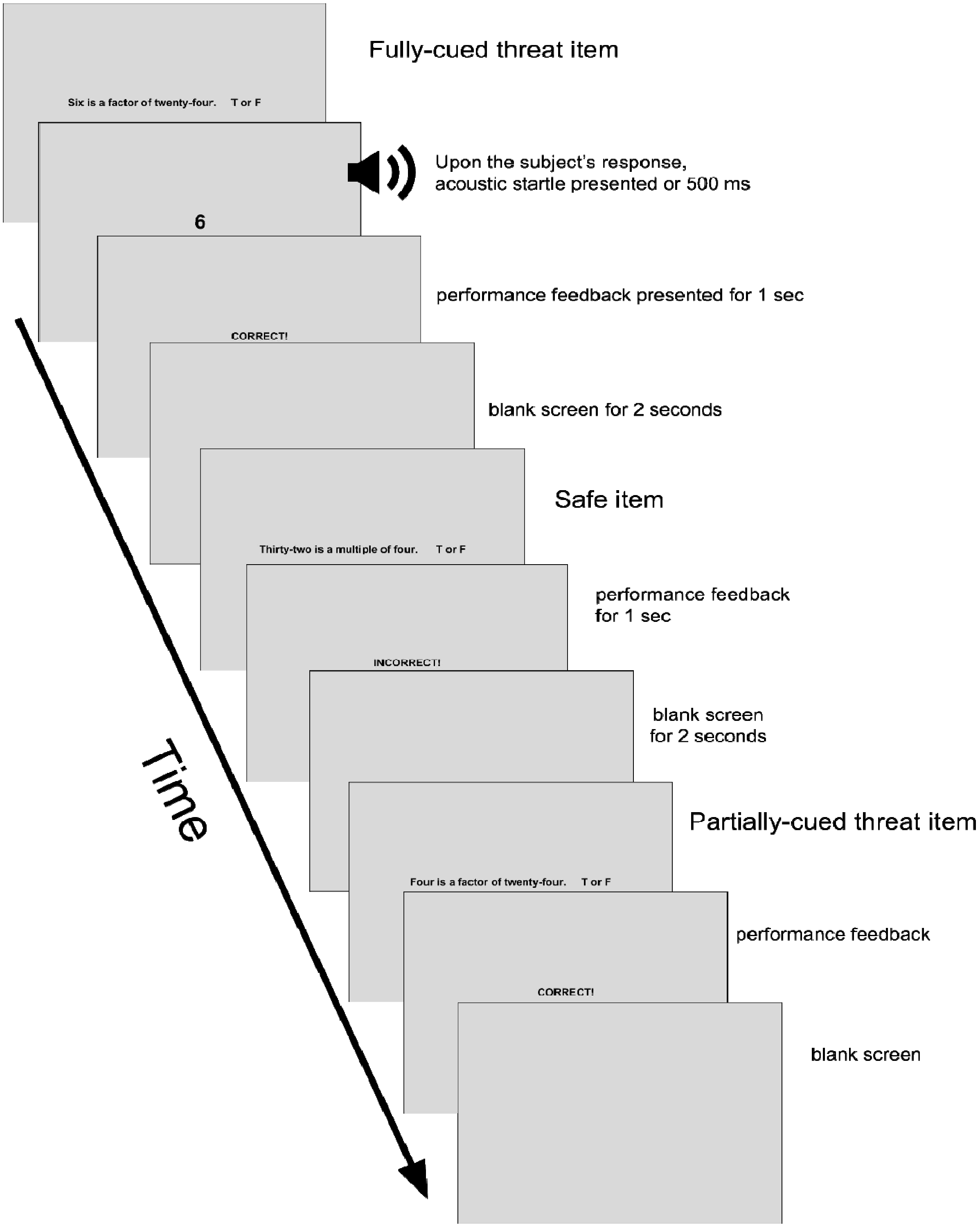

Figure 9. **Factor-Multiple Judgment Test Stimuli Sequence.**

Learning Task

The purpose of the Learning Task was (i) to acquaint subjects with the paradigm used during the Challenge Task and (ii) to train the fear conditioning response. At the onset of the Learning Task, the computerized test instructed subjects as follows.

> *The items in this test involve making True-False judgments regarding statements like ‘Seven is a factor of thirty-five. T or F’ and ‘Forty is not a multiple of ten. T or F’. Each item has an* ***invisible number****. In these examples, the invisible numbers are five (7 ** ***5*** *= 35) and four (10 ** ***4*** *= 40).*

Then subjects were instructed to speak aloud the invisible number as they completed a few practice items. Subjects then read the following instructions.

> *Your task is to complete every item as quickly and accurately as possible. You will see either ‘CORRECT’ or ‘INCORRECT’ after your response to each test item.*

Subjects next completed another small number of practice items, speaking aloud the invisible number in each item. Subjects then read about the experiment’s threat factor.

> *You will be startled by a loud noise a few times during this test. The loud noise will be paired with the number six.* ***At least 50%*** *of items with the number six will be paired with the loud noise. So,* ***if you see “six”*** *on the screen, you may be startled with the loud noise. Also,* ***if six is the invisible number*** *then you may be startled by the loud noise. Items that do not have a visible or invisible six will* ***never*** *be paired with the loud noise.*

Next, subjects read the following.

> *Your performance will be monitored during this test. If you fail to complete enough items correctly, you will be required to complete the test again.*

This last sentence was deceptive; no subjects were asked to repeat the test. This manipulation was used to increase performance motivation, simulating a high-stakes testing scenario.

All the preceding instructions were used to fulfill the first purpose of the Learning Task: familiarize subjects with the (rather complicated) conditions that would be employed during the Challenge Task. To fulfill the second purpose of the Learning Task – train the fear conditioning response (i.e., “six” = threat) – subjects completed a block of

25 items uninterrupted by instructions. During this final part of the Learning Task, subjects were not required to speak aloud the invisible number in the test items and the acoustic startle was administered four times. After completing this final block of items, subjects continued to the Depletion Task.

Regarding this study's internal validity, it is uncertain whether the Learning Task described was long enough to fulfill its purpose. Because fear conditioning is a learning process, evoking robust fear responses demands a rather high number of learning trails. During conditioning, responses to a conditioned stimulus (e.g., the number "six") grow more robust each time the conditioned stimulus is paired with the unconditioned stimulus (e.g., an acoustic startle). More trials are better, but apparently there are no validated benchmarks for the exact number of trials to use. Moreover, this study attempted to condition subjects to fear a *partially cued* stimulus (i.e., the invisible number six [e.g., "Thirty is a multiple of five. T or F"]). To my knowledge, many researchers have conditioned their subjects to fear *fully cued* stimuli (e.g., the visible number "six"), but no one has attempted to condition a partially cued stimulus. So, without references to (i) previous research on fear conditioning to partially cued stimuli or (ii) validated benchmarks for the proper number of trials to use in fear conditioning studies generally, I cannot address with certainty the internal validity of this study's design.

Depletion Task

Following the learning phase, subjects used paper and pencil to complete either the depleting or the nondepleting crossing out *e*'s task (see "Measures" section above). Group assignment was randomized. None of my subjects completed more than 2.5 of the 3 pages during the 10-minute time limit. Compliance was monitored for each subject and, in fact, all subjects obeyed the rules outlined by their respective condition.

Challenge Task

Challenge Task item types are illustrated in Table 2. The distribution of item-types during the Challenge Task is listed on Table 3. During the Challenge Task, subjects completed two blocks of 45 items with a short rest in between blocks. At the onset of each item-block, subjects were reminded (i) to respond as quickly and accurately as possible, and (ii) that they would be required to repeat the block if they performed poorly. As in the Learning Task, this last point was deceptive; no participants were required to repeat any item blocks. The threat of being required to repeat the item block was used to simulate high-stakes testing conditions. All subjects completed the same test items presented in the same order.

Reflection Task

Following the Challenge Task item blocks, subjects were asked to complete a paper-and-pencil questionnaire about their experience of the Challenge Task. The questionnaire contained the following three self-report items:

(i) What do you estimate was your percent correct during the test?

____%

(ii) How often were you surprised by the acoustic startle?

____ % of the times it sounded

(iii) How difficult were the test items?

(very easy) 1 2 3 4 5 6 7 8 9 10 (very difficult)

Finally, subjects were informed that the experiment was complete and they were free to ask questions. As they were escorted to the door, subjects were reassured that their course credit would be applied as soon as possible. On average, the entire experiment lasted 50 minutes.

Table 2. **Examples of Items from Six Threat-Difficulty Combinations.**

Threat Level	Difficulty	Example	
Safe	Easy	Fifteen is not a multiple of seven.	T or F
	Hard	Negative twenty is a multiple of three.	T or F
Partially Cued Threat	Easy	Twenty-four is a multiple of four.	T or F
	Hard	Nine is not a factor of negative fifty-four.	T or F
Fully Cued Threat	Easy	Six is not a factor of forty-seven.	T or F
	Hard	Negative thirty-two is a multiple of six.	T or F

***Table 3*. Challenge Task Distribution of Item-Types.**

	Easy	Hard
Safe	19 [7]	19 [7]
Partially Cued Threat	13 (4)	13 (3)
Fully Cued Threat	13 (3)	13 (4)

Notes: Numbers outside brackets / parentheses indicate the total number of items analyzed across the two item-blocks. Numbers in [] brackets indicate how many of these items were preceded by the acoustic startle; these items were removed prior to analysis. Numbers in () parentheses indicate how many of these items were paired with the acoustic startle; these items were analyzed separately. In total, 90 items (two blocks of 45 items each) were presented during the Challenge Task.

CHAPTER 3

RESULTS

Group Comparisons

After collecting data from 8 pilot subjects (excluded from analysis), 72 subjects participated in the experiment. Of these 72, 9 subjects were excluded due to poor data collection quality; the eye-tracking data for these subjects were either very noisy or contained a large amount of missing data. Of the 63 remaining, 11 subjects failed to achieve over 75% correct during the Challenge Task; because correct responses are the unit of analysis in this study, these 11 low performers were excluded from analysis. Thus, the analyses reported below are comprised of data from 52 subjects (25 depleted condition, 27 nondepleted). All 52 subjects completed each of the experimental phases outlined in Chapter 2.

Independent-samples *t*-tests revealed no significant differences between depleted and nondepleted groups across the variables listed in Table 5. This indicates that there were no observed between-group differences in the use of spectacles, math (arithmetic division) ability, speed during the Depletion Task, self-reported accuracy during the Challenge Task, self-reported surprise during acoustic startles, and self-reported difficulty of Challenge Task items. The lack of group differences on these variables is a desirable outcome; it supports an argument for internal validity. However, it does not argue for or against the effectiveness of the study’s Depletion Task.

Experimental Manipulation Check

Did the depletion manipulation work? As noted in Chapter 2, there were three notable differences between how I implemented the Depletion Task and how it was originally implemented. First, in contrast with the original use of the crossing out *e*'s task (Baumeister, Bratlavsky, Muraven, & Tice, 1998; Experiment #4), I printed equally legible materials for all subjects rather than photocopying the depletion group's materials in a hard-to-read manner. Second, my materials were three pages of *Lorem Ipsum* instead of one page from a statistics textbook. Third, I stopped subjects after 10 minutes on-task whereas Baumeister and colleagues did not impose a time limit on their subjects.

I did not include a previously validated manipulation check of depletion effects (e.g., the cold-pressor test). However, I did collect data on response time and accuracy during the Challenge Task (see Tables 5 and 6). In previous studies (e.g., Johns et al., 2008; Stewart et al., 2009; Vohs et al., 2005; Vohs et al., 2008), depleted subjects showed decreases in persistence (faster response times) and accuracy (more wrong answers) across a variety of task-types that included math-based challenges. If my subjects showed similar group differences, this would argue for the efficacy of my version of the Depletion Task. However, an independent samples *t*-test did not reveal any statistically significant between-group differences in response time or accuracy on any of the test item-types used in this study's Challenge Task.

The crossing out *e*'s task is intended to deplete working memory resources for one group but not the other. Hypotheses 2 and 3 of this study rely on the efficacy of this experimental manipulation. However, I found no evidence that my version of this depletion manipulation worked as it was intended.

***Table 4*. Group-Level Descriptives.**

Group	Glasses	French Kit	Crossing e's	Refl #1	Refl #2	Refl #3
Depleted	24%	21.7 (+/- 2)	2.0 (+/- 0.1)	79% (+/- 2%)	76% (+/- 6%)	4.4 (+/- 0.5)
Nondepleted	37%	20.4 (+/- 3)	2.1 (+/- 0.1)	78% (+/- 3%)	69% (+/- 6%)	4.8 (+/- 0.5)

Notes. (+/-) = Standard errors. *Glasses* = percent of subjects wearing spectacles during the experiment. *French Kit* = average number of correctly answered division problems; test was administered prior to Learning Phase. *Crossing e's* = average number of pages completed during the 10-minute depletion phase. *Refl* = questionnaire given during Reflection Phase. *Refl #1* = average self-reported percent correct on Challenge Task items. *Refl #2* = average self-reported frequency of feeling "surprised" by the acoustic startle during Challenge Task. *Refl #3* = average self-reported difficulty of Challenge Task items.

***Table 5.* Group-Level Accuracy Across Challenge Task Item-Types.**

	Safe / Easy	Safe / Difficult	PCT / Easy	PCT / Difficult	FCT / Easy	FCT / Difficult
Total Items	**19**	**19**	**13**	**13**	**13**	**13**
Depleted	17.2 (+/- 0.2)	15.3 (+/- 0.3)	11.4 (+/- 0.2)	10.1 (+/- 0.3)	11.6 (+/- 0.3)	9.2 (+/- 0.3)
Nondepleted	17.8 (+/- 0.3)	15.9 (+/- 0.3)	11.1 (+/- 0.2)	9.5 (+/- 0.2)	11.8 (+/- 0.3)	9.9 (+/- 0.4)

Notes. "PCT" = Partially Cued Threat. "FCT" = Fully Cued Threat. Unbolded numbers indicate the group's average number of correct responses for each item type. *(+/-)* = Standard errors.

***Table 6.* Group-Level Response Times Across Challenge Task Item-Types.**

	Safe / Easy	Safe / Difficult	PCT / Easy	PCT / Difficult	FCT / Easy	FCT / Difficult
Depleted	3.5 (+/- 0.2)	4.0 (+/- 0.1)	3.1 (+/- 0.1)	4.1 (+/- 0.2)	4.3 (+/- 0.2)	4.6 (+/- 0.3)
Nondepleted	3.4 (+/- 0.1)	3.9 (+/- 0.1)	3.4 (+/- 0.2)	3.9 (+/- 0.2)	4.2 (+/- 0.3)	4.7 (+/- 0.3)

Notes. "PCT" = Partially Cued Threat. "FCT" = Fully Cued Threat. Numbers indicate the average response time in seconds for each item type. *(+/-)* = Standard errors.

Pupil Signal Processing

During the Challenge Task, pupil diameter was measured at the thousandths-of-a-millimeter scale, 60 times per second. Prior to analysis, all pupil signals were smoothed using a Savitzky-Golay filter (Savitzky & Golay, 1964). Evoked pupil response (EPR) amplitudes were extracted for each subject on each item.

Pupillometry is used to measure mental effort (i.e., cognitive load). Theoretically, cognitive load is present only when subjects are actively processing the task details. It is possible that subjects gave partial or no effort (they guessed) on some items and gave full effort on other items. To minimize measurement error due to subjects guessing, incorrect responses were excluded from analysis. Using only correct responses does not completely protect against this source of measurement error, but it is a good practice.

The largest EPRs in this study were during items paired with the acoustic startle (see the far-right of Figure 10). Because the acoustic startle stimulus immediately evokes pupil dilation, the primary analysis excluded items that were either paired with or preceded by the acoustic startle. The acoustic startle was used throughout the experiment to associate / condition a cue (the number six) with fearful anticipation. This sense of threat – and not the immediate effect of the acoustic startle – was a variable of interest in this analysis.

For each item of the Challenge Task, event markers for item-onset (the moment when the item appeared on screen) and response-onset (the moment when the subject pressed a key indicating a T or F response) were embedded into the eye-tracking data stream. This type of event marking, managed by the stimulus-presentation software, allows the analyst to center item-level signal processing on either item- or response-onset. The time-window used for this analysis was centered on response-onset.

EPR amplitudes were defined as the greatest difference between a low-point (smaller pupil diameter) and a subsequent peak (bigger pupil diameter) within a 5-

second time window that ran from 3 seconds before response-onset until 2 seconds after response-onset. For both depleted and nondepleted groups, the average time from item-onset to response-onset was approximately 4 seconds (Table 6) and the period between response-onset and next item-onset was at least 3 seconds (Figure 9). So, each EPR measure represents the subjects reaction to one (and only one) test item.

After EPRs were extracted for each correct response to a nonstartle item, the responses were averaged within subject by item type (e.g., all safe-easy items, all safe-difficult items). Figure 10 represents these aggregated EPRs averaged within experimental groups.

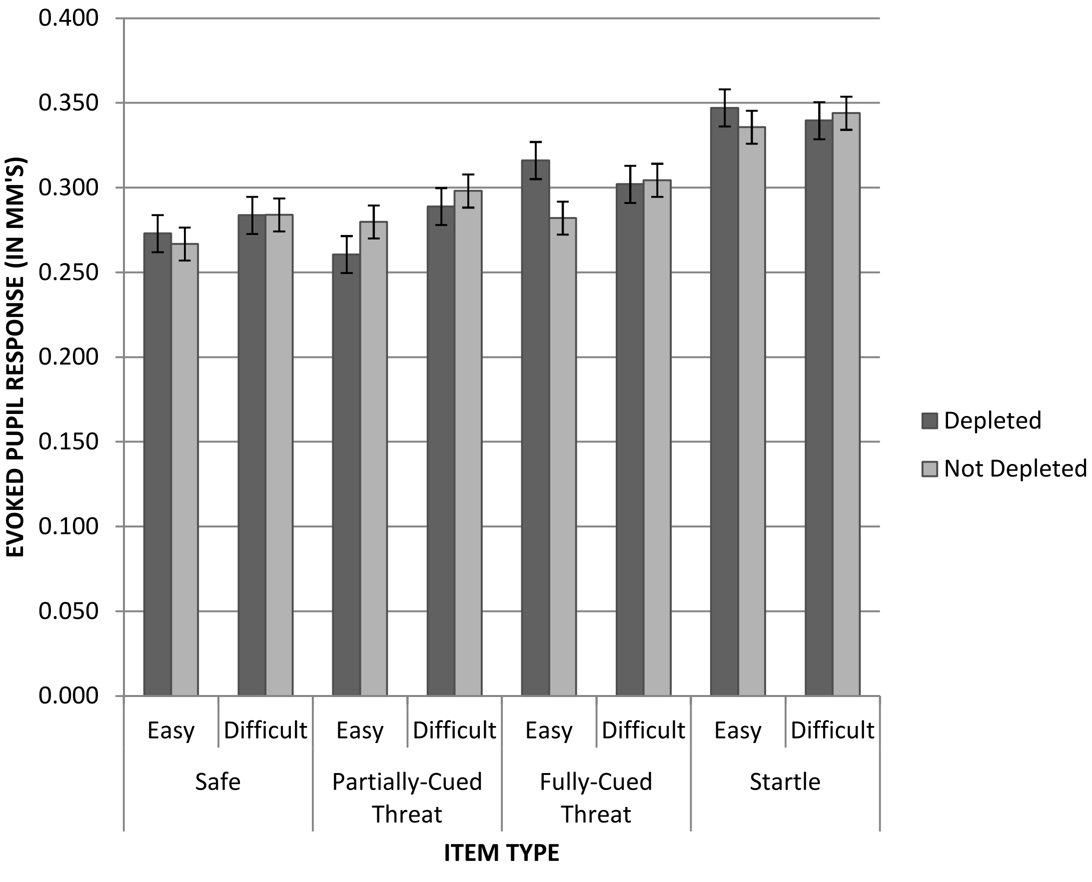

***Figure 10.* EPR Magnitudes Across All Conditions.** Error bars represent standard errors. The “Partially Cued Threat” and “Fully Cued Threat” conditions represent EPR magnitudes for items that were not paired with the acoustic startle. The “Startle” condition was excluded from the RMANOVA listed on Table 8.

Hypothesis Testing

To test the hypotheses listed in Table 1, this study employed a 2 X (2 X 3) mixed experimental design. The between-subjects (moderating) factor was *Depletion*, where 25 subjects were in the depleted and 27 in the nondepleted group. The within-subjects (main and interaction) factors included two levels of *Difficulty* (easy = items with all positive numbers, difficult = items with a negative number) and three levels of *Threat* (safe = items without an explicit or implied six, partially cued threat = items with an implied six, fully cued threat = items with "six" explicitly written into the stimulus).

Repeated-measures ANOVA was used to test for effects of between- and within-subject conditions. In these tests, the output of interest was the estimated variance explained by a factor (i.e., partial eta-squared). Effect sizes are not reported here when F-tests failed to meet a standard measure of statistical significance (i.e., when *alpha* > 0.05). For the purposes of this study, effect sizes were evaluated as follows: $eta^2 = 0.01$ is a small effect, $eta^2 = 0.05$ is a moderate effect, $eta^2 = 0.14$ is a large effect, consistent with Cohen's (1998, p. 283) suggested evaluations.

Accepting a 5% chance of Type I error (*alpha* = 0.05) and assuming a group sample size of 25, a power analysis on the within-between interaction effects of depletion yields the following chances of Type II errors. For the interaction with *Difficulty* (two measures [easy and difficult] yielding one numerator degree of freedom), the probabilities of Type II errors are 78.7% for small effects (*beta* = 0.213), 7.5% for moderate effects (*beta* = 0.925), and 0.1% for large effects (*beta* = 0.999). For the interaction with *Threat* (three measures [safe, partially cued and fully cued threat] yielding two numerator degrees of freedom), the probabilities of Type II errors are 83% for small effects (*beta* = 0.170), 24% for moderate effects (*beta* = 0.760), and 0.7% for large effects (*beta* = 0.993). In other words, this study generally has the power to detect large and moderate but not small effects of depletion.

As shown in Table 7, the main effect of Difficulty was statistically significant explaining 8% of variance in EPR magnitudes; the main effect of Threat was statistically significant, explaining 12% of variance in EPR magnitude; the Difficulty X Threat interaction was not statistically significant; and the various moderating effects of Depletion were not statistically significant. The main effects of Difficulty (*eta*^{2} = 0.08) and Threat (*eta*^{2} = 0.12) reflect moderate-to-large effect sizes. As Table 8 suggests, these effects are due to the fact that EPRs were consistently larger under difficult and fully cued threat conditions. The results of this analysis are discussed further in the next chapter.

Table 7. **Results of 2 X (2 X 3) RMANOVA.**

Source	***f***	***F***	***p***	***Observed Power***	***Partial Eta Squared***
Difficulty	1	4.42	0.04	0.54	0.08
Difficulty X Depletion	1	0.69	0.41	0.13	
Threat	2	6.63	0.00	0.85	0.12
Threat X Depletion	2	2.27	0.12	0.40	
Difficulty X Threat	2	0.76	0.45	0.17	
Difficulty X Threat X Depletion	2	1.14	0.32	0.23	

Notes: Results were generated by SPSS. In contrast to Figure 10 above, only three levels of threat were tested for statistical significance; items paired with the acoustic startle (far-right of Figure 10) were excluded from this RMANOVA.

Table 8. **Average EPR Magnitudes by Item-Type, Aggregated Across Groups.**

	Safe	PCT	FCT
Easy	0.270 (+/- 0.009)	0.270 (+/- 0.011)	0.298 (+/- 0.014)
Difficult	0.284 (+/- 0.008)	0.294 (+/- 0.008)	0.303 (+/- 0.010)

Note: (+/-) = standard errors. Numbers outside parentheses represent EPR averages, in millimeters, across all subjects in the study.

CHAPTER 4

DISCUSSION

As stated in Chapter 1, the methodological contributions of this study were three-fold. First, working memory resources were manipulated as a between-subjects factor, where (prior to the Challenge Task) one group completed a depleting task and the other group completed a nondepleting task; this approach could extend our understanding of how pupil reactions dynamically reflect working memory capacity limits (see e.g., Granholm et al., 1996; Meer et al., 2010). Second, this study pioneered an approach to fear conditioning where both implicit (partially cued threat) and explicit (fully cued threat) stimuli were conditioned; this approach could deepen our understanding of how cognitive and emotional processes interact at a basic level. Third, a previously unexplored type of mental arithmetic task (i.e., factor-multiple judgments) was employed; the presence of “invisible numbers” in these test items provides a face-valid tool for investigating implicit cognitive (math-based) processes.

The results of this experiment argue that cognitive and affective load do, as suggested by Polt (1970; see Figure 4), act as two weights on a single scale. Increases in cognitive load and affective load contribute independently to the magnitude of pupil dilation responses. In this study, the two load-sources did not moderate one another. These findings support Hypothesis 1 (see Table 1 and Figure 6). This finding may imply that cognitive load and affective load are two forms of the same thing – mental arousal – and thus have similar and additive effects on pupil diameter. Alternatively, cognitive load and affective load may be distinct forces that evoke pupil dilations via distinct and additive nervous pathways (e.g., sympathetic and parasympathetic nervous systems). This is an open question for future studies to address.

There were no statistically significant moderating effects of resource depletion found in this study. In prior studies, the crossing out *e*'s task led to significant between-group differences in task persistence, response accuracy, and physiological reactivity (Hagger et al., 2010), none of which were found in this study. The general cause of these null effects is unclear; however, it is noteworthy that – given this study's three-way experimental design – a sample size of 198 (two groups of 99 subjects) would have been required to detect small depletion effects. This study's sample size only yielded the power to detect moderate-to-large depletion effects. Another limitation of this study was its lack of a previously validated manipulation check following the Depletion Task. As a result, it is unclear whether the observed null effects reflect (i) a true lack of depletion effects due to a problem with the Depletion Task implementation used in this study, or perhaps (ii) an inability of pupil responses to reveal anything more than small depletion effects. The dual task resource depletion paradigm has never been used in prior EPR studies, so the current findings present another open question for future studies to address.

The null effects of depletion indicate no direct support for either Hypothesis 2 or Hypothesis 3 (see Table 1 and Figures 7 and 8). However, Hypothesis 3 was partially supported by the fact that under the most obviously threatening conditions – during fully cued threat items and items paired with the acoustic startle – depleted subjects had larger EPRs to easy than difficult items. This pattern was not evident among nondepleted subjects (see Figure 10). Similarly, prior dual task depletion studies (e.g., Wright, Stewart, & Barnett, 2008) found that, relative to nondepleted subjects, depleted subjects' cardiovascular reactivity was greater under threatening and easy but not difficult conditions. This pattern of findings may suggest that depletion simultaneously fosters physiological sensitivity to easily processed threats while it dampens physiological sensitivity to highly complex threats. If evident in the current study, this hypothesis would have emerged as a Difficulty X Threat X Depletion effect. It is noteworthy, again, that this study's sample size does not yield the power needed to detect significant but small three-way interactions; perhaps items that are more difficult and complex than those used in this study would have yielded a detectable (moderate or large) interaction. To test this, future factor-multiple judgment studies could employ items with two-digit multipliers (e.g., "Seventeen is a factor of one hundred forty-four. T or F") as a *very difficult* condition.

How difficult were the Factor-Multiple Judgment Test items? On average, the observed EPRs in this study were 0.28 mm, which is similar in magnitude to previously published EPRs during moderately difficult tasks (e.g., 4-digit span; see Figure 1). Averaging across all Challenge Task items, 11 subjects failed to achieve 75% accuracy; the remaining 52 subjects averaged 90% accuracy and rated the item difficulty as moderate (4.6 out of 10, where 1 is very easy and 10 is very difficult). The Challenge Task items were not normed on previous samples of college-aged participants. If future studies gave these items to a large-enough sample (e.g., $n > 200$), then Item Response Theory could provide judgments of difficulty at an item-by-item level. Specifically, a two-parameter Item Response Theory analysis would provide measurements for each item in terms of difficulty (probability of failing the item) and discrimination (ability of the item to distinguish high- from low-ability test-takers).

While no hypotheses were specified, some interesting patterns emerged regarding differences between this study's threatening conditions – partially cued (i.e., when "six" was an implied multiplicand) vs. fully cued (i.e., when "six" was an explicit multiplier). The Partially Cued Threat EPRs were similar in magnitude to the Safe EPRs;

in contrast, the Fully Cued Threat EPRs approximated the magnitude of Startle EPRs (see Figure 10). How does this finding bear on fear conditioning research? It likely suggests that implicit threat cues need to be conditioned for a longer period than was used during this study's Learning Task. In fact, I believe future researchers would be wise to *reverse* the relative length of the Learning Task (one block of 25 items) and Challenge Task (two blocks of 45 items) used in this study. In other words, the Learning Task should have a lot of items and the Challenge Task should be brief. The fact that this study's Partially Cued Threat EPRs were similar in magnitude to the Safe EPRs does not, however, suggest that all implied threat cues are insufficient to evoke observable fear responses. For example, deception tests (Webb, 2008) effectively use threat implications to distinguish between guilty and innocent test-takers. Future studies can investigate this issue further using various types of partially cued threat and by varying the time devoted to conditioning the implied threat cue.

Before concluding, it should be stated that the findings of this study are preliminary and generalization of these findings is limited to people who have roughly the same language and math abilities as the students in my sample. In particular, the null effects of Depletion cannot be taken at face value. Due to the loss of 21 of the original 72 subjects, this study was under-powered for the analysis of between-within interactions. On the other hand, this study's most definitive findings (e.g., support of Hypothesis 1) beg for a series of replication studies that draw from different (and bigger) population samples as well as the use of other (but similar) Challenge Tasks.

Future Directions

The findings of this study suggest a number of post-hoc hypotheses that are worth further investigation. Each of these hypotheses is listed on Table 9.

Table 9. **Post-Hoc (Speculative) Hypotheses**

H1.1	Fear conditioning is compatible with working memory tasks when the threat-contingency is relevant to the task. If, however, the threat-contingency is irrelevant to the task, then fear conditioning is prevented.
H1.2	When subjects have previously been conditioned to fear a stimulus, and when that stimulus is used as a source of threat in a cognitive task, the effects of difficulty and threat (during the task) on pupil responsivity are additive.
H2	Cognitive and affective load are distinct response drivers that have the same effect on the pupil (i.e., dilation). Cognitive load evokes pupil responses via parasympathetic nervous activation, whereas affective load evokes pupil responses via sympathetic nervous activation.
H3.1	Under threatening conditions, depleted task-EPRs are bigger to easy than difficult items whereas nondepleted task-EPRs are bigger to difficult than easy items.
H3.2	Depletion does not create a “ceiling” on fear-EPRs. Rather, easy conditions sensitize and difficult conditions desensitize depleted pupils to fear-conditioned stimuli.
H3.3	Under safe conditions, nondepleted task-EPRs are bigger to very difficult than moderately difficult items but depleted task-EPRs are not.

First, the study’s most compelling finding is that difficulty and threat are compatible EPR drivers, which presents a point of contrast with studies showing that engagement in a working memory task prevents fear conditioning (e.g., Straube et al., 2011). As discussed in Chapter 1, I think this contrast depends on whether the threat-contingency is relevant to the working memory task (see H1.1 on Table 9). Even if that hypothesis is wrong, it is reasonable to expect that difficulty and threat are independent (nonmoderating) drivers of pupil responsivity during test taking (see H1.2 on Table 9).

Second, allow me to briefly speculate as to why difficulty and threat are independent drivers of pupil responsivity. It does not seem likely that cognitive load and affective load are two forms of the same thing (mental arousal). Cognition and emotion are widely understood as distinct processes. Qualitatively different stimulus-types create cognitive and affective load. And, outside of pupillometry research, cognitive and affective load evoke qualitatively different responses. So, instead of reducing cognitive and affective load to merely two forms of arousal, I propose that task-EPRs and fear-EPRs could be facilitated by distinct sets of nerves (see H2 on Table 9).

Third, it remains unknown whether resource depletion exerts any effect on EPRs. However, the right half of Figure 10 reveals an interesting (yet small) group difference. When subjects are clearly threatened or directly annoyed, depletion appears to reverse the direction of difficulty effects on pupil responses (see H3.1 on Table 9). Notice also that the biggest EPRs shown on the right half of Figure 10 represent depleted pupil responses. This argues against the idea that depletion flatly dampens (as in Figure 2) the effects of threat on pupil responses (see H3.2 on Table 9). In contrast to its right half, the far-left side of Figure 10 suggests null depletion effects. Under clearly safe conditions, the groups look the same. However, this does not rule out the possibility that depletion lowers the “ceiling” on task-EPRs. Granholm and colleagues (1996) showed that nondepleted participants hit a task-EPR ceiling at around the ninth digit of a 13-digit span task. Because depletion is known to reduce persistence on difficult tasks, it seems likely that depletion lowers the task-EPR ceiling (e.g., from a maximum of 9-digit spanning to a maximum of 7- or perhaps even 5-digit spanning). Looking at Figure 1, the task-EPRs found in this study correspond to those evoked by a 5-digit span task. Even if my version of the crossing out *e*’s task was an effective Depletion Task, my Challenge Task items were not difficult enough to reveal the lowered task-EPR “ceiling”. Using test items that are very difficult, this depletion effect could be revealed (see H3.3 on Table 9).

Lastly, I believe that basic research on the interaction of working memory and emotion could be greatly enhanced by the use of numbers that (i) have been conditioned to / paired with either a threatening or a pleasing stimulus, and (ii) can be reliably evoked as an implicit part of a mental math task. Many working memory tasks involve the manipulation of numbers. While the effects of partially cued threats in this study were marginal at best, I think that a longer Learning Task would have effectively conditioned my “invisible numbers”. It is premature to make clear predictions about how partially cued threats – when properly conditioned – would interact with depletion and difficulty factors. However, I believe these tools could help basic researchers develop the theoretical framework necessary to understand a number of psychological applications. For example, knowledge about the effects of implicit threat cues could elucidate how and why deception detection tests (e.g., Webb, 2008) work the way they do.

REFERENCES

Baumeister, R.F., Heatherton, T.F., & Tice, D.M. (1994). *Losing control: How and why people fail at self-regulation.* San Diego, CA: Academic Press.

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In J. Caccioppo, L.G. Tassinary, & G. Berntson (Eds.), *The Handbook of Psychophysiology* (pp. 142-162). Hillsdale, NJ: Cambridge University Press.

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin, 91*, 276-292.

Borrego, R.L., & Gardner, R.M. (1986). Classical conditioning of pupil constrictions. *Perceptual and Motor Skills, 62*, 315-322.

Hess, E. (1975). *The tell-tale eye.* New York: Van Nostrand.

Carter, R.M., Hofstotter, C., Tsuchiya, N., & Koch, C. (2003). Working memory and fear conditioning. *Proceedings of the National Academy of Sciences, 100*(3), 1399-1404.

Chapman, C.R., Oka, S., Bradshaw, D.H., Jacobson, R.C., & Donaldson, G.W. (1999). Phasic pupil dilation response to noxious stimulation in normal volunteers: Relationship to brain evoked potentials and pain report. *Psychophysiology, 36,* 44-52.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Conway, A.R., Kane, M.J., & Engle, R.W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Science, 7*(12), 547-552.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Service.

Goldinger, S., He, Y., & Papesh, M. (2009). Deficits in cross-race face learning: Insights from eye movements and pupillometry. *Journal of Experimental Psychology*: *Learning, Memory, and Cognition, 35*(5), 1105–1122.

Granholm, E., Asarnow, R., Sarkin, A., & Dykes, K., (1996). Pupillary responses index cognitive resource limitations. *Psychophysiology, 33,* 457-461.

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

Hagger, M., Wood, C., Stiff, C., & Chatzisarantis, N. (2010). Ego-depletion and the strength model of self-control: A meta-analysis. *Psychological Bulletin, 136*, 496-525.

Hess, E., & Polt, J. (1960). Pupil size as related to interest value of visual stimuli. *Science*, *132,* 349-350.

Hess, E., & Polt, J. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science*, *143,* 1190-1192.

Heinrich, W. (1896). Die Aufmerksamkeit und die Funktion der Sinnesorgane. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane, 9*, 342-388.

Hourdaki, E., Giakoumaki, S., Armenaki, A., Theou, K., Grinakis, V., & Bitsios, P. (2003). Fear-inhibited light reflex: Effects of conditioned stimulus intensity and fear stimulus modality. Poster presented at *International Society on Brain and Behaviour: 1st International Congress on Brain and Behaviour***.** Retrieved 2/17/13 from http://www.annals-general-psychiatry.com/content/2/S1/S103

Johns, M., Inzlicht, M., & Schmader, T. (2008). Stereotype threat and executive resource depletion: Examining the influence of emotion regulation. *Journal of Experimental Psychology: General, 137*, 691–705.

Just, M., & Carpenter, P. (1993). The intensity dimension of thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology,* 47(2), 310-339.

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science, 154*, 1583-1585.

Kane, M.J., Hambrick, D.Z., & Conway, A.R. (2005). Working memory capacity and fluid intelligence are strongly related constructs. *Psychological Bulletin, 131*(1), 66-71.

Klingner, J. (2010). *Measuring cognitive load during visual tasks by combining pupillometry and eye tracking*. (Ph.D. dissertation). Stanford University Computer Science Department. Retrieved 2/17/13 from http://graphics.stanford.edu/papers/klingner-dissertation/klingner-dissertation.pdf

Meer, E., Beyer, R., Horn, J., Foth, M., Bornemann, B., Ries, J., Kramer, J., Warmuth, E., Heekeren, H., & Wartenburger, I., (2010). Resource allocation and fluid intelligence: Insights from pupillometry. *Psychophysiology, 47*(1), 158-169.

Nunnally, J., Knott, P., Duchnowski, A., & Parker, R. (1967). Pupillary response as a general measure of activation*. Perception & Psychophysics, 2,* 149-150.

Polt, M. (1970). Effect of threat of shock on pupillary response in a problem-solving situation. *Perception* & *Motor Skills, 31*, 587-593.

Savitsky, D., & Golay, M. J. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry, 36(8),* 1627-1639.

Schiff, J. (1875). *La pupille consideré comme esthésiométre* (translated by R. G. de Choisity). Paris: J. B. Bailliére.

Schiff, J., & Foa, P. (1874). La pupille considerée comme esthésiométre (translated by R. G. de Choisity). *Marseille Médical, 2*, 736-741.

Sehlmeyer, C., Schoning, S., Zwitserlood, P., Pfleiderer, B., Kircher, T., Arolt, V., & Konrad, C. (2009). Human fear conditioning and extinction in neuroimaging: A systematic review. *PLoS ONE, 4*(6), 1-16.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General, 125,* 4–27.

Stanners, R., Coulter, M., Sweet, A., & Murphy, P. (1979). The pupillary response as an indicator of arousal and cognition*. Motivation and Emotion, 3,* 319–340.

Stewart, C., Wright, R., Hui, S., & Simmons, A. (2009). Outcome expectancy as a moderator of mental fatigue influence on cardiovascular response. *Psychophysiology, 46*, 1141–1149.

Straube, T., Trippe, R., Schmidt, S., Weiss, T., Hecht, H., & Miltner, W.H.R. (2011). Dissociation of acquisition and expression of fear conditioned responses under working memory load. *Emotion, 11*(1). 209-213.

Vohs, K., Baumeister, R., & Ciarocco, N. (2005). Self-regulation and self-presentation: Regulatory resource depletion impairs impression management and effortful self-presentation depletes regulatory resources. *Journal of Personality and Social Psychology, 88*, 632–657.

Vohs, K., Baumeister, R., Schmeichel, B., Twenge, J., Nelson, N., & Tice, D. (2008). Making choices impairs subsequent self-control: A limited-resource account of decision making, self-regulation, and active initiative. *Journal of Personality and Social Psychology, 94*, 883–898.

Webb, A. (2008). *Effects of motivation and item difficulty on oculomotor and behavioral measures of deception*. (Ph.D. dissertation). University of Utah, Department of Educational Psychology.

Wright, R., Junious, T., Neal, C., Avello, A., Graham, C., Herrmann, L., Junious, S., & Walton, N. (2007). Mental fatigue influence on effort-related cardiovascular response: Difficulty effects and extension across cognitive performance domains. *Motivation and Emotion, 31*, 219–231.

Wright, R., Martin, R., & Bland, J. (2003). Energy resource depletion, task difficulty, and cardiovascular response to a mental arithmetic challenge. *Psychophysiology, 40*, 98–105.

Wright, R. A., Stewart, C. C., & Barnett, B. R. (2008). Mental fatigue influence on effort-related cardiovascular response: Extension across the regulatory (inhibitory)/nonregulatory performance dimension. *International Journal of Psychophysiology, 69*, 127–133.