THE EFFECTS OF CAFFEINE ON COGNITIVE FATIGUE

A Thesis
Presented to
The Academic Faculty

By

Sunni Haag Newton

In Partial Fulfillment
Of the Requirements for the Degree
Master of Science in Psychology

Georgia Institute of Technology

December, 2009

The Effects of Caffeine on Cognitive Fatigue

Approved by:

Dr. Phillip L. Ackerman, Advisor
School of Psychology
Georgia Institute of Technology

Dr. Paul Corballis
School of Psychology
Georgia Institute of Technology

Dr. Ruth Kanfer
School of Psychology
Georgia Institute of Technology

Date Approved: October 12, 2009

ACKNOWLEDGEMENTS

I would like to thank the following people:

My parents, Bob and Peggy Haag, whose constant love and support is the root of all my past, present, and future success and happiness

My husband, Tommy Newton, whose unerring ability to both cheer me up and calm me down has proved invaluable throughout the process of working on this thesis

My sisters, Kaysea and Laney Haag, who were quick to provide an amusing story and/or a sympathetic ear when I needed it

My thesis advisor, Dr. Ackerman, whose helpful suggestions greatly increased the quality of this thesis

The members of my thesis committee, Dr. Ackerman, Dr. Corballis, and Dr. Kanfer for the generous contributions of both their time and their knowledge

My colleagues, especially Julie Nguyen, Katie McNulty, Charles Calderwood, and Stacey Shapiro, whose intelligence, kindness, and humor have improved my grad school experience immeasurably

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Prior caffeine research has examined the effects of caffeine on performance using simple, lower-level cognitive tasks. The present study extended this work to investigate the effects of caffeine on performance and self-report mood measures during execution of a complex cognitive task. In a between-subjects design, 116 participants were administered either caffeinated or non-caffeinated chewing gum. Results showed higher fatigue and negative affect (NA) levels and lower positive affect (PA) and task performance levels in the placebo condition. These findings replicate prior findings on mood effects of caffeine; also, they extend the limited results on performance effects of caffeine by demonstrating moderate support for improved complex cognitive task performance after caffeine intake. Furthermore, these results show the efficacy of gum for caffeine administration in research.

CHAPTER 1

INTRODUCTION

The widespread use of caffeine in the general population is exemplified by more than just long customer lines at the local Starbucks; researchers have estimated that as much as 87% of the US population aged two and older consumes caffeine in some form on a daily basis (Frary, Johnson, & Wang, 2005). Indeed, caffeine is the most widely consumed central nervous system stimulant in the world (Committee on Military Nutrition Research, Food and Nutrition Board [CMNR], 2001), and records suggest that humans have ingested the products of plants containing caffeine since prehistoric times (James, 1991).

Although caffeine has been, and remains, very popular, the acceptability of its use in the general population has been controversial for some time. As James (1991) noted: "[T]he introduction of caffeine beverages into countries that had not previously known them sometimes elicited deep moral outrage and efforts to repress the use of these new substances" (p. 15). In 1675, King Charles II denounced English coffee houses as "seditious meeting places" (cf James, 1991). In 1911, the US government sued the Coca-Cola Company for "marketing and selling an adulterated beverage that was injurious to health because it contained a deleterious ingredient, namely, caffeine" (Benjamin, Rogers, & Rosenbaum, 1991, p. 593). (Interestingly, this lawsuit was the reason behind the decision of the Coca-Cola Company to hire Harry Hollingworth to conduct a scientific investigation of the behavioral effects of caffeine.) In general, although present opinions on the topic are less extreme than they have been in the past, controversy

continues regarding the use of caffeine, particularly among vulnerable populations (e.g., young children, pregnant women).

Perhaps the most important reason for the consumption of caffeine by so many individuals relates to the long-standing belief that caffeine enhances mental effectiveness. As Kraepelin (cited in Hollingworth, 1912) noted nearly 100 years ago:

> We know that tea and coffee increase our mental efficiency in a definite way, and we use these as a means of overcoming mental fatigue…In the morning these drinks remove the last traces of sleepiness and in the evening when we still have intellectual tasks to dispose of they aid in keeping us awake. (p. 3)

In the remainder of the introduction, I present a review of both the relevant literature pertaining to the effects of caffeine on mood states and performance and the literature on cognitive fatigue. First, I introduce some general research findings on caffeine and cognitive fatigue, provide background information about the reasons for widespread interest in research on caffeine and cognitive fatigue, and address how these two broad topic areas are combined in the current study. Second, I provide basic information on caffeine and psychopharmacologic actions and long-term health consequences associated with caffeine use. Third, I summarize the literature on the effects of caffeine on mood and performance, the effects of caffeine dosage, and the effects associated with various task types. Fourth, I introduce the notion of cognitive fatigue as a consequence of prolonged action, discuss various definitions and issues of measurement, and summarize empirical evidence relating caffeine and cognitive fatigue to performance. In the fifth and final section of the introduction, I address several salient methodological issues in this field of study, including the determination of caffeine dose,

the selection of a vehicle for caffeine delivery, and the choice of an operational definition for a typical caffeine consumer.

The immense popularity of caffeine has prompted decades of investigations on its effects, going back as early as 1912 when Hollingworth conducted a study on the effects of caffeine on human performance and sleep (Hollingworth, 1912). Research over the decades has yielded two well-established findings with respect to the effects of caffeine on human behavior. First, caffeine is generally associated with increased levels of alertness. Second, caffeine tends to temporarily improve performance on sustained attention and psychomotor tasks (Christopher et al., 2005). These effects are typically found among alert individuals soon after caffeine is consumed (Smith et al., 2005). Recent research has also suggested a different set of effects for individuals who are already fatigued at the time that they consume caffeine. For example, Smith et al. (2005) found significant improvements in alertness, vigilance performance, and speed of encoding new information for individuals tested at the beginning of a four-hour session and faster simple reaction time (RT) and fewer long responses on choice RT tasks for the same individuals tested at the end of the four-hour session. In this experiment, caffeine was administered one hour into the session or one hour into the session and again three hours into the session, depending on the experimental condition. Effects found at the end of the testing session were not observed earlier in the session, and were posited to reflect the ability of caffeine to counteract behaviors associated with increasing fatigue. These differences in the time courses of various caffeine effects are thought to be influenced by variations in brain neurotransmitter activity related to whether the individual is alert or fatigued at the time of assessment (Smith et al., 2005).

Similarly, Smit and Rogers (2000) found that an increase in boredom reported by individuals in the placebo condition toward the end of the testing session was prevented by all caffeine doses (doses of 12.5, 25, 50, or 100 mg) for those individuals in the experimental conditions. Childs & de Wit (2008) found that among overnight sleep-deprived participants, 200 mg caffeine delivered via capsule led to improved mood and prevention of a reaction-time increase (on simple and two choice RT tasks) that occurred in the placebo group (see Table 1 for a list of caffeinated beverages and their estimated caffeine contents).

Interest in cognitive fatigue has also become an increasingly important aspect of educational research. In the US, high school students are often engaged during every waking hour: attending school during the day and completing homework and participating in extracurricular activities until bedtime – a pattern that education researchers refer to as "overscheduling" (Melman, Little, and Akin-Little, 2007). Also of interest are the effects of cognitive fatigue in longer tests of critical importance, such as the SAT (Christopher, 2006). Some research findings related specifically to standardized testing have suggested that feelings of fatigue may be reduced by allowing breaks of five minutes per half hour or ten minutes per hour, while other research has indicated that varying task type during the testing session may reduce the need for this frequency of breaks (Liu, Allspach, Feigenbaum, Oh, and Burton, 2004).

Educational interventions to reduce fatigue have also been applied to the school day in general. Research suggests that adolescents tend to be sleep-deprived when high school classes start prior to 8:15 am (Wahlstrom, 2003). In one school district, moving the start time for high schools from 7:15 am to 8:40 am was associated with improved

Table 1.

Caffeine content estimates for various beverages

Beverage	Size (oz)	Caffeine content (mg)	Caffeine (mg) per oz. (or per shot)
Starbucks Espresso, small (brewed @ location	**1 shot**	**58.1**	**58.1**
Einstein Bros® Espresso, double (brewed @ location)	**2 shots**	**185.0**	**92.5**
Starbucks regular coffee (brewed @ location)	**16.0**	**259.3**	**16.2**
Einstein Bros regular coffee (brewed @ location)	**16.0**	**206.3**	**12.9**
Dunkin' Donuts regular coffee (brewed @ location)	**16.0**	**143.4**	**9.0**
Red Bull® (canned energy drink)	**8.3**	**66.7**	**8.0**
Starbucks Doubleshot™ (canned energy drink)	**6.5**	**105.7**	**16.3**
Starbucks Frappuccino® Mocha (canned beverage)	**9.5**	**71.8**	**7.6**
Mountain Dew® (canned soda)	**12.0**	**45.4**	**3.8**
Pibb® Xtra (canned soda)	**12.0**	**34.6**	**2.9**
Pepsi® (canned soda)	**12.0**	**31.7**	**2.6**
Diet Pepsi® (canned soda)	**12.0**	**27.4**	**2.3**
Coca-Cola® Classic (canned soda)	**12.0**	**29.5**	**2.5**
Diet Coke® (canned soda)	**12.0**	**38.2**	**3.2**

Sources: McCusker, Goldberger, & Cone, 2003; McCusker, Goldberger, & Cone, 2006.

attendance rates and more hours of sleep per night for students, on average (Wahlstrom, 2003). Other research has investigated which times of day are best for students in terms of eliciting maximal attention and achievement (Klein, 2004). These findings suggest a gradual increase in academic achievement from 8:00 am to noon, a significant performance decline from noon to 1:00 pm, and an increase back to pre-noon levels between 1:00 pm and 2:00 pm. These various programs of research indicate that researchers in the educational field are both aware of and interested in reducing the effects of fatigue for students.

Students often anecdotally report using caffeine or other strategies to stay alert while completing homework or during testing. A recent popular media story quoted an Atlanta area college student as saying that "99.9 percent of the people I know [consume caffeine-loaded energy drinks] for most tests and all big exams" (Hendrick, 2006). Use of caffeine products has grown so rampant that certain varieties have been banned in some schools (Associated Press, 2007). A recent article on energy drinks pointed out the lack of US regulations on caffeine content limitation and labeling for these products, increasing incidences of caffeine overdose associated with rapid consumption of energy drinks, and removal of several types of energy drinks from stores after multiple adverse reactions were reported (Reissig, Strain & Griffiths, 2009).

Somewhat surprisingly, there have been no studies investigating the use of caffeine to alleviate fatigue during complex task performance, such as in the educational testing context. One recent study broached this topic area by investigating the effects of caffeine on perceived mood state, concentration, and arousal during a college lecture (Peeling & Dawson, 2007). In this study, the authors used a within-subject design to

assess students' reactions to a 75-minute lecture on exercise rehabilitation both with and without previous consumption of a low caffeine dose (100 mg. delivered in a pill capsule). The significant findings were that students in the caffeine condition reported higher levels on the following adjectives during the lecture as compared to placebo: "awake", "clear-minded", "energetic", "alert", and "anxious" (Peeling & Dawson, 2007, p. 334). Students in the caffeine condition also reported feeling significantly more able to concentrate during the lecture as compared to placebo. Unfortunately, this study included a very small sample size (N = 10) and failed to utilize a performance measure to assess information retained during the two lectures. A more compelling finding would have been a significantly higher level of information retained during the lecture after caffeine consumption as compared to the lecture after placebo. Nevertheless, this study does indicate that caffeine might influence factors relevant to attention and learning, and provides an investigation of these caffeine effects in a real-world setting (Peeling & Dawson, 2007).

Given the historical basis for using caffeine to enhance mental states and the finding that caffeine exerts additional beneficial effects for fatigued as compared to alert individuals (Smith et al., 2005), it appears theoretically and practically useful to investigate the effects of caffeine in the context of ability testing. From a theoretical perspective, an investigation of caffeine effects in this context permits assessment of the generalizability of effects on performance obtained in simple tasks (e.g., RT or vigilance tasks). From a practical perspective, examination of the effects of caffeine on complex task performance under conditions of induced cognitive fatigue may provide crucial information relevant to the widely-held but untested assumption that caffeine reduces

fatigue (and thereby improves performance) in educational and organizational settings. Thus, the purpose of the current study was to assess the effects of caffeine consumption on complex task performance among individuals under the condition of induced cognitive fatigue.

1.1 Caffeine: Pharmacologic actions and long-term health consequences

Caffeine occurs naturally in green coffee beans, maté, and cola nuts, and is most often consumed by humans in beverages such as coffee, tea, and cola. At room temperature, caffeine is a white, odorless powder with a bitter taste. Pharmacologically, caffeine is thought to affect the body by blocking the adenosine receptors in the brain (James, 1991). The function of adenosine is to regulate various biological processes throughout the body, including the cardiovascular and nervous systems. Adenosine is associated with decreased activity in all of the following functions: blood pressure, central nervous system activity, respiration, urine output, and intestinal peristalsis. By blocking adenosine receptors, caffeine prevents adenosine from having its depressing effect on all of the processes previously listed.

About 99% of caffeine consumed in beverage form (i.e, coffee, tea, or cola) is absorbed within 45 minutes of ingestion; the speed of absorption can be increased by exposing caffeine to the oral mucosa (i.e., via the chewing gum formulation that was used in the present study). The mean half-life (the time it takes for half of a given amount of the substance to be eliminated from the body) of caffeine in the plasma of healthy individuals is approximately 5 hours, with a range of 1.5 to 9.5 hours (CMNR, 2001). Smoking reduces the half-life (meaning that caffeine will be eliminated from the body more rapidly) while use of oral contraceptives increases half-life (meaning that caffeine

will be eliminated from the body more slowly). Most caffeine is metabolized in the liver, and it easily crosses cell membranes, including the blood-brain barrier. Caffeine also influences fluid homeostasis by increasing urine output within one hour of ingestion. Significantly increased urine output has been found in individuals between three and 24 hours after ingestion (CMNR, 2001). Data are inconclusive as to whether this creates a total body water deficit, but it may be a concern when an individual's water balance could already be compromised (e.g., as a result of military operations which take place in hot or high-altitude environments).

The long-term health consequences of caffeine consumption remain controversial. A substantial body of research on the relationship between caffeine use and specific diseases, such as benign breast disease and particular cancers, shows no evidence of an association (CMNR, 2001). Extensive research has also been conducted on the relationship between caffeine consumption and cardiovascular disease and osteoporosis (CMNR, 2001). With regard to cardiovascular disease, there has been no clear epidemiological evidence to suggest a causal relationship between caffeine consumption and hypertension, coronary heart disease, or occurrence of arrythmias. Similarly, there have been no clear trends linking caffeine consumption to negative outcomes with regard to bone density or occurrences of fractures (CMNR, 2001). In summary, there has been no conclusive evidence to date showing that regular use of caffeine is associated with negative medical outcomes in arenas such as cancer, bone health, and cardiovascular disease. Furthermore, several potential health benefits of regular caffeine consumption were summarized in a recent review paper: regular caffeine intake is associated with lowered risk of type 2 diabetes and Parkinson's disease, and is thought to preserve long-

term cognitive functioning by reducing ischaemic brain damage (damage caused by lowered glucose and oxygen supplies to the brain) through its action in blocking adenosine receptors (Rogers, 2007).

Caffeine has been postulated to be associated with numerous negative reproductive outcomes, such as spontaneous abortions, premature births, and low infant birth weights (CMRN, 2001). Researchers who have conducted reviews of the caffeine and pregnancy literature concluded that negative outcomes are highly unlikely to be associated with caffeine intake of less than 300 mg/day (Christian & Brent, 2001; Nawrot, Jordan, Eastwood, Rotstein, Hughenholts, & Feeley, 2003). At doses higher than 300 mg/day, the potential negative effects of caffeine on reproductive outcomes are less clear. Because researchers have not completely ruled out the hypothesis that there are deleterious consequences of caffeine on the fetus, doctors often recommend that pregnant women limit their caffeine intake.

1.2 Performance and mood effects of caffeine

Recent reviews of the effects of caffeine on mood and behavior, conducted by Smith (2002), Lorist and Tops (2003), Ruxton (2008), and Fredholm, Battig, Holmen, Nehlig, and Zvartau (1999), indicate that moderate levels of caffeine consumption (i.e., less than 300 mg/day) have positive effects on mood and certain types of performance, and that most caffeine consumers self-regulate intake to maximize positive effects and minimize negative effects (e.g., anxiety, jitteriness). Smith (2002) reviewed studies conducted between 1967 and 2002 that investigated the effects of moderate caffeine use on adult mood, performance, and sleep. For this review, Smith (2002) divided studies on caffeine and performance efficiency into two sections: studies before 1990 and studies

after 1990. Because studies prior to 1990 had already been reviewed in the literature (see Lieberman, 1992), Smith (2002) focused primarily on research conducted after 1990. Although Smith (2002) noted that the studies reviewed varied substantially in the methodologies used, he drew six conclusions about the effects of caffeine based on his review of the research literature: 1) caffeine exerts a positive effect on alertness and fatigue reduction; 2) caffeine leads to improved performance on tasks that require sustained response and on vigilance tasks; 3) the effects of caffeine on complex tasks are more difficult to assess, partially because these effects are likely a result of the interaction of caffeine and other variables, such as personality and time of day; 4) caffeine withdrawal has little or no effect on performance, but may have a negative impact on mood; 5) the regular use of caffeine appears to be beneficial, as it leads to higher mental functioning; and 6) most individuals are capable of regulating their caffeine intake so as to maximize its positive effects. Smith (2002) concluded that at levels consumed by most people (i.e., less than 300 mg/day), caffeine has primarily positive effects on human behavior.

A more recent review conducted by Ruxton (2008) echoed some of Smith's (2002) conclusions. This review (Ruxton, 2008) included 41 studies conducted between 1992 and 2007 that employed a double-blind, placebo-controlled design and used healthy adults as subjects. This set of studies was broken down into subsets and various mood, performance, and health outcomes were assessed. Sixteen studies used non-sleep-deprived subjects and investigated mood and cognitive performance outcomes; fourteen of these reported benefits of caffeine consumption, including increased alertness, increased short-term recall, decreased reaction times, improved mood, and decreased

fatigue. Caffeine doses in these studies varied, but most used a single dose ranging from 37.5 to 400 mg. The effects of caffeine in sleep-deprived subjects were less compelling; in this review, only three of seven studies on sleep-deprived subjects demonstrated beneficial effects of caffeine consumption. Obviously the extent of the sleep deprivation would be an important factor in interpreting this finding. With regard to physical performance, the majority of relevant studies indicated that caffeine doses ranging from 175 to 420 mg positively influenced physical performance measures such as endurance, power, and fatigue tolerance (Ruxton, 2008). Ruxton also addressed concerns that caffeine consumption can lead to dehydration; the relevant studies reviewed suggested that low to moderate caffeine intake was not associated with dehydration risk, even in conditions of extreme exercise. Taken together, Ruxton puts forth a range of caffeine consumption levels thought to maximize the benefits and minimize the risks associated with caffeine consumption; this suggested range is 38 to 400 mg/day.

Lorist and Tops (2003) conducted a review of empirical research investigating the relationships among caffeine consumption, fatigue, and cognition. These authors surveyed research investigating the influence of caffeine on various neuromodulator systems. Specifically, behavioral, electroencephalogram (EEG), and event-related brain potential (ERP) measures of performance were examined. Lorist and Tops (2003) found evidence to support a U-shaped dose-response curve for caffeine, such that doses lower than 500 mg were associated with positive performance effects while doses higher than 500 mg were associated with negative performance effects as well as increases in self-reports of anxiety and tension. Lorist and Tops (2003) also suggested that mental

processes are negatively accelerating as a function of increasing time-on-task, and that caffeine may serve to attenuate this slowing by blocking adenosine receptors.

The effects of caffeine on mood states show a similar pattern of findings as performance, with low or moderate doses associated with positive mood effects and high doses associated with negative mood effects. Fredholm et al. (1999) reviewed the research literature on caffeine dependence. Dependence can be defined as "a state of affairs when administration of the drug is sought compulsively, leading to disrupted behavior if necessary to secure its supply. Use continues despite the adverse psychological or physical effects of the drug" (Rang et al., 1995, cited in Fredholm et al., 1999, p. 104). The main focus of this review was on caffeine dependence, with special attention given to the effects of caffeine on specific neuronal brain substrates. Fredholm et al. (1999) concluded that caffeine doses of 20 – 200 mg were linked to positive subjective reports of mood. At doses of this magnitude, participants generally reported feeling energetic, self-confident, alert, able to concentrate, and motivated to work. These findings stand in the absence of withdrawal effects. At higher doses individuals reported feelings of anxiety, although what constitutes a higher dose for a given individual is difficult to ascertain due to interindividual differences. Fredholm et al. (1999) also suggested that most individuals adjust their caffeine intake so as not to reach the level of caffeine that would lead to feelings of anxiety. Dosage in these studies was measured in terms of mg caffeine consumed per day.

Further evidence on caffeine effects as a function of habitual usage was obtained by Haskell, Kennedy, Wesnes, and Scholey (2005) in a study of caffeine effects on cognition and mood among habitual consumers and non-consumers. A consumer was

defined as someone whose mean daily caffeine intake exceeded 50 mg/day while a non-consumer was defined as someone whose mean daily caffeine intake was less than 50 mg/day. Using a placebo-controlled, double-blind design, the authors administered three drinks each (placebo, 75 mg caffeine, and 150 mg caffeine) to 24 habitual caffeine consumers (mean caffeine intake = 217 mg/day) and 24 habitual non-consumers (mean caffeine intake = 20 mg/day). Each of the three previously listed drinks was administered in one of three separate sessions, with 48 hours between sessions, and measures of mood and cognitive performance were taken thirty minutes post-drink in each session. Haskell et al. (2005) reported that some mood effects of caffeine were more prevalent among habitual caffeine consumers than among habitual non-consumers. Self-reports of alertness increased significantly for habitual consumers at both 75 and 150 mg doses, but there were no significant increases in self-reported alertness for non-consumers at either dose. Additionally, both consumers and non-consumers exhibited significant decreases in self-reported mental fatigue, but this decrease occurred among consumers only at the 150 mg dose and among non-consumers only at the 75 mg dose (Haskell et al., 2005).

The review conducted by Ruxton (2008) included a similar conclusion regarding more prevalent positive effects with caffeine consumption for regular consumers as compared to non-consumers. In this review, the studies comparing groups of habitual consumers with habitual non-consumers primarily showed that the habitual consumers exhibited larger performance and mood effects after caffeine consumption as compared to habitual non-consumers. This conclusion is considered in the context of the argument that regular caffeine consumption leads to tolerance to caffeine and mood effects; if this were the case, the opposite pattern of results should have been found.

Rogers (2007) suggests several potential reasons for why non-consumers may be non-consumers: 1) they have experienced little or no benefit from prior caffeine use; 2) they already experience high levels of functioning without caffeine; 3) they experience negative effects upon caffeine consumption. Additionally, there are religious groups whose members refrain from caffeine consumption; specifically, coffee and tea consumption among Mormons and Seventh Day Adventists is strongly discouraged (Troyer, 1988). While some of this differentiation may be based on personal preferences and/or perceived benefits or lack thereof, there is also some support for a genetic basis for individual variability in response to caffeine. Researchers found a significant relationship between two linked polymorphisms on the adenosine (A_{2a}) receptor gene and self-reported anxiety after administration of a 150 mg caffeine dose to infrequent caffeine users. This same genetic variation has also been implicated in panic disorders, suggesting that the anxiety-inducing mechanisms related to panic disorders may be the same as those responsible for strong anxiety reactions to caffeine consumption in some individuals (Alsene, Deckert, Sand & de Wit, 2003).

1.2.1 Caffeine dosage.

As mentioned previously, the effects of caffeine on mood and performance appear to be related to dosage and habitual use. Smith (2002), for example, noted the importance of distinguishing between the effects of caffeine consumed normally (throughout the day) from food and drink sources, and caffeine delivered in a very large single dose (e.g., much more than would normally be consumed from food and drink sources). Most caffeine studies deliver caffeine in a single large dose, equal to or larger than the amount of caffeine a consumer would typically ingest in a day. Since caffeine is most often

ingested in a number of smaller doses uniformly distributed throughout the day (Barone & Roberts, 1996) rather than in one single dose, it is important to understand whether findings from studies where caffeine is administered in a single large dose will generalize to more typical patterns of caffeine consumption.

In an effort to specifically address this issue, Brice and Smith (2002) found that beneficial effects of caffeine on mood and performance were obtained both when a single 200 mg dose of caffeine was administered and when four hourly doses of 65 mg caffeine were administered. Performance and mood effects of caffeine have been found even at extremely low doses. Smit and Rogers (2000) found that caffeine doses as low as 12.5 mg led to significant reductions in RT (a reduction of approximately 3% in mean RT) on a simple RT task, as well as prevented an increase in self-reported boredom that was found in the placebo group toward the end of testing. Smit and Rogers (2000) concluded that the dose-response curve is relatively flat between doses of 12.5 mg and 100 mg.

In a study of the effects of repeated doses of caffeine, Smith et al. (2005) reported that when only placebos were given, performance and mood decrements were found in accordance with predicted effects of prolonged testing (2.5 hours of testing). Several simple, lower-level tasks were used in this study (e.g., a focused attention task, a categoric search task, a simple RT task, and a repeated digits vigilance task); self-reports of mood (e.g., drowsy vs. alert, happy vs. sad) were also collected. Effects of caffeine found early in testing (when participants were assumed to be alert) included improvements on vigilance tasks (d = .81), increased ratings of alertness (d = .70), and improved encoding of new information (d = .58). Effects of caffeine found later in testing (when participants were presumed to be fatigued) included improvements in performance

(increased number of hits) on the repeated digits task (f = .33), decreases in mean reaction times for categoric search tasks (f = .40) and simple RT tasks (f = .74), a decreased frequency of attentional lapses (f = .52), and increases in ratings of alertness (f = .60). Additionally, the decreases in RT on a categorical search task, decreases in simple RT, and decreased frequency of attentional lapses on a categoric search task were significantly larger for a higher caffeine dose (3.0 mg/kg body weight; the dose is calculated individually depending on the participant's body weight) than for a lower caffeine dose (1.5 mg/kg body weight). These findings are reflected in the linear dose-response relationship (for doses ranging from 1.5 to 3.0 mg/kg body weight) found for caffeine on measures of simple reaction time and frequency of attentional lapses as well as on self-reported alertness (Smith et al., 2005).

More specific findings about the impact of repeated caffeine doses are reported by Hewlett and Smith (2007). These authors administered either one or two doses of caffeine to alert participants (each dose was 1 mg caffeine/kg body weight) and then tested them on alertness and vigilance task performance. They found a significant effect of caffeine on self-reported alertness after task performance; the magnitude of the effect was related to the number of doses, such that alertness was approximately 14% higher for one dose as compared to no dose, 26% higher for two doses as compared to no dose, and 11% higher in two doses as compared to one dose. A similar finding was found for vigilance task performance, such that performance was 12% higher for one dose as compared to no dose, 19% higher for two doses as compared to no dose, and 6% higher for two doses as compared to one dose. These findings confirm the linear dose-response relationships found in prior research between both caffeine and alertness and caffeine and vigilance

task performance, at least up to a certain dose (2 mg/kg body weight; Hewlett & Smith, 2007).

The Ruxton (2008) review suggests that these findings in support of a linear dose-response relationship of caffeine may be somewhat isolated. Fifteen studies addressed this topic; only two of them (one of which is the Smith et al., 2005 paper discussed above) found support for a linear dose-response relationship. The other thirteen studies found that the majority of caffeine's beneficial effects were realized after the first dose, and that subsequent doses carried little or no additional benefit. The extent to which benefits can be conferred by later caffeine doses was not addressed by the current study, but represents an interesting area for future research.

A dose of caffeine also impacts individuals differently as a result of individual differences in body weight. Some studies (e.g., Christopher et al., 2005; Smith et al., 2005; Yeomans, Ripley, Davies, Rusted, & Rogers, 2002) have dealt with this issue by delivering caffeine doses that are proportional to body weight (i.e., 2 mg/kg, which would equal approximately 180 mg caffeine for a 200 lb., or 91 kg., individual). While this methodology allows for greater confidence in the standardization of the dosage for all subjects, there were several reasons against taking this approach in the current study. First and foremost, one of the objectives of this study was to make use of gum as a caffeine delivery vehicle, and this method of delivery does not allow for the manipulation of the dosage to be proportional to body weight. Second, there is a precedent in the literature for using a discrete caffeine dose (e.g., Brice & Smith, 2002; Haskell et al., 2005; Lieberman, Tharion, Shukitt-Hale, Speckman, & Tulley, 2002; Rogers, Richardson, & Dernoncourt, 1995; Smit & Rogers, 2000). Third, weighing participants

has the potential to bring about feelings of discomfort and embarrassment, and the possible benefits of using this approach did not outweigh the potential risks. I concluded that this weight-based dosing method was beyond the scope of precision necessary for the current study, and the discrete dosing method was employed.

While consideration of the experimental treatment dose and delivery method is important, researchers must also take care in selecting a placebo substance when implementing a placebo-controlled design. With regard to the selection of this placebo substance, Hollingworth (1912) emphasized the importance of preventing subjects from knowing whether they were getting placebo or caffeine. He stated that the two substances he used were completely indistinguishable from one another, and further suggested that subjects' inability to discern when they were getting caffeine helped diminish expectancy effects. Because the current study employed a between-subjects design, it was not as important that the experimental and placebo treatments be indistinguishable, but rather that they both be unfamiliar to participants so the two types of gum would not be identified as something they had chewed before.

In summary, several important conclusions can be drawn from extant research on caffeine dosage. First, it appears that, all other things equal, doses ranging from as low as 12.5 mg up to several hundred mg appear to exert a positive effect on subjective mood states and performance outcomes. Second, the individual's pattern of daily caffeine usage may alter the impact of a given dosage (e.g., 100 mg), resulting in different patterns of effects of caffeine in the context of performing various simple tasks. Third, while weight-based caffeine doses are more precise, the constraints and goals of the current study lend themselves more readily to a discrete caffeine dosing method. Fourth,

given the widespread belief that caffeine exerts positive effects on mood and performance, it is important to consider whether individuals in experimental vs. placebo groups show any similarities that could be the result of expectancy effects. That is, the placebo condition must be compelling in order to minimize the potential confound of expectancy effects.

1.2.2 Task type.

The effects of caffeine on performance have been studied using a variety of different tasks, the majority of which are simple tasks (e.g., lower-level RT or psychomotor tasks). Commonly used tasks include categorical search tasks (Brice & Smith, 2002; Christopher et al., 2005; Smith et al., 2005), choice RT tasks (Brice & Smith, 2002; Hindmarch, Quinlan, Moore, and Parkin, 1998; Hindmarch, Rigney, Stanley, Quinlan, Rycroft, and Lane, 2000; Lieberman et al., 2002; Patat et al., 2000; Scholey & Kennedy, 2004), repeated digits vigilance tasks (Christopher et al., 2005; Scholey & Kennedy, 2004; Smith, Kendrick, Maben, & Salmon, 1994; Smith et al., 2005), and simple reaction time tasks (Brice & Smith, 2002; Scholey & Kennedy, 2004; Smit & Rogers, 2000; Smith et al., 1994). Results of such studies generally show that caffeine improves performance on vigilance and other simple tasks that benefit from a high level of alertness (Smith, 2002). However, as Smith (2005) noted, the use of primarily simple, lower-level tasks makes it difficult to generalize these findings with respect to more complex cognitive processes.

A few studies have examined the effects of experimental caffeine manipulations and regular caffeine consumption on more complex tasks. Haskell et al. (2005) found that a caffeine manipulation (dose was equal to either 75 or 150 mg) was positively associated

with performance increases on a sentence verification accuracy task (f = .50). Peeling & Dawson (1997) found that a 100 mg dose of caffeine increased self-reported alertness and ability to concentrate during a college lecture. Kohler, Pavy, & Van Den Heuvel (2006) found that among sleep-deprived subjects, caffeine administration (dose = 200 mg) led to faster and more accurate responses on a grammatical reasoning task as compared to a group whose treatment was chewing a tasteless, odorless substance. Caffeine administration (3.5 mg/kg body weight) was found to be associated with increased systematic processing of the arguments presented in a persuasive message (Martin, Hamilton, McKimmie, Terry & Martin, 2007). Results obtained by Jarvis (1993) suggested positive effects of regular caffeine consumption on incidental verbal memory and visuo-spatial reasoning tasks. Additionally, researchers have found that low (i.e., less than or equal to 135 mg caffeine/day) caffeine consumption was associated with more accurate performance on a time estimation task (f = .33) than high caffeine consumption (i.e., more than 135 mg caffeine/day) or no caffeine consumption (Stine, O'Connor, Yatko, Grunberg, & Klein, 2002).

Hollingworth (1912) observed that the effects of caffeine on lower order, motor processes (i.e., simple tasks) tend to occur quickly, but are transient, while the effects of caffeine on higher order mental processes (i.e., complex tasks) take longer to become apparent but are more persistent. Hollingworth (1912) found that the duration of the effects of caffeine on tests of motor speed and coordination (i.e., simple tasks) ranged from two to four hours while the effects of caffeine on tests of association and choice (i.e., complex tasks) lasted into the next day. With regard to action time (i.e., the time it takes for the effects of caffeine to appear), performance on more simple tasks (i.e., motor

speed and coordination tests) was affected by caffeine within 45 to 90 minutes after caffeine ingestion while performance on complex tasks (i.e., association and choice tasks) was affected by caffeine within two to three hours after caffeine ingestion (Hollingworth, 1912). The same range of caffeine doses was used for all tests; in general, larger caffeine doses were associated with the shorter action times. Additionally, some of the action times were prolonged when the caffeine dose was taken with food. In general, the effects of caffeine on more simple tasks (i.e., motor speed and coordination tests) were observed more quickly than the effects of caffeine on more complex tasks (i.e., association and choice tasks), regardless of dose.

The first objective of the present study was to extend the domain of caffeine research from relatively simple tasks to a more complex verbal task. Given the benefits to performance on simple, lower-level cognitive tasks, as well as modest support for positive performance effects on moderately complex cognitive tasks, caffeine administration was expected to have a positive effect on performance on a complex verbal task.

1.3 Cognitive Fatigue

One of the most popular effects of caffeine is its putative influence on reducing cognitive fatigue. Although this reduction of cognitive fatigue is often closely associated with perceptions of greater mental alertness and sustained performance over time, a substantial research literature has evolved to suggest that cognitive fatigue, alertness, and sustained performance represent only partially overlapping constructs and may have different determinants and consequences. To better understand the relations among these

constructs and the influence of caffeine on them, I provide a brief review of the cognitive fatigue literature, with a focus on the definition and measurement of cognitive fatigue.

1.3.1 Defining cognitive fatigue.

A review of the research literature indicates several common features in various definitions of cognitive fatigue. Specifically, cognitive fatigue is typically defined with respect to a number of contextual factors, including: 1) the importance of time as a variable related to fatigue, 2) the individual as being continuously engaged in some sort of activity, and 3) the existence of a functional relationship between elapsed time and activity. In addition, cognitive fatigue is typically associated with several types of behavior changes, including changes in work performance, physiological processes, and subjective feelings (Cameron, 1973).

Van der Linden, Frese, and Sonnentag (2003) defined cognitive fatigue as "a psycho-physiological state resulting from sustained performance on cognitively demanding tasks and coinciding with changes in motivation, information processing, and mood" (p. 484). They also suggested that a common result of cognitive (or mental) fatigue is a decrease in task engagement and an increased resistance to further exert or allocate effort to the task (van der Linden et al., 2003). Consistent with other definitions, the van der Linden et al. (2003, p. 484) definition emphasizes the role of time ("sustained performance"), engagement in an activity ("cognitively demanding tasks"), a link between the activity and elapsed time ("resulting from sustained performance on cognitively demanding tasks"), and expected effects on affective, cognitive, and conative outcomes ("changes in motivation, information processing, and mood").

Another fatigue conceptualization is provided by Arai (1912), in the context of her investigation of the effects of mental fatigue on physical processes, feelings of fatigue, and intellectual efficiency. These studies consisted primarily of a mental multiplication task, but also included other tasks such as memorizing nonsense syllables. The author herself was a subject of the experiments, and used the technique of introspection to provide an insightful description of her personal experience of the feelings of mental fatigue: "I found in them emotions of repugnance at the thought of certain forms of mental activity, amounting sometimes to a sort of mental nausea; feelings of dullness or stupidity…cravings for certain familiar forms of mental relaxation, feelings of sleepiness, heavy feelings in the head…" (p. 72). She defines mental fatigue as "the unfavorable and destructive effect which continuous mental work produces on mental functions" (p. 29). This definition encompasses the four key factors described previously, as well as a reference to the negative performance outcomes of mental fatigue.

Arai's definition emphasizes the multiple intraindividual sources of mental fatigue, building on Thorndike's (1900) two theories of mental fatigue and a thorough review of the caffeine literature prior to 1912. Thorndike's first theory of mental fatigue (the mechanical theory) proposes that sustained mental work leads to a gradual decrease in the efficiency of mental functions, and that the decrease which occurs is proportional to the amount of mental work done. This theory emphasizes the notion that mental fatigue results from a decrease in mental efficiency across extended time-on-task. In contrast, Thorndike's second theory (the by-product theory) posits that mental fatigue is a complex phenomenon, and that the mind produces by-products that include feelings of

weariness, headache, and sleepiness as a result of mental activity. These by-products in turn decrease the mind's ability to do mental work. Thus, the by-product theory suggests that mental fatigue is not just a result of sustained energy expenditure but rather a consequence of the detrimental effects that such sustained energy expenditures produce. Using a series of laboratory experiments designed to contrast these two theories, Thorndike found support for the by-product theory, concluding that "the causes of fatigue are not mere decrease of energy, but highly complex by-products of work" (Arai, 1912, p. 14).

1.3.2 Delineating the construct space.

The construct of cognitive fatigue has also been differentiated from boredom. For example, Myers (1937) suggested that while mental fatigue and boredom may have the same performance outcome (decreases in the quality and/or quantity of output), they are quite different. He suggested that when an individual begins a mental task, most or all of his/her attention is allocated to the task and not towards other things which could occupy his/her attention, and that this direction of attention does not require much effort. As initial task interest fades, and no other potential interests to support continued task performance are available to replace them, boredom sets in. Specifically, Myers (1937) defined boredom as "the outcome of a failure to find interests which can maintain spontaneous or voluntary attention" (p. 298). It can be thought of as a fatigue of specific interests.

This concept of boredom can be juxtaposed with Myers' definition of fatigue: "a general impotence to concentrate attention and to act purposefully, intelligently and creatively" (p. 299). According to Myers (1937), fatigue occurs over the time course of

prolonged mental activity in a predictable pattern: first, a fatigue of specific interests (i.e., boredom), then a fatigue of general self-activity, and lastly, feelings of irritation, worry, anxiety, and an inability to recall memories, to consider relevant issues, and to view a problem in its entirety and reach a solution. As proposed by Myers (1937), feelings of boredom would act as a sufficient but not necessary condition preceding feelings of fatigue, and could serve as a safeguard against further mental activity which may result in severe mental fatigue.

Myers (1937) also suggested that feelings of cognitive fatigue may serve as the cause, rather than the consequence, of performance decrements after sustained mental activity, and that "surely the affective processes, conscious and unconscious…must be a potent factor in determining the signs and symptoms of fatigue" (Myers, 1937, p. 299). By this argument, feelings of fatigue may not reflect actual cognitive fatigue (which would be expected to result in performance decrements), but rather function to warn the individual against maintaining the current level of activity. If this is the case, one would expect to see subjective reports of fatigue prior to the onset of fatigue-related performance decrements. Ackerman, Kanfer, Wolman, and Haag (2008) obtained empirical support for Myers' conceptualization in a study of prolonged complex verbal task performance. As expected, self-reports of subjective fatigue increased over the course of the four hour testing session, but test performance at the aggregate level remained stable.

1.3.3 Measuring cognitive fatigue; individual differences in cognitive fatigue.

The measurement of cognitive fatigue is a major issue in research on the effects of caffeine on performance. Since cognitive fatigue is typically induced by increasing time-

on-task, measures of cognitive fatigue are often inferred by assessing changes in performance or self-reported subjective states over time-on-task. Fatigue researchers typically use a performance measure as the criterion rather than a measure of physiological changes or a self-report of subjective feelings; Cameron (1973) attributed the lack of focus on physiological changes to the difficulty involved both in measuring physiological changes and in linking them to other more practical and real-world relevant variables such as output, accidents, or performance quality. Research using performance measures has not shown dramatic, reliable decreases in performance with increasing time-on-task; this lack of findings has been attributed to individual differences in motivational processes used to compensate for fatigue effects on performance (Cameron, 1973). For example, given a sufficiently high incentive for task performance, an individual may exert and maintain a higher level of effort under high task demands than he/she would, given no incentive for performance. Cameron (1973) suggested that individual differences and situational factors may partially account for the lack of consistent performance deficits under conditions of fatigue.

Cameron (1973) noted that "if a task is boring and repetitive, or the circumstances provide little stimulation, the natural process is likely to take precedence, and low arousal will result" (p. 644). Consistent with earlier conceptualizations by Thorndike (1912) and Myers (1973), Cameron suggested that boredom serves as an antecedent for changes in arousal that result in mental fatigue and performance decrements. From an arousal perspective, Cameron's explanation is consistent with the Yerkes-Dodson law, which suggests an inverted U-shaped curve in the graph of stress or arousal (on the x-axis) and performance (on the y-axis), with high and low levels of stress associated with inefficient

performance and moderate levels of stress associated with maximal performance (Revelle, Amaral, & Turriff, 1976; Revelle, Humphreys, Simon, & Gilliland, 1980). In other words, if a task is boring and/or incentives for performance are low, then increasing time-on-task is expected to result in performance decrements as a consequence of the influence that boredom has on decreasing arousal and increasing mental fatigue. In contrast, if a task is interesting and/or incentives for performance are high, then arousal and effort are expected to be higher, and increasing time-on-task is less likely to yield performance decrements, up to a point (i.e., this expectation would not hold if the individual is excessively hungry or sleepy).

As Thorndike (1912) suggested, the effects of cognitive fatigue on performance must also be considered in the context of individual differences in reactions (e.g., mood, feelings, etc.) to a situation and capacity to adapt to the situation. A variety of cognitive, affective, and conative variables may interact to influence interpretations of a situation as boring or mentally fatiguing (Ackerman & Kanfer, 2006). Cameron (1973) echoed this idea, stating that it cannot be determined in advance whether a situation will be fatiguing, and that the uncertainty in attempting to do so is compounded by individual differences. This idea has implications for the current study in that substantial variations in individuals' subjective reactions to the fatiguing situation were expected.

In a study of individual differences in pilot performance, Davis (1946) observed that participants differed in the manner in which they transitioned from the early to the late stage of performance testing. Using the Cambridge Cockpit flight simulator, 355 experienced pilots repeated 10-minute sets of maneuvers periodically during a 50-minute simulation. Davis (1946) found that approximately 75% of all participants (termed the

"normal" group) showed little or no change in performance over the course of the experiment. In contrast, approximately 15% of all participants (termed the "overactivity" group) showed strong task reactions across the session, such as sustained attempts to improve their performance, emotional distress, preoccupation with the task even after it was finished, and flushing of the face. The remaining 10% of all participants (termed the "withdrawal" group) appeared to lower their standards of performance over the course of the session, were not emotionally disturbed by errors later in the session, and reported becoming bored and even day-dreaming during the session. Davis (1946) concluded that the test was indeed fatiguing, and that the skills needed to carry out the flight simulation became "disorganized" as a result of the fatiguing situation (p. 28). These results also indicate that individuals differ in their subjective reactions to the same fatiguing situation.

Recent evidence suggests that one might be better served by making different predictions for specific groups of individuals depending on the strategies employed during testing, rather than expecting consistent response patterns across all individuals. The same groups identified by Davis (1946) were found by Ackerman et al. (2008) in their investigation of the effects of four hours of continuous ability testing on cognitive fatigue and related outcomes. In this study, participants were classified into one of four groups based on the extent to which their performance data included "drops" (decreases of more than 20% from the second to the fourth test within a given hour) and/or "spurts" (increases of more than 20% from the second to the fourth test within a given hour). The normal group, which included 30% of participants, exhibited neither spurts nor drops during the course of testing, indicating that they either were unaffected by fatigue or were able to increase effort just enough to counteract fatigue effects without overshooting and

improving performance dramatically. The overactivity group, which included 44% of participants, exhibited a spurt and no drops, indicating that they increased effort at some point during testing, most likely to counteract fatigue effects, and in doing so improved their performance. The withdrawal group, which included 10% of participants, exhibited a drop and no spurts, suggesting that they experienced fatigue effects, did not sufficiently increase effort to counteract them, and consequently experienced a decrement in performance. Ackerman et al. (2008) also conceptualized a fourth group which was hypothesized but not found by Davis (1946). The combined overactivity/withdrawal group, which included 14% of participants, showed at least one drop and at least one spurt during testing. It is difficult to hypothesize about the efforts (or lack thereof) of this group, as group members displayed several different patterns, including a drop followed by a spurt, a spurt followed by a drop, and two drops followed by a spurt.

Another study which illustrates the occurrence of different patterns of performance with increasing time-on-task for individuals was conducted by Martyn (1913). In this study, three subjects performed mental multiplication (two single-digit figures multiplied together, then that product multiplied by a third single-digit figure) for one hour per day for five days. Other tests, including measures of muscular capacity, spatial threshold, rate of respiration and pulse, and quickness and accuracy of perception, were administered before and after the periods of mental multiplication. She concluded that "the signs of fatigue differ according to the individual; that the response which the individual makes in the first stages of fatigue depends on the stability of dispositions previously acquired, on the method of working…" (Martyn, 1913, p. 445). A different pattern of performance effects emerged for each of the three subjects. One subject

showed acceleration in rate of performance, such that any decreases in rate (termed loss of incitement) were more than compensated for by acceleration in rate attributed to recovery from fatigue (termed recuperation). A second participant showed substantial loss of incitement with very little accompanying recuperation, which resulted in an overall decreasing rate of performance. A third participant exhibited relatively stable performance over the hour, with little evidence of either loss of incitement or recuperation (Martyn, 1913). These patterns of responses seem to parallel the overactivity, withdrawal, and normal (respectively) patterns identified by Davis (1946).

Although the studies investigating the effects of time-on-task on performance employ different tasks and small samples, the findings suggest that there may be individual differences in both subjective feelings of cognitive fatigue as well as reactions to these feelings. Combined with broad conceptualizations of mental fatigue by Thorndike and others, it is reasonable to expect that individual differences in feelings of fatigue and reactions to these feelings will exert differential influences on performance over time-on-task. The present study evaluated the effects of cognitive fatigue on both subjective feelings of fatigue and related constructs, as well as on performance on a complex cognitive task, by focusing on individual differences in both the experience of and performance effects related to cognitive fatigue.

1.3.4 Time-on-task, task demands, and cognitive fatigue.

An important observation made by Arai (1912) was the differentiation between practice effects and fatigue effects. She defined practice effects as "the favorable influence which the continued exercise of a function exerts on its own efficiency and which does not entirely disappear within a few minutes" (p. 80). In her work on the

mental multiplication task, which she performed for four to twelve hours a day over numerous days, she observed that on the first day, numerous fluctuations occurred in the time taken to do one problem. She concluded that these fluctuations were due to lack of practice on the task rather than fatigue. As she became more accustomed to the task, time irregularities decreased and the performance curve became smoother. In the early days of testing, fluctuations were observed near the start of the testing period, while in later days of testing, fluctuations were observed closer to the end of the testing period. Arai (1912) attributed fluctuations occurring earlier in the session to lack of practice, and fluctuations occurring later in the session to fatigue. Arai (1912) cautioned that if performance effects are to be examined in the context of fatigue, the participant must be at or near the limit of improvements related to practice. The notion of practice effects was not a substantial consideration for this study, as the task that was employed (the Cloze task, see the following section for a description) is one which typically exhibits a shallow practice curve (Ackerman et al., 2008).

A rough estimate of the amount of practice required may be obtained from studies conducted by Axel Oehrn in 1889 (reported by Arai, 1912). In these studies, Oehrn attempted to determine the work-curves of specific mental functions. For each mental function, he assessed the average time taken to reach what he called the point of maximum efficiency, which reflects the time at which the fatigue effect begins to outweigh the practice effect. These times ranged from 24 minutes for learning nonsense syllables to 60 minutes for memorizing lists of numbers.

Arai (1912) also noted that “difficult and disagreeable work brings about a decrease in the efficiency of the function exercised” (p. 114). One participant (whose data

were analyzed extensively) took almost twice as long to do mental multiplication problems at the end of a twelve-hour period as compared to the beginning. A group of eleven individuals showed a 24% increase in the time taken to do mental multiplication problems over the course of two hours. The author concluded that feelings of fatigue are somewhat correlated with a state of mental inefficiency, but cautioned that an individual can feel fatigued without having his/her performance affected by fatigue. Similar to the conclusion drawn decades later by Cameron (1973), Arai (1912) suggested that feelings of fatigue serve as a warning sign to stop work, and the associated symptoms are often experienced well before actual performance is negatively affected.

In a review of the literature on test length and cognitive fatigue, spanning from 1896 to 2006, Ackerman and Kanfer (2006) looked at method of performance assessment, task characteristics and cognitive fatigue, broad theories of cognitive fatigue, subjective effects of cognitive fatigue, trait determinants of cognitive fatigue, state variables related to cognitive fatigue, predictors of subjective effects of cognitive fatigue, and relationships between subjective perceptions of fatigue and performance effects. Similar to Cameron (1973), Ackerman and Kanfer (2006) concluded that mean decrements in cognitive test performance are likely to be small with respect to increases in time-on-task, for a period of up to several hours. Further, extant research suggests that individuals respond in several different ways to a fatiguing situation, such that some individuals are likely to increase task effort as a result of feeling fatigued, which will lead to maintenance of, or an increase in, performance over time, while others will decrease task effort upon feeling fatigued, which may lead to performance decrements. Still other individuals may first increase and then later decrease task effort. These findings hold only

if the task is dependent on mental effort. Performance under task conditions that engender cognitive fatigue is further complicated by individual differences in trait (i.e., ability, personality, motivation, susceptibility to caffeine), state (sleep deprivation, hunger, sickness), and experiential (i.e., previous involvement with similar stimuli or situations) variables (Ackerman & Kanfer, 2006).

While there are likely to be individual differences in experiences of, and reactions to, fatigue, there are numerous elements of a task which are more likely to cause fatigue overall. In order to examine the effects of caffeine on fatigued individuals, the task selected for the present study was one thought to be likely to induce fatigue with increasing time-on-task. To this end, special consideration was given to selecting a task that includes factors identified by Ackerman and Kanfer (2006) as particularly influential in increasing cognitive fatigue. Specifically, a modified version of the Cloze task (Taylor, 1953; see Ackerman, Beier, and Bowen, 2000 for a review of the Cloze test) was used in the current study; this task was constructed to place high demands on intellectual functioning, provide limited opportunities for knowledge of results/feedback, be rated low in terms of being intrinsically interesting or enjoyable, emphasize verbal content rather than math content, be constructed as experimenter-paced rather than learner-paced (i.e., rate of stimulus presentation is not controlled by the subject), and contain relatively homogeneous content (Ackerman & Kanfer, 2006).

The Cloze task (Taylor, 1953) was modified to fit all of the above qualifications. If the task was sufficiently fatiguing, mood effects should have become evident in the absence of caffeine administration. Based on findings from Ackerman et al. (2008), reports of subjective fatigue and NA were expected to increase with extended time-on-

task, reports of PA should have decreased with increasing time-on-task, and performance effects should have varied based on the strategy used by each individual (i.e., strategies employed by the normal, withdrawal, overactivity, and combined overactivity/withdrawal groups described previously).

1.3.5 Summary of Cloze and Completion study results.

Recent findings by Ackerman and colleagues provide normative performance information for the Cloze and Completion tasks over a four hour period. The Cloze task involves presenting participants with a passage that is approximately 250 words in length from which every 5th word has been deleted and replaced with a blank. The first and last sentences of the passages are left intact. The participant's task is to fill in the missing words. The Completion task is similar to the Cloze task except that the passage is read aloud in its entirety immediately before the participant attempts to fill in the missing words (the passage with the blanks is hidden from the participant's view while the passage is read aloud). In general, these tasks assess verbal fluency, with the Completion tasks having a memory component as well. The findings from the Ackerman et al. (2008) study are particularly relevant because the methodology employed is very similar to that used for the current study.

In the study by Ackerman et al. (2008), 99 participants were given 16 Cloze and 16 Completion tests, for a total of 32 tests. The Cloze and Completion tests were administered in four booklets of four Cloze and four Completion tests each, presented in alternating order. Each booklet took approximately one hour to administer. So within each hour, four Cloze tests and four Completion tests were given. Self-reports of subjective feelings were also collected throughout the testing session: prior to each hour

of testing and at the conclusion of the session, for a total of five collections of self-report data. Results obtained by Ackerman et al. (2008) on Completion task performance showed non-significant changes in mean performance when analyzed over the entire session. In contrast, mean performance on the Cloze test showed a significant but small improvement over the course of the entire session. When analyzed within each hour of testing, Cloze test performance showed significant performance decrements from Test #2 to Test #4. These effects were more robust in Hours 2, 3, and 4 as compared to Hour one (Hour one $t(98) = 1.62$, ns; Hour two $t(98) = 2.75$, $p<.01$; Hour three $t(98) = 4.01$, $p<.01$; Hour four $t(98) = 2.90$, $p<.01$). No similar patterns were found for Completion test performance. These results suggest that Cloze test performance was affected by fatigue while Completion test performance was not. As a result of this finding, I decided to use 32 Cloze tests in the present study rather than 16 Cloze and 16 Completion tests. An added benefit of this change was that the total test time was reduced by the time required to read the 16 Completion passages aloud.

A possible explanation for this difference in findings is that when individuals fill out the Completion test they have already heard the passage read aloud, while on the Cloze test they have not. Thus, the Cloze test is a more difficult task because the participant has less knowledge of the context of the passage than when he/she has heard the entire passage read aloud. Terman (1906) explained this difference (his explanation, though not directly concerning modern-day Cloze and Completion tests, addresses an almost identical pair of tasks), suggesting that: reading the passage aloud "rob[s] the test of its puzzle nature…[after having the passage read aloud, the participant] therefore knew the general sense and had a much narrower field to hunt over in the search for suitable

words" (cited in Ackerman et al., 2000, p. 106). Another potential explanation is that when the passage is read aloud, the participant can sit still and rest, so that the time during which the passage is read aloud (1-2 minutes) provided an opportunity for recovery, thus alleviating fatigue to some degree before each Completion test is completed. It also appears that recovery took place during the four-minute self-report questionnaire which was administered between each set of tests (i.e., between each hour of testing).

In contrast to the relative stability of performance over time-on-task, subjective reactions to testing were adversely affected over time. Self-reports of fatigue ($F(3, 294) = 11.01$, $p < .01$) and NA ($F(3, 294) = 10.48$, p, $<.01$) increased from Hour one to Hour four, and PA decreased ($F(3, 294) = 35.20$, p, $<.01$) over the course of testing. Findings obtained in the Ackerman et al. (2008) study suggested that the Cloze task is subject to fatigue effects on performance, but that these effects may be attenuated by even brief periods of recovery (i.e., four minutes to complete the subjective feelings questionnaire). In the current study, the number and length of breaks and questionnaires was limited in order to inhibit recovery.

1.4 Methodological issues

1.4.1 Determination of caffeine dose.

Identifying the appropriate caffeine dosage to elicit mean behavioral effects has proven difficult in previous research. Ideally, the dose of caffeine administered would be sufficiently large to produce a mean effect but not large enough to pose health risks or to exceed the amount of caffeine a person would typically ingest in a day through typical vehicles (i.e., coffee, tea, caffeinated sodas, energy drinks).

The appropriate dose for the current study was determined in part by using findings from previous research on patterns of typical daily caffeine consumption in the U.S. population. Frary et al. (2005) used data from the US Department of Agriculture Continuing Survey of Food Intakes by Individuals from 1994 – 1996 and 1998 to form overall estimates of caffeine consumption in the population. They suggested that the mean daily caffeine intake for habitual caffeine consumers in the U.S. during these periods was 193 mg. daily (which is approximately equal to the caffeine content of a 12.oz Starbucks coffee; see Table 1). Barone and Roberts (1996) formed population estimates for daily caffeine intake using archival data from the 1989 Market Research Corporation of America Survey, the 1977 – 1978 Nationwide Food Consumption Survey, and the 1993 Winter Coffee Drinking Study. They found that caffeine consumers' average daily caffeine intake was approximately four mg/kg body weight. Since the weight of an average US adult is 74 kg (for females) and 86 kg (for males), this would result in an average daily caffeine estimate of 296 mg for females and 344 mg for males, for an average of 320 mg per day. The higher estimates obtained by Barone and Roberts, compared to those of Frary et al. may be due to the fact that the Frary et al. estimates were based on caffeine intake of both adults and children, whereas the Barone and Roberts (1996) study developed estimates based on data taken exclusively from adults.

Nawrot et al. (2003) examined the possibility that caffeine consumption negatively affects human health by reviewing studies resulting from their comprehensive review of the literature on caffeine use in humans. Based on their review, Nawrot et al. (2003) concluded that daily caffeine consumption of up to 400 – 450 mg/day was associated with no adverse health effects. These findings, taken with those findings

related to caffeine dosage that were discussed earlier, suggest that caffeine dosages between 200 and 250 mg represent an effective dose for most adults; that is, large enough to yield an effect, comparable to the daily intake of an average caffeine consumer, and well under the lower end of the range of daily intake values said to be associated with no adverse health effects (i.e., 400 mg/day from Nawrot et al., 2003).

1.4.2 Caffeine delivery.

Previous research has used a variety of methods to administer caffeine, including pills and capsules (Childs & de Wit, 2008; Smit & Rogers, 2000; Lieberman et al., 2002), beverages that naturally contain caffeine (e.g., coffee, tea; see Hindmarch et al., 2000; Hindmarch et al., 1998), powder dissolved in a beverage that does not contain caffeine (e.g., juice, decaffeinated coffee, decaffeinated tea; see Haskell et al., 2005; Brice & Smith, 2002; Christopher et al., 2005), and caffeinated gum (Kamimori, Johnson, Thorne, & Belenky, 2005; Kamimori et al., 2002; McLellan, Bell, Lieberman, & Kamimori, 2005; Smith, 2009; Syed, Kamimori, Kelly, & Eddington, 2005). The administration of caffeine through beverages and powder dissolved in non-caffeinated beverages affords a more naturalistic experience for participants (as opposed to a pill/capsule or gum), as well as the ability to compare effects associated with caffeine alone and effects associated with coffee, tea, etc. In contrast, capsules and pills represent more obtrusive manipulations but allow for more precise measurement of dose and standardized administration.

Several recent studies (Kamimori et al., 2002; Kamimori et al., 2005; McLellan et al., 2005; Smith, 2009; Syed et al., 2005) have used gum as the vehicle for caffeine delivery. The primary advantage of using gum over traditional caffeine delivery vehicles

(i.e., pills, powders) is that the gum has a significantly faster rate of absorption while providing a nearly equal amount of caffeine to the systemic circulation when compared to a caffeine capsule (Kamimori et al., 2002). Eighty-five percent of the caffeine contained in the gum is delivered after five minutes of chewing (Syed et al., 2005). The highest caffeine plasma concentrations were found between 44.2 and 80.4 minutes after administration in a group of 84 healthy, non-smoking males; these peak concentrations were reached in nearly half the time with the gum formulation compared to the capsule formulation (Kamimori et al., 2002). In other studies, the caffeine gum has been associated with maintenance of vigilance performance, running times, and marksmanship performance after overnight sleep deprivation (Kamimori et al., 2005; McLellan et al., 2005).

In a 2009 study, Smith used caffeine gum to investigate mood and performance effects of small caffeine doses. In this study, participants chewed placebo gum, chewed caffeine gum, or did not chew gum. The caffeine dose used in this study was 40 mg. Smith found that caffeine gum exerted positive effects towards the end of the study, when participants had become fatigued, compared to the placebo gum and no gum groups. In assessments of mood carried out after a one-hour testing session, participants who had chewed the gum reported higher levels of alertness and hedonic tone (happiness) as compared to the other two groups. So the low-dose caffeine gum was associated with mood effects that are generally found to occur with caffeine consumption. For the performance component of the study, several low-level, reaction-time and vigilance-type tasks were used. These types of tasks generally show performance improvements associated with caffeine administration. On all tasks except for the simple RT task, the

caffeine gum group performed significantly better during the one-hour testing session than the placebo gum and no gum groups. So Smith demonstrated that low caffeine doses administered via gum produce similar mood and performance effects to those found with larger doses delivered in more traditional methods (Smith, 2009).

The current study used Stay Alert® caffeine gum for caffeine administration. This method of caffeine delivery, increasingly popular in military and law enforcement settings as a performance enhancer for use by individuals operating under conditions of sleep deprivation, affords more rapid administration of caffeine. To date, the caffeine gum has been used primarily in military studies, and these studies have been limited in number. Other brands of caffeine gum are sold in various drug stores, and Stay Alert® can be purchased from online retailers, making this vehicle for caffeine delivery readily available to the general population. However, caffeine gum has been used in only one research study not involving a military context (Smith, 2009). The use of this delivery method in the current study offers potentially greater generalizability to other real-world settings where such gum is already available to consumers. Looking beyond the current research focus on the use of caffeine gum in military settings, this product has potential benefits for other segments of the population (e.g., students, employees in jobs where long shifts are required, such as doctors, nurses, and truck drivers, etc.).

Given the rapid absorption rate of caffeine delivered in gum, the effects of caffeine on subjective feelings and performance were expected to begin occurring within 10-15 minutes of administration. Based on previous findings by Kamimori et al. (2002), peak effectiveness of the manipulation should occur approximately one hour after administration. Additionally, it is unlikely that the caffeine effects will "wear off" over

the course of four-hour testing sessions, as the elimination half-life of caffeine is estimated to be between 3 and 7 hours (Nawrot et al., 2003). Performance decrements may still occur, however, if the potential fatiguing effects of the task become greater than the potential positive performance effects of the caffeine during the course of the study.

1.4.3 Operational definition of habitual caffeine consumers.

The current study recruited regular caffeine users. This precedent has been set by other research (Brice & Smith, 2002; Christopher et al., 2005; Haskell et al., 2005; Hindmarch et al., 2000), in part because individuals who refrain from caffeine use may do so because of an abnormal sensitivity and/or atypical physiological reactions to its use (Rogers et al., 1995; Rogers, 2007). Additionally, researchers have found effects of caffeine on performance in certain tasks (i.e., simple RT, rapid visual information processing) to be more robust for high (i.e., more than 200 mg/day) rather than low (i.e., less than 100 mg/day) caffeine consumers, suggesting that individuals are more likely to consume higher amounts of caffeine on a daily basis if they experience greater positive effects from doing so (Smit & Rogers, 2000).

Similarly, Attwood, Higgs, and Terry (2007) found that high (i.e., > 200 mg/day) but not moderate (i.e., < 200 mg/day) caffeine consumers exhibited significantly faster reaction times on simple and choice reaction time tasks after caffeine administration as compared to after placebo. Participants were also asked to describe any felt effects of the substance they ingested during the experiment. Written responses were coded as positive, negative, or no effect. Positive effects included alertness, cheerfulness, increased well-being, and decreased drowsiness. Participants' responses to this questionnaire indicated that high caffeine consumers were also significantly more likely to report having

experienced positive effects of caffeine, while moderate caffeine consumers were significantly more likely to report having experienced no effects of caffeine. These authors concluded that the increased likelihood for high caffeine consumers to experience positive effects of caffeine is a likely factor contributing to their higher usage of caffeine (Attwood et al., 2007). The criteria for inclusion in the study in terms of caffeine consumption was represented as a range, based on previous research that has employed a range of daily consumption levels in determining which participants to include.

In a study conducted by Smit and Rogers (2000) on the effects of low caffeine doses on cognitive performance and mood, the authors divided participants into lower and higher caffeine consumption groups. Low caffeine consumers were defined as consuming less than 100 mg/day, while higher caffeine consumers were defined as consuming more than 200 mg/day (Smit & Rogers, 2000). Smith et al. (2005) excluded participants who did not consume any caffeine on a daily basis. Yeomans et al. (2002) studied the effects of caffeine on performance and mood as a function of the level of caffeine abstinence. Individuals were included in the study based on self-reports of daily caffeine intake that indicated daily caffeine usage falling between 195 – 650 mg/day. Self-report measures of daily caffeine usage were obtained by asking participants how many cups of coffee they drank in a day, what type of coffee they drank, how many cups of tea they drank in a day, and how many cans of cola they drank in a normal day. Participant responses were used to make an estimate of daily caffeine intake by multiplying each answer by the average caffeine content for each of those drinks (Yeomans et al., 2002). Haskell et al. (2005) defined caffeine consumers as those who drank tea and/or coffee, and consumed at least 50 mg of caffeine per day. The range of

daily caffeine consumption in this study was 60-800 mg/day, with a mean of 217 mg/day. It is clear from this sampling of caffeine studies that qualifications for typical caffeine consumers vary widely. It also seems that some of the studies (e.g., Haskell et al., 2005; Yeomans et al., 2002) include consumers with an extremely wide range of average daily caffeine consumption. In the current study, daily caffeine consumption was calculated based on information provided by participants over the phone prior to enrollment in the study; consistent with past research, only participants whose daily caffeine usage exceeds 200 mg/day were included in the study.

It was anticipated that the use of self-report measures to assess daily consumption of caffeinated beverages to calculate usage may have been problematic if participants were simply asked how often they consume caffeinated beverages. For example, someone who drinks a regular (i.e., a "grande®",16 oz.) cup of Starbucks coffee each day receives approximately 259.3 mg caffeine from this beverage (McCusker, Goldberger, & Cone, 2003). In contrast, an individual who consumes three 12-oz. cans of Coca-Cola Classic is getting approximately 88.5 mg caffeine from these three drinks (McCusker, Goldberger, & Cone, 2006). Although the Coca-Cola drinker consumes three caffeinated beverages as opposed to the Starbucks drinker's one caffeinated beverage, the Starbucks drinker is getting more than three times the amount of caffeine that the Coca-Cola drinker is getting.

To address this concern, I used a design similar to that used by Yeomans et al. (2002). During the phone scheduling/introduction to the study, participants were asked how many of each of the following beverages they consume in a typical day: caffeinated regular coffee, caffeinated espresso-based coffee, caffeinated iced tea, caffeinated hot tea,

caffeinated soda, and caffeinated energy drinks. Based on the caffeine content estimates put forth in scientific articles (e.g., Barone and Roberts, 1996; McCusker et al., 2003; McCusker et al., 2006) and other sources (i.e., websites such as Starbucks.com and EnergyFiend.com), average estimates for the caffeine contents of these six categories were calculated. Participants' answers were used to compute a daily caffeine intake estimate using these estimated caffeine content values. If their daily caffeine intake estimate was greater than 200 mg., they were invited to schedule an appointment for the study. If not, they were informed that they were ineligible to participate. The lower limit of 200 mg/day was intended to prevent participants from getting substantially more caffeine during the study than they would typically have in the context of a normal day. An upper limit was not set, as there appeared to be little precedent to do so based on the designs employed in previous studies.

1.4.4 Selection of dose.

Another consideration was the dose or amount of caffeine that was given to participants. Based on prior research, a dose of approximately 200 mg of caffeine per administration seemed optimal. This dosage was selected because it is consistent with average estimates of daily caffeine consumption for the US population (estimates were 193 mg/day for Frary et al., 2005 and 320 mg/day for Barone & Roberts, 1996), well below the daily consumption found to be associated with no adverse health effects (400 mg/day, Nawrot et al., 2003), and research suggests that several significant mood and performance effects emerge with a 200 mg dose as compared to a 100 mg dose, but few significant effects emerge with a 300 mg dose as compared to a 200 mg dose (Lieberman et al., 2002). It is also used as the only or maximal dose in many studies on the effects of

caffeine on humans (Kamimori et al, 2002; Kamimori et al., 2005; Syed et al., 2005; Brice & Smith, 2002). It seemed important not to include individuals whose daily caffeine intake is substantially less than the intended dose they would receive during the experimental sessions, so that individuals who self-regulate their caffeine intake to avoid reaching levels associated with negative effects (i.e., anxiety) would not receive a dose larger than they would typically consume, independent of participation in the study. Participants were administered either a single dose of 200 mg caffeine or placebo, in gum form, one hour into the testing session.

1.5 Summary of hypotheses

In summary, the purpose of this study was to investigate the influence of caffeine on complex task performance and associated self-reports of mood and fatigue. A between-subjects design, in which participants performed the Cloze task repeatedly in either a placebo or a caffeine condition, was employed.

Findings reported by Davis (1946) and Ackerman et al. (2008) indicate that individuals under conditions of cognitive fatigue react by employing different strategies. Performance varies as a result of the strategy chosen by each individual, so different patterns of performance were predicted for individuals based on the strategy they used. Also, findings by Ackerman et al. (2008) indicate that Cloze task performance over a four-hour period is relatively stable at the aggregate level.

Hypothesis 1a) In the placebo condition, performance on the Cloze test will vary between participants, such that some individuals will exhibit stable performance, some will show performance improvements with increasing time-on-task, indicative of increased effort to combat fatigue effects, some will show performance decrements,

indicative of fatigue effects and/or decreased effort, and some will show mixed periods of performance improvements and performance decrements.

Hypothesis 1b) At the aggregate level, performance on the Cloze test is expected to remain relatively stable, but individuals are expected to fall within these specific identifiable groups based on the strategies they use during testing.

In prior research, administering the Cloze task to participants over extended time-on-task led to increasing self-reports of fatigue and NA, and decreasing self-reports of PA (Ackerman et al., 2008). Similar results have also been found for placebo condition participants in caffeine studies (e.g. Smith et al., 2005). So it was expected that self-reports of positive mood would decrease while self-reports of negative mood and fatigue would increase with increasing time-on-task. More precisely:

Hypothesis 2) In the placebo condition, self-reports of fatigue and NA will increase, and self-reports of PA will decrease, over time-on-task.

Research has indicated that caffeine has beneficial performance effects on simple, lower-level cognitive tasks (e.g., reaction time tasks, vigilance tasks, psychomotor tasks), and has demonstrated modest support for positive performance effects on somewhat more complex cognitive tasks (e.g., time estimation tasks). Also, given that caffeine should alleviate feelings of fatigue, the drops (performance decrements) and spurts (performance improvements) that were expected to be seen as a result of fatigue in the placebo condition should have been less frequent in the caffeine condition. That is:

Hypothesis 3) In the caffeine condition, Cloze task performance will improve and drops and spurts will occur less frequently, in comparison to the placebo condition, such that there will be fewer performance decrements, fewer performance

improvements, and higher overall performance at the aggregate level after administration of caffeine. The expected effect size is medium, in terms of Cohen's d.

The effects of caffeine on subjective mood states are well-established: caffeine consumption engenders increased feelings of alertness and decreased feelings of fatigue. Caffeine has also been associated with increased feelings of ability to concentrate, motivation to work, and energy. Specifically,

Hypothesis 4) In the caffeine condition, self-reports of fatigue and NA will decrease, and self-reports of PA will increase, compared to the placebo condition.

CHAPTER 2

METHOD

2.1 Participants

An a priori power analysis was conducted in order to determine an appropriate sample size for providing sufficient power to detect an effect of the size that could be reasonably expected based on prior research. This analysis was done for the repeated-measures ANOVA and for the independent samples t-test, as these were the two most frequently proposed analyses in the analysis plan. An analysis of effect sizes based on values reported in the literature revealed average effect sizes of $d = .6$ and $f = .5$. Power analyses were done with a range of effect sizes below these average values, as it was expected that the present study might yield lower effect sizes due to the use of a more complex task, and using power values ranging from .80 to .95. It was determined that a sample size of 100 (50 per group) would yield power somewhere in the .80 to .90 range for effect sizes of at least $d = .5$ and $f = .4$. Specifically, for the repeated-measures ANOVA, a sample size of 13 (in each group) was suggested in order to achieve power of .90 with $f = .4$. And for the dependent samples t-test, a sample size of 102 (51 per group) was suggested in order to achieve power of .80 with $d = .5$. More participants than the proposed 100 were actually run in order to allow for higher levels of power.

One hundred and sixteen volunteers (61 men and 55 women) were recruited from the Georgia Institute of Technology undergraduate student population enrolled in a psychology class during the recruitment period. Volunteers for the study were recruited

using the School of Psychology Experimetrix program. Qualifications for participation in the experiment were clearly stated in the Experimetrix posting. Specifically, eligible participants had to be aged 18 years or older, with normal or corrected-to-normal vision, who consumed the specified amount of caffeinated beverages on a daily basis, and who were not pregnant or thought that they might have been pregnant. Even though research has not suggested consistent, causal links between regular caffeine consumption and adverse outcomes, doctors and researchers still tend to recommend that pregnant women limit caffeine intake to a lower amount than is recommended for the population at large. For this reason, this study excluded women who were pregnant or who thought that they might have been pregnant.

Interested and eligible individuals were instructed to call the laboratory to get more information and schedule an appointment. Potential participants who called the laboratory were given information and asked a few brief questions over the phone to assess their typical daily consumption of various caffeinated beverages. This information was then used to form an estimate of daily caffeine intake for each individual. Consistent with screening procedures used by Haskell et al. (2005), Smit and Rogers (2000), and Smith et al. (2005), individuals who reported caffeine consumption levels that were estimated at or above 200 mg/day using a standardized estimation form were included in the study. Participants signed a consent form prior to starting the study. It outlined the general components of the study (i.e., telling them that they would be asked to chew gum that may or may not be caffeinated and complete several hours of cognitive tasks, as well as reiterating that they must be typical caffeine consumers and must not be pregnant). Participants were compensated with four and a half extra credit research hours. The mean

caffeine consumption level among the sample was 268.38 mg/day (*SD* = 122.52 mg/day) for weekdays and 229.20 mg/day (*SD* = 156.30 mg/day) for weekend days. The mean age of the sample was 20.2 years (*SD* = 2.28 years).

2.2 Procedure

A between-subjects design consisting of one, 4.5-hour testing session was used. Participants were run in groups of two to twelve, and each group was randomly assigned to either the placebo or experimental (caffeine) condition. Fifty-eight participants (32 men and 26 women) were run in the experimental condition and 58 participants (29 men and 29 women) were run in the placebo condition. Participants completed an initial pre-session questionnaire assessing typical caffeine consumption habits and sleep, food intake, and caffeine consumption prior to the session. Next, participants alternated between a self-report questionnaire assessing current mood states and subjective feelings, a set of eight Cloze tests, another questionnaire, another set of eight Cloze tests, and so on, followed by a final self-report questionnaire at the end of the session, for a total of four sets of eight Cloze tests and five self-report questionnaires (one before each set of Cloze tests and one at the end of the session). Participants were given a five minute break after they completed the second set of Cloze tests, approximately half-way through the testing session. Please see Figure 1 for a layout of the session. Sessions were held on weekday and weekend mornings; the timing was standardized such that all sessions were conducted from 8:30 am to 1:00 pm.

The caffeine condition involved the administration of caffeine one hour after the start of testing. The primary purpose of delaying caffeine administration to one hour after the start of testing was to establish baseline performance and self-report data for

Pre-Session Questionnaire & Cloze Instructions (approx. 10 minutes)
includes 6 items on sleep, food, and caffeine consumption immediately prior to the session, 9 items on caffeine consumption in general, and Cloze test instructions and example

Interim Questionnaire A (approx. 8 minutes)
includes 46 self-report items

Cloze Booklet A (48 minutes)
includes 8 Cloze tests

Gum Administration (approx. 10 minutes)

Interim Questionnaire B (approx. 8 minutes)
includes 46 self-report items

Cloze Booklet B (48 minutes)
includes 8 Cloze tests

Five Minute Break (approx. 5 minutes)

Interim Questionnaire C (approx. 8 minutes)
includes 46 self-report items

Cloze Booklet C (48 minutes)
includes 8 Cloze tests

Interim Questionnaire D (approx. 8 minutes)
includes 46 self-report items

Cloze Booklet D (48 minutes)
includes 8 Cloze tests

Task Perceptions Questionnaire (approx. 10 minutes)
includes 56 self-report items: 46 are the same as previous interim questionnaires, also includes items on perceptions of gum, performance during session, plans immediately following session, and regular caffeine consumption habits

Debriefing and assignment of credit (approx. 8 minutes)

Figure 1.

Layout of study session

comparison between the placebo and experimental groups. Caffeine was delivered via Stay Alert® gum (see Kamimori et al., 2005; McLellan et al., 2003-2004; Smith, 2009; Syed et al., 2005 for studies using this gum), a caffeinated chewing gum that contains 100 mg of caffeine per piece. The gum was administered after completion of the first set of Cloze tests. The actual dose ingested by the participants was closer to 170 mg than the full 200 mg. The gum contains 100 mg per piece, but per research reported by Syed et al. (2005), within five minutes of chewing approximately 85% of the total dose is delivered. Participants were instructed to chew two pieces of this gum for five minutes and then to discard the gum in a provided wastebasket, so in practice, the caffeine dose was actually about 170 mg.

The placebo condition was defined as one in which no caffeine was administered during the session. A gum similar in size, shape, and taste was intended to be used as the placebo, but no comparable gum was found. Instead, the researcher set out to find a gum that would be unfamiliar to participants. It was determined that because a between-subjects design was being used and no participant would chew both types of gum, there was no need to make the experimental and placebo gums comparable. The main concern was that no participant be able to identify the placebo as a type of gum that they had chewed before. The gum that was chosen was Beechies® Peppermint Chewing Gum. This gum was not readily available in most stores, and a pilot sample of participants was not able to identify this gum as one they had chewed before. Participants followed the same procedure with the placebo as with the caffeine gum. Participants were told that they would be administered either caffeinated or non-caffeinated gum during the session.

2.3 Measures

A brief questionnaire with self-report items concerning sleep during the night before the session, caffeine and food consumption on the day of the session, and caffeine consumption in general was given prior to the start of testing. Self-reports of mood and fatigue were gathered before completion of each booklet of eight Cloze tests and at the conclusion of the session, for a total of five collections of self-reported mood and fatigue data. See Figure 1 for a list of all tests and self-report measures that were administered, the order in which they were administered, and the number of items in each. Candidate measures from which relevant items were selected for administration include the Profile of Mood States (POMS; McNair, Lorr, & Droppleman, 1992), the Activation-Deactivation Checklist (AD-ACL; Thayer, 1989), and the Rating Scale Mental Effort (RSME; Zijlstra, 1993). The RSME, for example, has two mental fatigue items, two physical fatigue items, and one item each on resistance to further effort, boredom, and visual fatigue (van der Linden et al., 2003). The relevant items from this scale were the two mental fatigue items and the items on boredom and resistance to further effort. The POMS contains relevant items such as “worn out”, “clear-headed”, “fatigued”, “unable to concentrate”, as well as irrelevant items such as “spiteful”, “lonely”, and “ready to fight” (McNair et al., 1992). Relevant items from each scale were combined to form 5 composite measures: fatigue, PA, NA, positive motivation, and confidence.

Testing consisted of multiple versions of a Cloze test (Taylor, 1953; see Ackerman et al., 2000 for a review of the Cloze test). The test was created primarily for assessing the readability of passages of text, but its creator concluded that it was also

useful for assessing the reader's mental ability and prior knowledge, as well as general language facility, specific knowledge, relevant vocabulary, ability to learn, attention, and motivation (Ackerman et al., 2000). The test was useful for the purposes of this study in that it possessed many of those factors discussed previously as thought to increase the likelihood of a task producing cognitive fatigue: high cognitive demands, not intrinsically interesting or enjoyable, low response to practice/learning effects, few breaks, verbal content, homogeneous content, and rate of stimulus presentation not controlled by subject. Furthermore, the Cloze test is considered to be associated with crystallized intelligence (Gc; Ackerman et al., 2000).

Thirty-two Cloze tests were administered in four sets of eight tests each; participants were given six minutes to work on each Cloze test. The six minute time limit was set to prevent the participants from finishing early and having time to rest. The order of administration of the four test sets was counterbalanced, and the individual tests were evenly distributed throughout the booklets so that each contained a roughly equivalent set of tests in terms of topic area and number of blanks. The four general topic areas for the passages are Science, History, English/Language Arts, and Human Interest. The passages were gathered from SAT prep books and college freshman-level textbooks (i.e., U.S. history, general chemistry, introduction to psychology; please see Appendix A for a list of Cloze test sources). The number of blanks in each passage ranged from 41 to 47 blanks.

Twenty-one of the 32 passages used in this study were used in the Ackerman et al. (2008) study. New passages were created to replace the 11 passages from the Ackerman et al. (2008) study that generated mean performance data which fell outside an

experimenter-determined acceptable range. For the tests that had been previously used as Cloze tests, tests with a mean percent correct below 50% or above 73% in the Ackerman et al. (2008) study were replaced for the current study. For the tests that had been previously used as Completion tests, the acceptable range was higher, given that this task is substantially easier than the Cloze test. Completion tests with a mean percent correct below 70% or above 80% in the Ackerman et al. (2008) study were replaced for the current study. Those Completion tests kept from the previous study were formatted the same but were used as Cloze tests instead of Completion tests (i.e., they were not read aloud before the participants filled them in).

The final (fifth) collection of self-report data was embedded within a larger questionnaire. A set of self-report items similar to those that had been included in the previous interim questionnaires was given; the only difference between this set and those on the previous questionnaires was that some items pertaining to specific feelings during the tests were changed to have past-tense wording. Additionally, this final questionnaire included items about the participant's strategy use, perceptions of his or her performance, typical caffeine consumption habits, and perceptions of the gum he or she chewed during the session.

CHAPTER 3

RESULTS

There are four subsections of results. The first contains information on the reliability of the self-report and performance scales, as well as checks for pre-manipulation differences between the two conditions on the variables of interest. The second section contains the bulk of the analyses, the planned analyses for each of the four hypotheses. The third section contains additional analyses related to comparisons between the two treatment groups. The fourth section contains analyses on the items contained in the final questionnaire. These items were not addressed in the hypotheses, and pertain to participants' perceptions of their performance during the session, their thoughts on the gum they chewed, and their plans for immediately after the session.

3.1 Analysis of scales and checks for pre-manipulation differences

The three self-report scales have internal consistency reliability above $\alpha = .80$: $\alpha = .81$ for the fatigue scale, $\alpha = .87$ for the PA scale, and $\alpha = .87$ for the NA scale. For the Cloze tests, alpha coefficients were calculated separately for each hour of testing, and also for the full set of 32 Cloze tests as a whole: $\alpha = .89$ for Hour 1, $\alpha = .92$ for Hour 2, $\alpha = .91$ for Hour 3, $\alpha = .92$ for Hour 4, and $\alpha = .97$ for the full set of Cloze tests. These reliability estimates were conducted using data from both the caffeine and placebo groups.

As a random assignment check, t-tests were used to compare mean levels of performance during the first hour and mean scores on the self-report measures at baseline in order to determine if there were any significant differences between the caffeine and

placebo groups on any of the dependent variables prior to the experimental manipulation. These analyses included all 116 participants for the performance measures and 114 participants for the self-report measures (two participants, one from each treatment group, were excluded from all self-report analyses due to substantial missing data). There were no significant differences between the two groups on the baseline (i.e., performance during the first hour and self-report measures taken prior to the start of testing) measures of Cloze performance, $t(114) = .72$, $p > .05$, $d = .13$, self-reported fatigue, $t(112) = -1.29$, $p > .05$, $d = .24$, self-reported PA, $t(112) = .50$, $p > .05$, $d = .09$, or self-reported NA, $t(112) = -1.18$, $p > .05$, $d = .22$ (see Table 2). These results imply that on all variables of interest, there were no significant differences between the experimental and placebo groups prior to the experimental manipulation.

3.2 Hypotheses 1 – 4, planned analyses

3.2.1 Hypothesis 1

Hypothesis 1 stated that while Cloze test performance at the aggregate level would remain constant in the placebo group with increasing time-on-task, individual differences in performance would be evident based on the strategies participants employed during the testing session. To test this, two general analyses were used. To assess the overall performance trend, I used a repeated-measures ANOVA with hour as the independent variable and hourly performance as the dependent variable. This analysis was conducted on only the 58 placebo group participants in order to investigate the aggregate level performance effects in the absence of caffeine effects. Please see Figure 2 for a bar graph of hourly mean Cloze test performance levels and 95% confidence intervals, and Figure 3 for a line graph displaying hourly mean Cloze test performance

Table 2.

Means and t-tests for differences in means for self-reported positive affect (PA), negative affect (NA), and fatigue (Fat) and Cloze test performance (Perf) for caffeine and placebo groups.

	Caffeine	Placebo	ind. Samples t-test t value	significance of t-value	*d*
BasePA	1.92	1.86	0.50	0.62	0.09
PostHr1PA	2.03	1.85	1.27	0.21	0.24
PostHr2PA	1.96	1.54	3.31	0.00**	0.63
PostHr3PA	1.66	1.39	2.50	0.01**	0.47
PostHr4PA	1.74	1.59	1.07	0.29	0.20
BaseNA	1.42	1.52	-1.18	0.24	-0.22
PostHr1NA	1.42	1.51	-1.04	0.30	-0.20
PostHr2NA	1.51	1.65	-1.36	0.18	-0.26
PostHr3NA	1.55	1.84	-1.98	0.05*	-0.37
PostHr4NA	1.61	1.78	-1.21	0.23	-0.23
BaseFat	2.90	3.10	-1.29	0.20	-0.24
PostHr1Fat	2.71	2.96	-1.54	0.13	-0.29
PostHr2Fat	2.71	3.35	-3.74	0.00**	-0.71
PostHr3Fat	3.05	3.67	-3.39	0.00**	-0.64
PostHr4Fat	3.18	3.59	-2.39	0.02*	-0.45
AvgHr1Perf	63.42	61.68	0.72	0.48	0.13
AvgHr2Perf	67.54	62.79	1.91	0.06	0.36
AvgHr3Perf	67.91	62.02	2.39	0.02*	0.45
AvgHr4Perf	67.55	62.49	1.91	0.06	0.36

* = sig. at alpha = .05 level; ** = sig. at alpha = .01 level

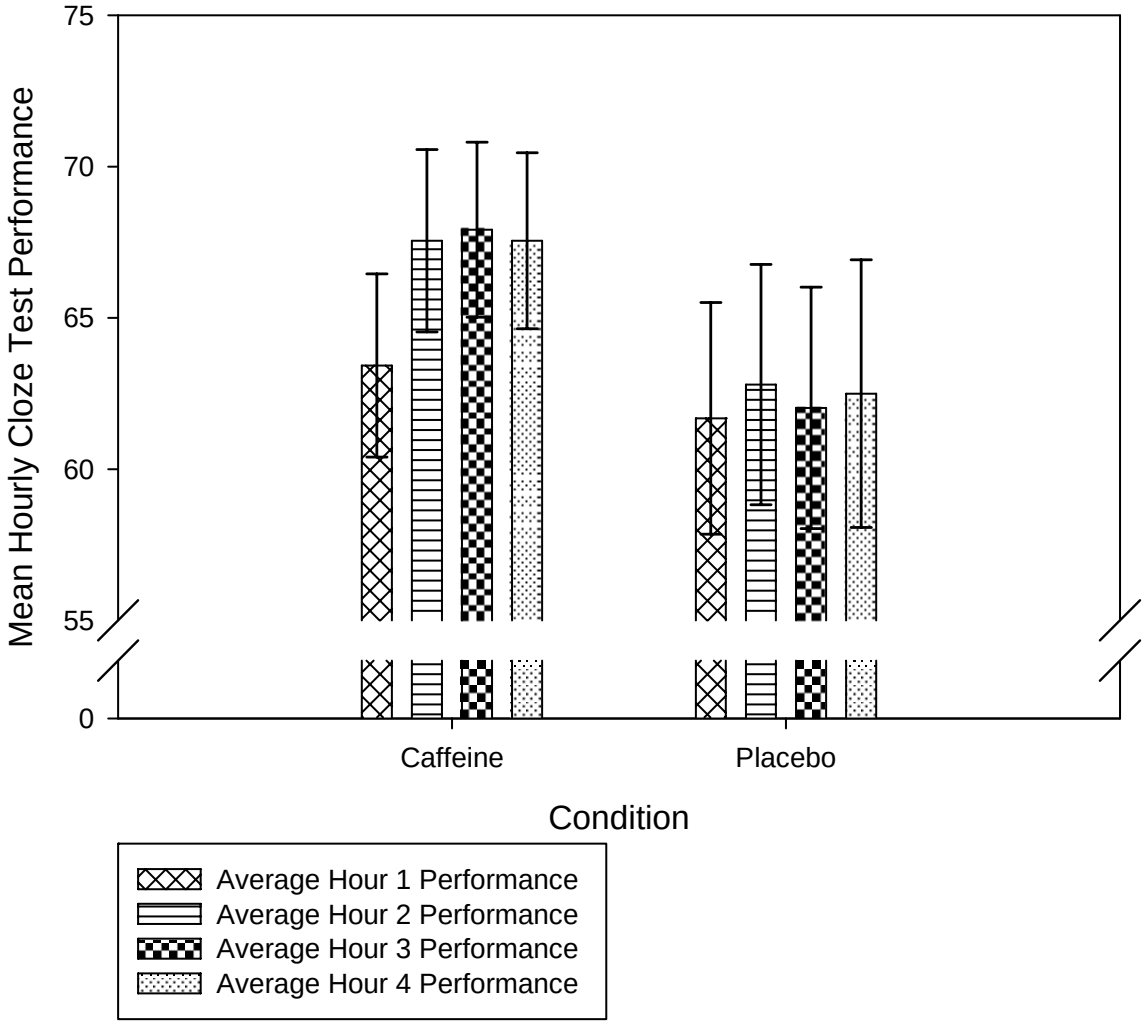

Figure 2.

Hourly mean performance for caffeine and placebo groups with 95% confidence interval bars for the means

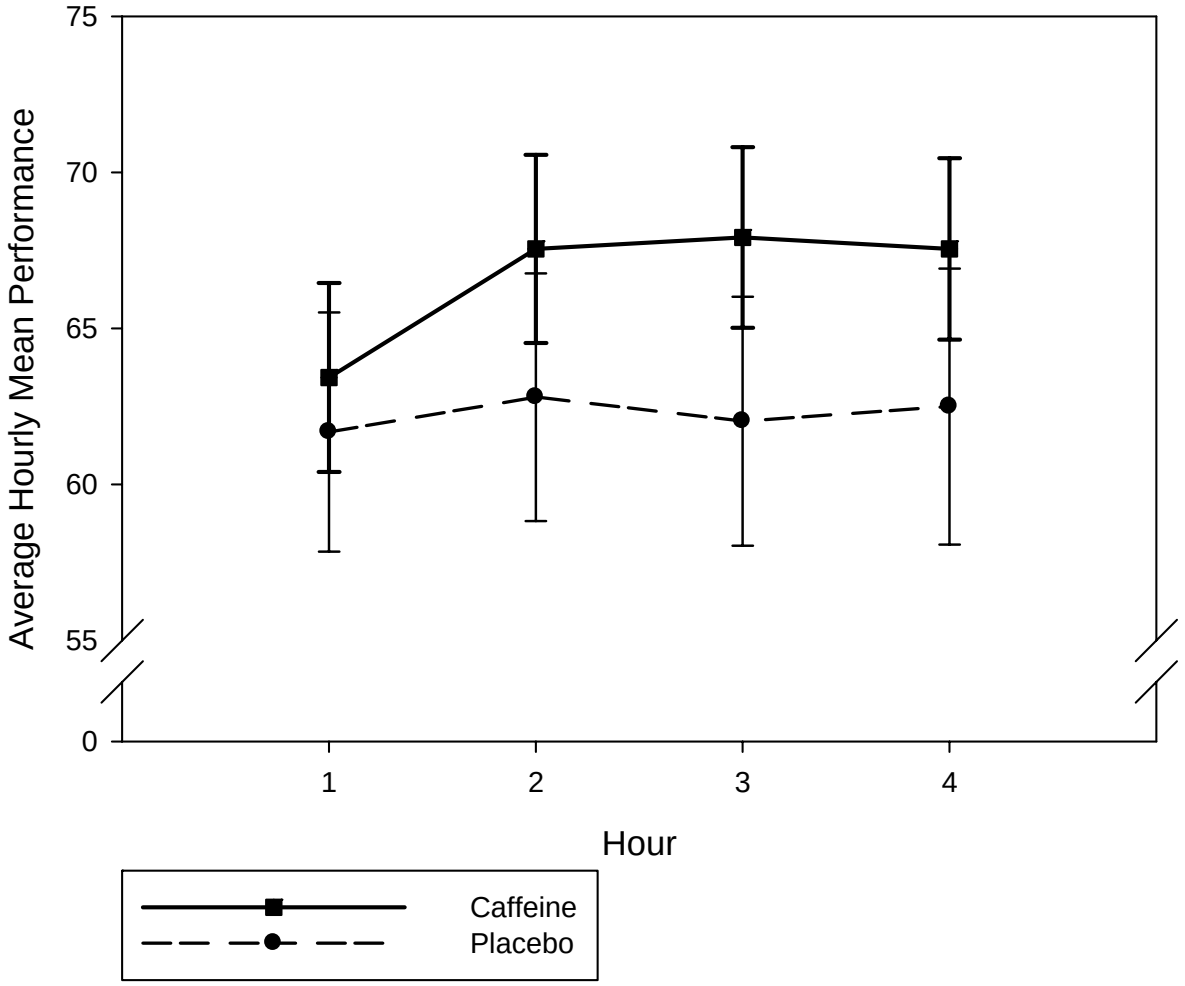

Figure 3.

Mean hourly Cloze test performance (Perf) for caffeine and placebo conditions with 95% confidence interval bars for the means

levels (please note that in order to minimize the number of figures, most figures contain information on both the caffeine and placebo groups). The results for this analysis suggested no significant differences between hourly performance levels, indicating constant aggregate level performance with increasing time-on-task among the placebo group participants, F (2.02, 115.38) = .53, $p > .05$, d = .26, *power* = .99. Information on the full analysis is contained in Table 3 (please note that in order to minimize the number of tables, most tables contain information on both the caffeine and placebo groups).

Because the expected value of the F ratio when the null hypothesis is true is approximately 1.0, an F ratio of < 1.0 is indicative of some bias in the estimation of mean squares between and/or within groups and necessitates some modifications in terms of computing and interpreting effect size estimates (Voelkle, Ackerman & Wittmann, 2007). In their (2007) methodological paper, Voelkle et al. caution against interpreting regular effect size estimates such as d or f in cases where the F ratio is < 1.0. Because the F value for the repeated-measures ANOVA on the Cloze test performance of the placebo group participants is < 1.0, alternate effect size estimates were calculated, per the recommendations and formulas set forth by Voelkle et al. (2007). ε^2 and ω^2 were calculated for this ANOVA; these values are intended to correct the biased sum of squares estimates that occur with other effect size values, such as d or f. These adjusted effect size estimates are more appropriate under the condition of an F ratio < 1.0, and are intended to prevent erroneous conclusions being drawn from large effect size estimates, which do not make sense in the context of an F ratio < 1.0. For this ANOVA, $\varepsilon^2 = -.0007$ and $\omega^2 = -.0007$. These values are very close to zero, and Voelkle et al. (2007) emphasize that for an F ratio < 1.0, the most accurate population effect size estimate is always zero.

Table 3.

Repeated-measures ANOVA results for self-report and performance measures for caffeine and placebo groups

Caffeine group participants (n = 57 for self-rpt measures, 58 for performance measures)																			
	# items	**alpha**	**base-line mean**	**base-line SD**	**post-hr 1 mean**	**post-hr 1 SD**	**post-hr 2 mean**	**post-hr 2 SD**	**post-hr 3 mean**	**post-hr 3 SD**	**F**	**MSE**	**ES**	***f***	***d* (for Anova)**	**Post-test mean**	**Post-test SD**	**t**	***d* (for t-test)**
Fatigue	12	0.81	2.90	0.87	2.71	0.87	2.71	0.91	3.05	0.94	6.02**	0.30	0.10	0.33	0.89	3.18	0.85	-1.95	0.52
PA	7	0.87	1.92	0.67	2.03	0.80	1.96	0.81	1.66	0.68	7.26**	0.26	0.12	0.36	1.02	1.74	0.80	n/a	n/a
NA	12	0.87	1.42	0.46	1.42	0.43	1.51	0.50	1.55	0.61	4.24**	0.13	0.07	0.27	0.76	1.61	0.56	n/a	n/a
Cloze Perf	32	0.97	63.42	11.52	67.54	11.46	67.91	11.01	67.55	11.06	14.78**	21.55	0.21	0.51	1.44	n/a	n/a	n/a	n/a

Placebo group participants (n = 57 for self-rpt measures, 58 for performance measures)																			
	# items	**alpha**	**base-line mean**	**base-line SD**	**post-hr 1 mean**	**post-hr 1 SD**	**post-hr 2 mean**	**post-hr 2 SD**	**post-hr 3 mean**	**post-hr 3 SD**	**F**	**MSE**	**ES**	***f***	***d* (for Anova)**	**Post-test mean**	**Post-test SD**	**t**	***d* (for t-test)**
Fatigue	12	0.81	3.10	0.75	2.96	0.82	3.35	0.92	3.67	1.00	19.29**	0.33	0.26	0.59	1.58	3.60	0.96	1.33	0.36
PA	7	0.87	1.86	0.67	1.86	0.66	1.54	0.51	1.39	0.46	15.27**	0.22	0.21	0.52	1.47	1.59	0.67	n/a	n/a
NA	12	0.87	1.52	0.45	1.51	0.58	1.65	0.65	1.84	0.91	7.68**	0.35	0.12	0.37	1.05	1.78	0.91	n/a	n/a
Cloze Perf	32	0.97	61.68	14.56	62.79	15.10	62.02	15.17	62.49	16.82	0.53	39.09	0.01	0.10	0.27	n/a	n/a	n/a	n/a

Notes: ES = partial eta squared values; t = t-test between 4th and 3rd hour; * = sig. at alpha = .05 level; ** = sig. at alpha = .01 level;

PA = positive affect; NA = negative affect

So while the traditional effect size estimates for this ANOVA suggest a small effect size difference between hourly mean performance levels for the placebo group participants (i.e., $d = .26$ and $f = .10$), it is more accurate to rely on the adjusted measures of association strength (i.e., $\varepsilon^2 = -.0007$ and $\omega^2 = -.0007$) which suggest that the effect size estimate is approximately zero.

To investigate individuals' use of the four hypothesized strategies, I calculated drops and spurts (this method is modeled after that employed by Ackerman and colleagues in their 2008 study of cognitive fatigue during testing). A drop is defined as a decrease of more than 20 percentage points between the 3rd and the 7th Cloze tests within a given hour. A spurt is defined as an increase of more than 20 percentage points between the 3rd and the 7th Cloze tests within a given hour. Examination of each participant's pattern of drops and spurts allowed me to infer the strategy that they implemented during the course of the testing.

The way in which drops and spurts were used to determine the strategy implemented by a given individual is shown in Table 4. Individuals who exhibited neither a drop nor a spurt were classified into the "normal" group. Those individuals who showed one or more spurts with no drop were classified into the "overactivity" group, while individuals who showed the opposite pattern (i.e., one or more drops with no spurt) were classified into the "underactivity" group. Lastly, those individuals who showed both one or more drops and one or more spurts were classified into the "mixed" group (combined overactivity/withdrawal).

Using performance data from all four hours of testing and following the classification scheme previously described, the 58 placebo group participants were

Table 4.

Classification scheme for strategy groups based on drops and spurts

	One or more spurts	**No spurts**
One or more drops	Mixed group	Underactivity group
No drops	Overactivity group	Normal group

classified into the four groups in the following manner: 25 participants in the normal group (43.1%), 16 participants in the overactivity group (27.6%), 13 participants in the underactivity group (22.4%), and 4 participants in the mixed group (6.9%). This information is summarized in Figure 4.

3.2.2 Hypothesis 2

Hypothesis 2 stated that in the absence of caffeine, self-reports of fatigue and NA would increase, and self-reports of PA would decrease, over time-on-task. To test this, I used a repeated-measures ANOVA with hour as the independent variable and self-report scores as the dependent variable. For the fatigue scale, in which the wording of some items differed slightly between the post-test questionnaire and all prior questionnaires, a dependent t-test was used to compare the self-report scores from the final interim questionnaire with those in the post-test questionnaire. Hypothesis 2 was tested using data from the 57 placebo group participants only, as the caffeine group provided self-report data at only one time point prior to caffeine administration. In these and all other tests using the self-report data, two participants (one from the placebo group and one from the experimental group) were excluded due to substantial missing data.

For self-reports of fatigue, a significant effect of hour was found for the 57 placebo group participants, $F(2.59, 144.97) = 19.29$, $p < .01$, $d = 1.58$. A paired samples t-test conducted on mean fatigue levels from the post-Hour 3 and post-test measurements suggested no significant difference, $t(56) = 1.33$, $p > .05$, $d = .36$. The changes from one measurement to the next were mostly in the predicted direction: mean self-reported fatigue levels decreased from baseline to post-Hour 1 ($d = .37$), increased from post-Hour 1 to post-Hour 2 ($d = 1.29$), increased from post-Hour 2 to post-Hour 3 ($d = 1.08$), and

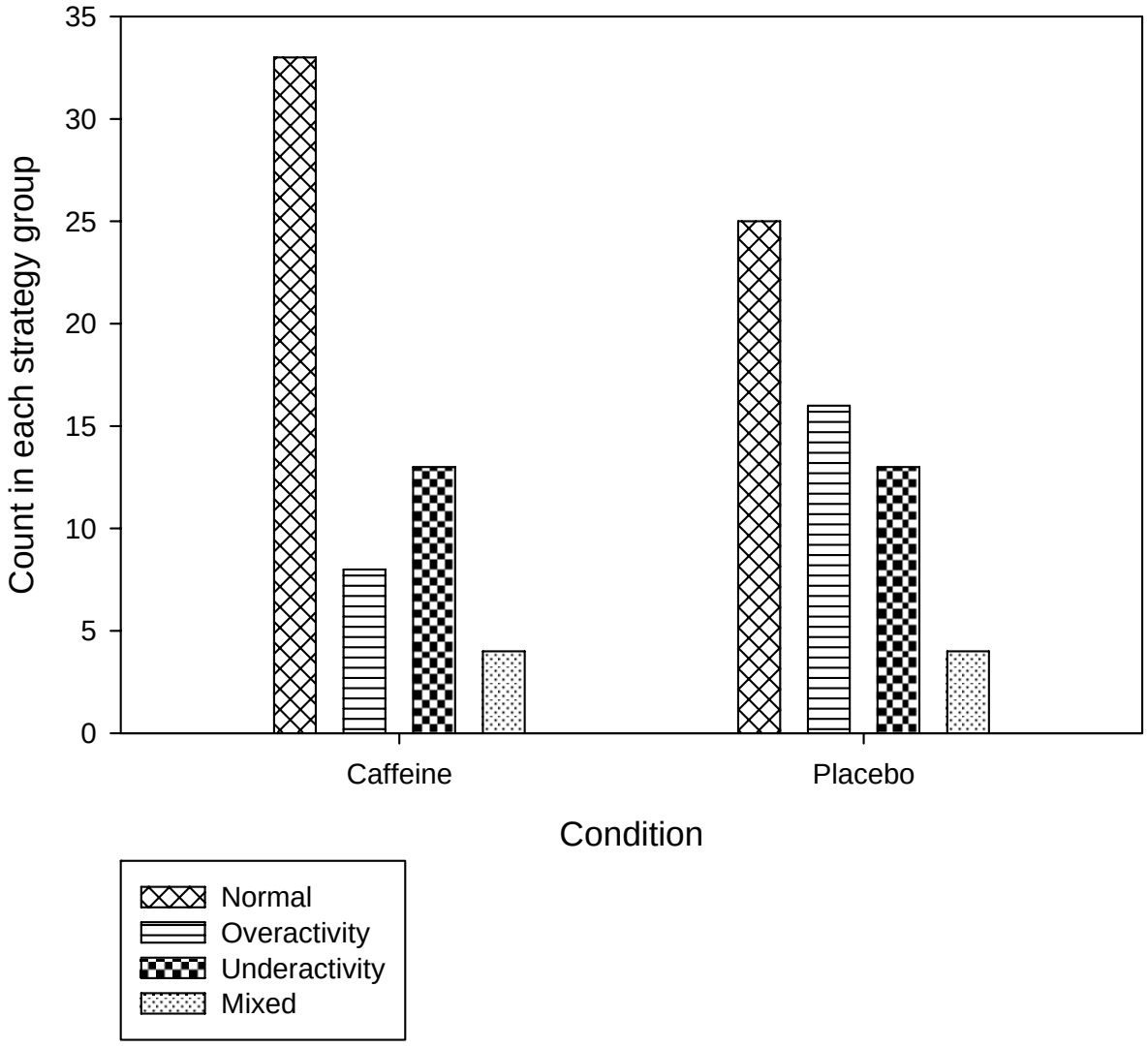

Figure 4.

Counts of participants in each strategy group for caffeine and placebo conditions

decreased from post-Hour 3 to post-test (d = .36). So the largest effects over the course of the session with respect to changes in fatigue levels felt by the placebo group participants occurred during the middle of the session. Information on the full analysis is provided in Table 3. Mean self-reported fatigue levels at each measurement time are displayed in Figure 5. Effect size estimates are displayed in Table 5.

For self-reports of PA, a significant effect of hour was found for the 57 placebo group participants, $F(2.91, 163.05) = 15.27$, $p < .01$, $d = 1.47$. This ANOVA included data from all five measurement points. Some of the changes from one measurement to the next were in the predicted direction: mean self-reported PA levels decreased minimally from baseline to post-Hour 1 (d = .01), decreased from post-Hour 1 to post-Hour 2 (d = 1.44), decreased from post-Hour 2 to post-Hour 3 (d = .77), and increased from post-Hour 3 to post-test (d = .80). The decreases in PA during the 2nd and 3rd hours of testing (with large and medium effect sizes, respectively) correspond to the hypothesized decrease in PA with increasing time-on-task for placebo group participants. The increase in PA (with a large effect size) during the last hour of testing is in the opposite direction from what was hypothesized, but replicates a finding from Ackerman et al. (2008). The suggested reason for this increase is that participants experienced positive feelings due to the end of the session drawing near. Information on the full analysis is provided in Table 3. Mean self-reported PA levels at each measurement time are displayed in Figure 6. The effect size estimates are reported in Table 5.

For self-reports of NA, a significant effect of hour was found for the 57 placebo group participants, $F(1.85, 103.79) = 7.68$, $p < .01$, $d = 1.05$. This ANOVA included data from all five measurement points. The changes in NA from one hour to the next largely

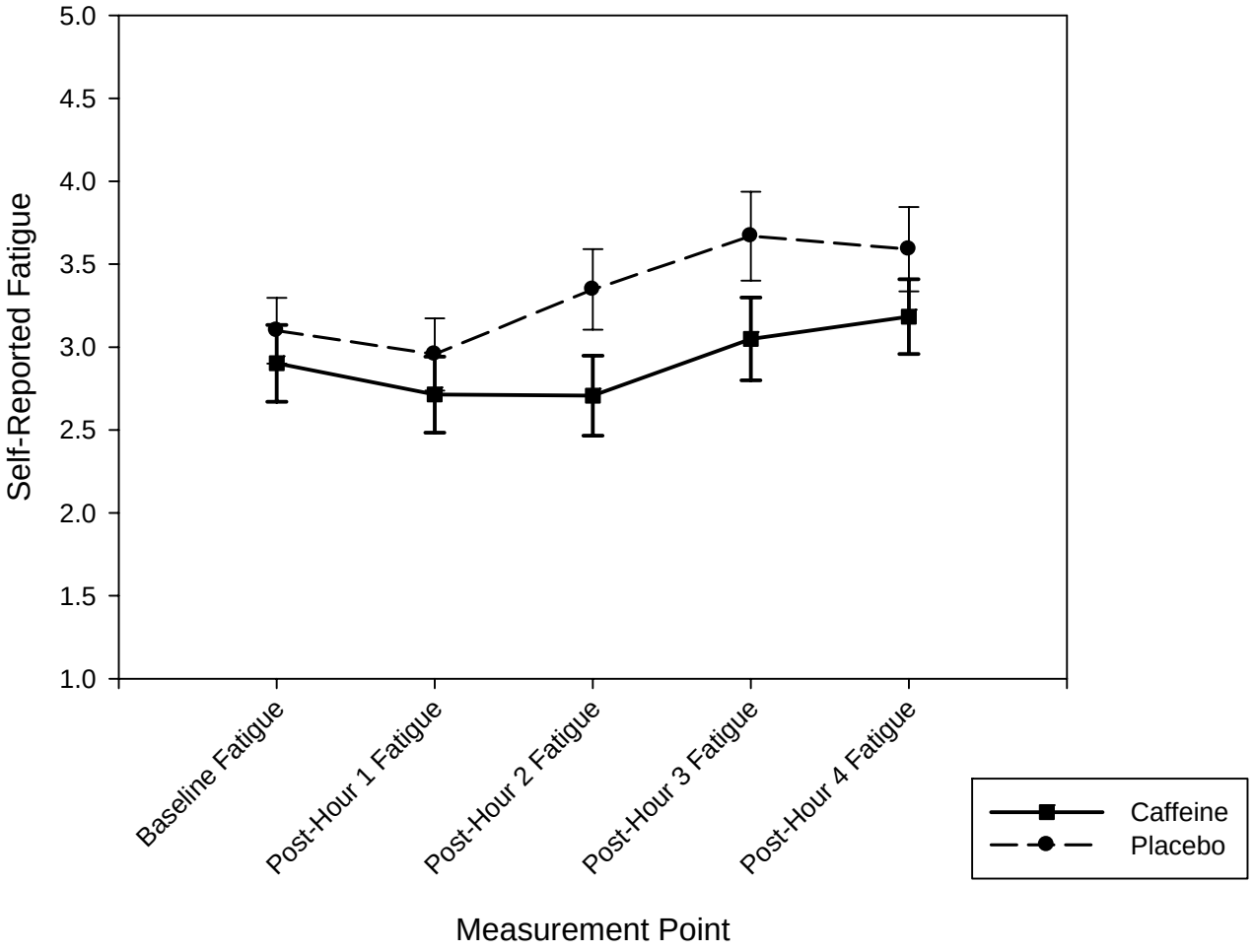

Figure 5.

Mean self-reported fatigue scores for caffeine and placebo conditions at five measurement points with 95% confidence interval bars for the means

Table 5.

Effect size estimates for pairwise hourly comparisons on performance, fatigue (Fat), positive affect (PA), and negative affect (NA) for the caffeine and placebo groups

Caffeine Group			**Placebo Group**	
Comparison	**Cohen's *d***		**Comparison**	**Cohen's *d***
Hour 1 vs Hour 2 Performance	1.33		Hour 1 vs Hour 2 Performance	-0.41
Hour 1 vs Hour 3 Performance	1.32		Hour 1 vs Hour 3 Performance	-0.12
Hour 1 vs Hour 4 Performance	1.13		Hour 1 vs Hour 4 Performance	-0.18
Hour 2 vs Hour 3 Performance	0.17		Hour 2 vs Hour 3 Performance	0.27
Hour 2 vs Hour 4 Performance	0.00		Hour 2 vs Hour 4 Performance	0.06
Hour 3 vs Hour 4 Performance	0.16		Hour 3 vs Hour 4 Performance	-0.14
BaseFat vs. PostHr1Fat	0.61		BaseFat vs. PostHr1Fat	0.37
BaseFat vs. PostHr2Fat	0.51		BaseFat vs. PostHr2Fat	-0.66
BaseFat vs. PostHr3Fat	-0.33		BaseFat vs. PostHr3Fat	-1.22
PostHr1Fat vs. PostHr2Fat	0.02		PostHr1Fat vs. PostHr2Fat	-1.29
PostHr1Fat vs. PostHr3Fat	-0.89		PostHr1Fat vs. PostHr3Fat	-1.77
PostHr2Fat vs. PostHr3Fat	-1.12		PostHr2Fat vs. PostHr3Fat	-1.08
PostHr3Fat vs. Post-testFat	0.52		PostHr3Fat vs. Post-testFat	-0.36
BasePA vs. PostHr1PA	-0.42		BasePA vs. PostHr1PA	0.01
BasePA vs. PostHr2PA	-0.13		BasePA vs. PostHr2PA	1.17
BasePA vs. PostHr3PA	1.02		BasePA vs. PostHr3PA	1.60
PostHr1PA vs. PostHr2PA	0.27		PostHr1PA vs. PostHr2PA	1.44
PostHr1PA vs. PostHr3PA	1.22		PostHr1PA vs. PostHr3PA	1.74
PostHr2PA vs. PostHr3PA	0.95		PostHr2PA vs. PostHr3PA	0.77
PostHr3PA vs. Post-testPA	-0.30		PostHr3PA vs. Post-testPA	-0.80
BaseNA vs. PostHr1NA	0.03		BaseNA vs. PostHr1NA	0.03
BaseNA vs. PostHr2NA	-0.34		BaseNA vs. PostHr2NA	-0.55
BaseNA vs. PostHr3NA	-0.49		BaseNA vs. PostHr3NA	-0.82
PostHr1NA vs. PostHr2NA	-0.58		PostHr1NA vs. PostHr2NA	-0.80
PostHr1NA vs. PostHr3NA	-0.71		PostHr1NA vs. PostHr3NA	-1.04
PostHr2NA vs. PostHr3NA	-0.22		PostHr2NA vs. PostHr3NA	-0.75
PostHr3NA vs. Post-testNA	-0.34		PostHr3NA vs. Post-testNA	0.43

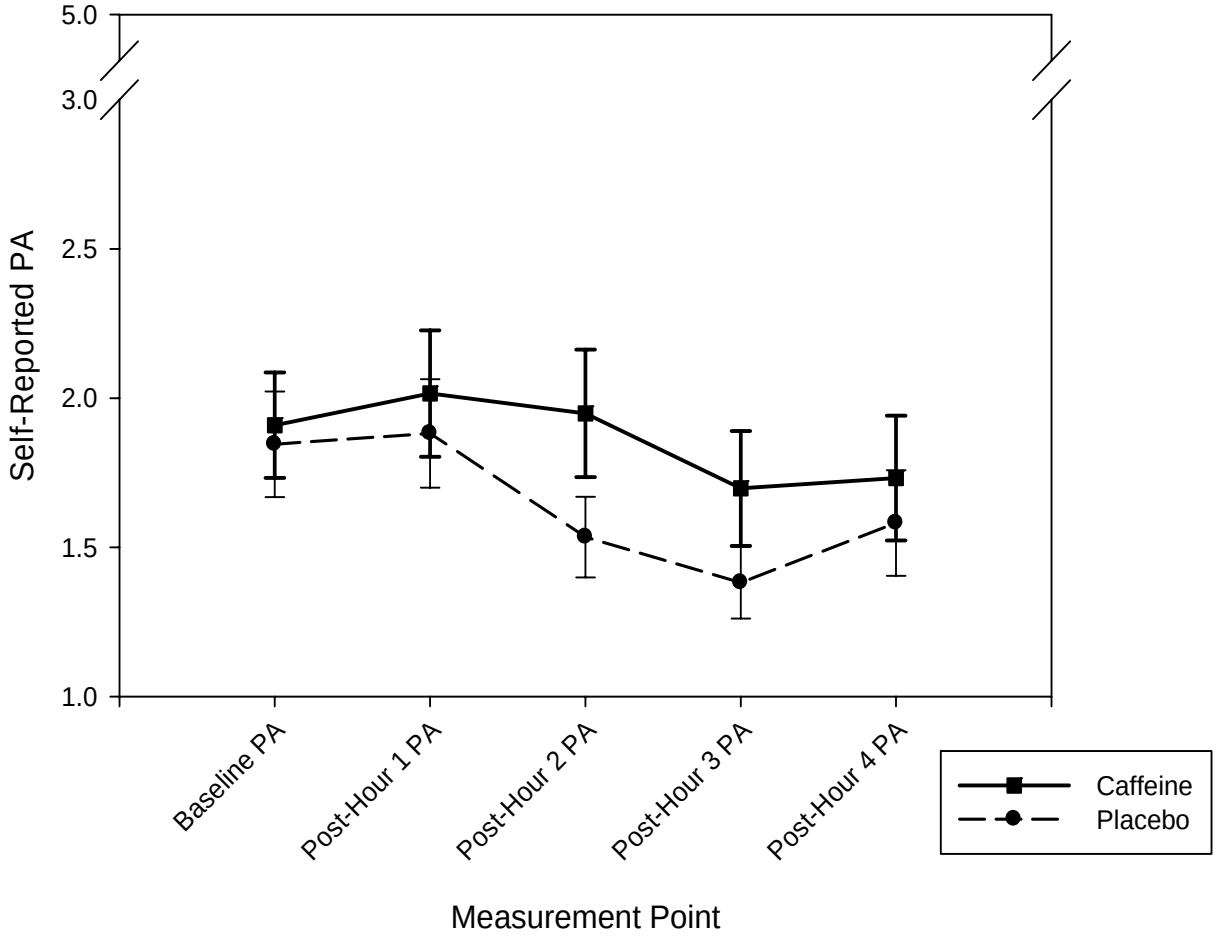

Figure 6.

Mean self-reported positive affect (PA) scores for caffeine and placebo conditions at five measurement points with 95% confidence interval bars for the means

followed the hypothesized changes: self-reported NA decreased slightly from baseline to Post-Hour 1 (d = .03), increased from post-Hour 1 to post-Hour 2 (d = .80), increased from post-Hour 2 to post-Hour 3 (d = .75), and decreased from post-Hour 3 to post-test (d = .43). The increases in NA during the 2nd and 3rd hours of testing (with large and medium effect sized, respectively) support the hypothesized changes in NA. The NA decrease (with medium effect size) in the final hour of testing mirrors the increase in PA found during the same time period, and is likely also a product of positive feelings associated with the end of the session. Mean self-reported NA levels at each measurement time are displayed in Figure 7. The full analyses for these results are included in Table 3. The effect size estimates are reported in Table 5.

3.2.3 Hypothesis 3

Hypothesis 3 stated that the individuals in the caffeine group would exhibit fewer performance decrements (i.e., drops), fewer performance improvements (i.e., spurts), and higher overall performance at the aggregate level, relative to the placebo condition, in Hours 2, 3, and 4. To test the first part of this hypothesis, I used Chi-square tests to compare the number of drops and spurts within each hour of testing to determine if there were significantly fewer spurts and significantly fewer drops per hour (for the last three hours only, since caffeine was administered to the caffeine group after the first hour) for the caffeine group as compared to the placebo group. As this analysis involved comparison of drop and spurt counts for the two conditions, all 116 participants (58 placebo group participants and 58 caffeine group participants) were included. Results suggested that there were no significant differences in frequency of drops or spurts between the caffeine and placebo groups within any of the hours. Please see Table 6 for

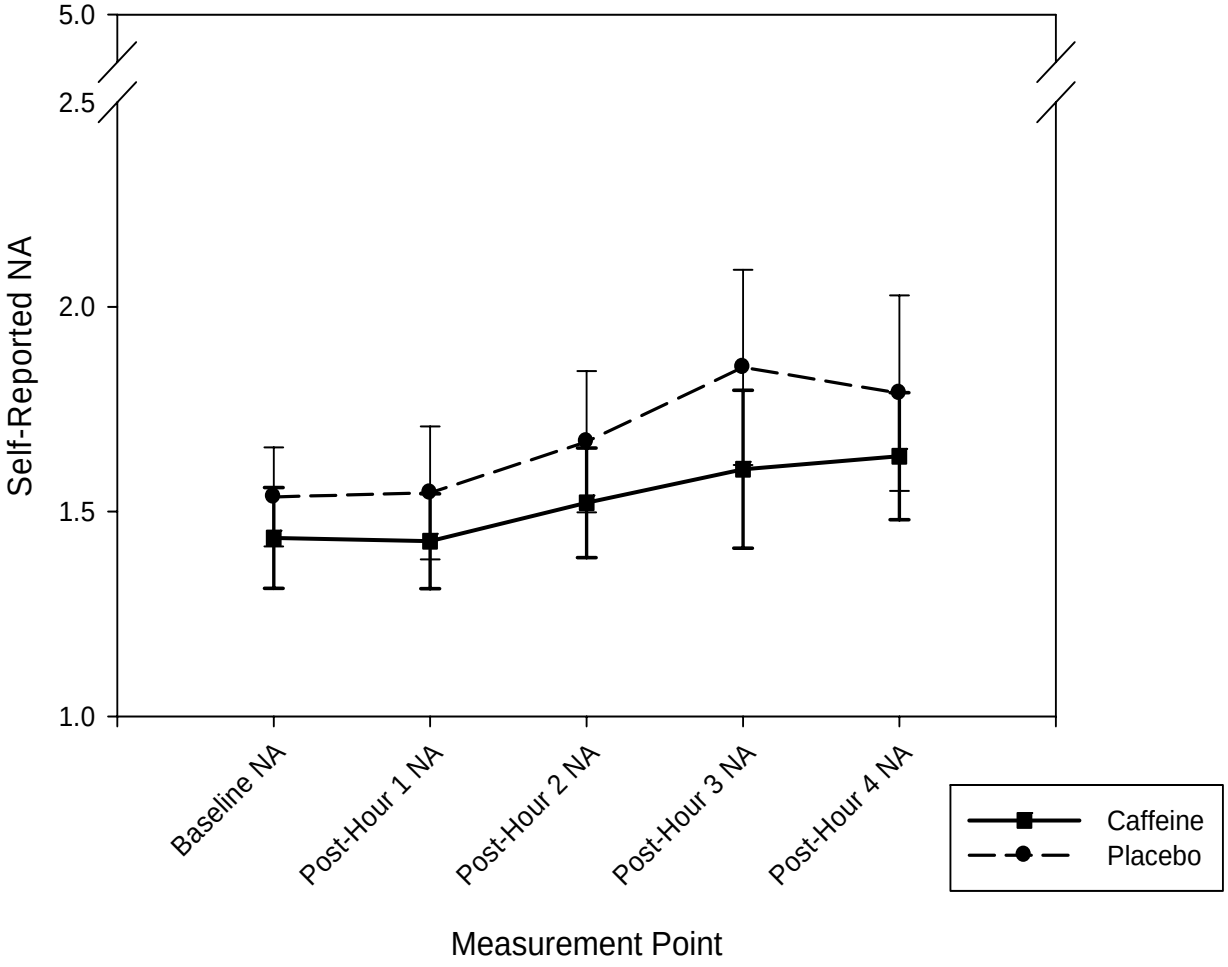

Figure 7.

Mean self-reported negative affect (NA) scores for caffeine and placebo conditions at five measurement points with 95% confidence interval bars for the means

Table 6.

Frequencies of drops and spurts within hours 2, 3 and 4 for the caffeine and placebo groups, with chi-square tests for differences between the two groups

	Caffeine	Placebo	Chi-square value	Significance of Chi-square
# drops in Hr2	4	4	0.00	1.00
# spurts in Hr2	3	3	0.00	1.00
# drops in Hr 3	4	6	0.44	0.51
# spurts in Hr3	4	4	0.00	1.00
# drops in Hr4	8	3	2.51	0.11
# spurts in Hr4	5	6	0.10	0.75

the full analysis, and Figure 4 for the strategy group distribution of the caffeine participants.

I also performed independent samples t-tests to compare the mean performance level within each hour (again, for the last three hours only) for the caffeine and placebo groups. Again, because this analysis involved comparing the two treatment groups, all 116 participants were included. For average performance during hour three, the caffeine group had a significantly higher mean performance level as compared to the placebo group, $t(114) = 2.39$, $p < .05$, $d = .45$. For average performance during hours two and four, the caffeine group did not perform significantly higher than the placebo group, $t(114) = 1.91$, $p > .05$, $d = .36$ and $t(114) = 1.91$, $p > .05$, $d = .36$ for hours two and four, respectively. The Hour 3 performance data shows a significant performance effect of caffeine, as well as a Cohen's *d* value that is close to the cutoff for a medium effect (i.e., $d = .50$). Hours 2 and 4 did not show significant performance effects, but the corresponding effect sizes are over the small threshold ($d = .20$), midway between a small and a medium effect. So with the significant effect and medium effect size in Hour 3, along with the small effect sizes for higher performance among caffeine group participants as compared to placebo group participants in Hours 2 and 4, this data provides at least moderate support for the notion that caffeine improves performance on a complex cognitive task. Please see Table 2 for the full analysis.

3.2.4 Hypothesis 4

Hypothesis 4 stated that the caffeine group would exhibit lower levels of self-reported fatigue and NA, and higher levels of self-reported PA, relative to the placebo group, in Hours 2, 3, and 4. To test this, I performed independent samples t-tests between

the mean self-reported values of fatigue, PA, and NA for the two groups for all questionnaires given after caffeine administration (please see Table 2 for analyses). These analyses included all 114 participants, 57 from the caffeine group and 57 from the placebo. For self-reported fatigue, the placebo group reported significantly higher mean levels of fatigue as compared to the caffeine group in the measurements taken after Hour 2, $t(112) = -3.74$, $p < .01$, $d = .71$, after Hour 3, $t(112) = -3.39$, $p < .01$, $d = .64$ and after Hour 4, $t(112) = -2.39$, $p < .05$, $d = .45$ There were no significant differences between self-reported fatigue levels in the two groups at baseline, $t(112) = -1.29$, $p > .05$, $d = .24$, or at the post-Hour 1 measurement, $t(112) = -1.54$, $p > .05$, $d = .29$. So after caffeine administration, the participants who had consumed caffeine reported lower mean levels of fatigue than those participants who had not consumed caffeine, and this difference was not present prior to caffeine administration.

With regard to PA, the caffeine group reported significantly higher mean PA levels as compared to the placebo group in the measurements taken after Hour 2, $t(112) = 3.31$, $p < .01$, $d = .63$, and after Hour 3, $t(112) = 2.50$, $p < .05$, $d = .47$. There were no significant differences between self-reported PA levels in the two groups at baseline, $t(112) = .50$, $p > .05$, $d = .09$ and at the measurement taken after the first hour of testing, $t(112) = 1.27$, $p > .05$, $d = .24$. Again, the two groups differed on PA levels after caffeine administration, with the caffeine group reporting significantly higher PA levels, and this difference was not present prior to caffeine administration. Additionally, the significant difference in PA levels that was seen in the measurements taken after Hours 2 and 3 was eliminated by the final measurement, as the groups did not report significantly different PA levels at this last measurement, $t(112) = 1.07$, $p > .05$, $d = .20$. Perhaps this is due to

an overriding effect of improved attitude related to the end of the session drawing near, which should have been experienced equally by the two groups. The effects for NA were less apparent. There were no significant differences in mean self-reported NA levels for the two groups except in the post-Hour 3 measurement, $t(112) = -2.00$, $p = .05$, $d = .37$ with significantly lower levels reported by the caffeine group as compared to the placebo group.

3.3 Additional analyses

The repeated-measures ANOVAs conducted with the placebo group participants to investigate the manner in which their Cloze test performance, self-reported fatigue levels, self-reported PA levels, and self-reported NA levels changed over time were repeated with the caffeine group participants to see if the changes over time on these variables of interest differed between the two groups. Effect size estimates were also calculated for the hourly changes in performance and the three self-report variables among the caffeine and placebo group participants in order to compare them with the effect size estimates for hourly changes in performance and the three self-report variables among the placebo group participants. The differences between the two conditions at each measurement point were assessed with the independent-samples t-tests, but it also seemed useful to investigate whether the time courses of change in these variables differed between the two conditions. Comparing the repeated-measures ANOVAs shows whether significant differences between hours occurred on each variable of interest for the two groups, and looking at the effect size estimates for hourly changes allows for comparisons of the time intervals in which large changes in the variables of interest occurred for the two groups.

To additionally investigate the overall performance of the caffeine group, I performed a repeated-measures ANOVA with hour as the independent variable and hourly performance as the dependent variable, using data from only the 58 participants in the caffeine group. This is a parallel test to that which was run with the placebo group participants in order to investigate their overall performance. This analysis indicated a significant effect of hour, $F(2.46, 141.09) = 14.78$, $p < .01$, $d = 1.37$ (see Table 3 for the full analysis). Effect sizes were calculated to assess the magnitude of the difference between the mean performance score in each possible pair of hours. These same effect size calculations were also done for the pairs of hourly performance data for the placebo group participants in order to allow for comparisons between the effect size estimates for the two groups. These effect size estimates can be found in Table 5. For the caffeine group participants, the effect size estimates for Hour 1 vs. Hour 2, Hour 1 vs. Hour 3, and Hour 1 vs. Hour 4 were all large by Cohen's standards ($d = 1.33$, 1.32, and 1.13, respectively). Effect size estimates for all other hourly comparisons among caffeine group participants failed to reach the small threshold ($d = .2$). For the placebo group, no effect size achieved the medium threshold ($d = .5$). Two hourly differences passed the small cutoff; the increase from Hour 1 to Hour 2 had an effect size of $d = .41$, and the decrease from Hour 2 to Hour 3 had an effect size of $d = .27$. Taken together, these effect size estimates suggest that for the caffeine group, there was a large difference between Hour 1 performance and performance during all subsequent hours. For the placebo group, there were small differences between Hour 1 and Hour 2 and between Hour 2 and Hour 3. Because the baseline comparison indicates that the two groups did not differ prior to caffeine administration, these differences in effect size patterns can be attributed to

performance effects of caffeine. Namely, the caffeine group experienced a large performance increase from Hour 1 to Hour 2 that did not subside during the session, while the placebo group experienced smaller performance changes during the session.

The repeated-measures ANOVA on fatigue in the 57 caffeine group participants revealed a significant effect of hour, $F(2.54, 141.97) = 6.02$, $p < .01$, $d = .89$. The paired-samples t-test on the post-Hour 3 and post-test measurements suggested no significant difference between self-reported fatigue levels from the post-Hour 3 and the post-test measurement, $t(56) = -1.95$, $p > .05$, $d = .52$ (please see Table 3 for the full analyses). Effect size estimates for the differences from one hour to the next reveal an interesting finding: caffeine appears to have delayed a large increase in feelings of fatigue by one hour. The mean self-reported fatigue levels for the caffeine group decreased from baseline to post-Hour 1 ($d = .61$), remained stable from post-Hour 1 to post-Hour 2 ($d = .02$), then increased from post-Hour 2 to post-Hour 3 ($d = 1.12$) and increased again from post-Hour 3 to post-test ($d = .52$). In the placebo group, there were increases from post-Hour 1 to post-Hour 2 and from post-Hour 2 to post-Hour 3, both of which had large effect size. In the caffeine group, an increase with a large effect size did not occur until the interval between post-Hour 2 and post-Hour 3. So the placebo group participants experienced a large effect sized increase in fatigue during the 2nd hour of testing; caffeine administration appears to have delayed this large increase in felt fatigue until the 3rd hour of testing for the placebo group participants. Effect size estimates for all hourly comparisons are reported in Table 5.

The repeated-measures ANOVA on PA in the 57 caffeine group participants revealed a significant effect of hour, $F(2.95, 165.09) = 7.26$, $p < .01$, $d = 1.02$ (please see

Table 3 for the full analyses). Effect size estimates for the differences in mean self-reported PA from one hour to the next suggest that, similar to the finding with fatigue, caffeine may have delayed a large drop in PA by one hour. Among caffeine group participants, mean PA levels increased from baseline to post-Hour 1 (d = .42), decreased from post-Hour 1 to post-Hour 2 (d = .27), decreased from post-Hour 2 to post-Hour 3 (d = .95), and then increased from post-Hour 3 to post-test (d = .30). The placebo group experienced a decrease in PA from the post-Hour 1 to the post-Hour 2 measurement with a large effect size of d = 1.44. The caffeine group also experienced a decrease in PA in this time interval, but it had a small effect size of d = .27. The caffeine group did experience a decrease in PA with a large effect size of d = .95, but this occurred in the interval from post-Hour 2 to post-Hour 3, which is one hour later than the occurrence of the first large decrease in the placebo group. Effect size estimates for all hourly comparisons are reported in Table 5.

The repeated-measures ANVOA on NA in the 57 caffeine group participants revealed a significant effect of hour, $F(2.90, 162.62) = 4.24$, $p < .01$, d = .76 (please see Table 3 for the full analyses). Effect size estimates for the differences in mean self-reported NA from one hour to the next suggest that, rather than changing the time course of feelings as was the case with fatigue and PA, caffeine seems to have led to a more constant, steady experience of NA. The caffeine group experienced changes in NA which had sizes around the small and medium thresholds; specifically, the caffeine group experienced a medium increase in NA from the post-Hour 1 to the post-Hour 2 measurement (d = .58) and a small increase in NA from the post-Hour 2 to the post-Hour 3 measurement (d = .22). In these same two time intervals, the placebo group experienced

larger increases in NA, with effect sizes of d = .80 and d = .75, respectively. Instead of delaying a large change in felt emotions, caffeine appears to have prevented a large change in the case of NA. Caffeine did not, however, alleviate all negative emotions, as the caffeine group did experience increasing NA over the course of the session. But these increases were smaller in magnitude than those experienced by the placebo group. Effect size estimates for all hourly comparisons are reported in Table 5.

A separate set of additional analyses was conducted in an effort to understand the nature of any expectancy effects that may have occurred. Participants were classified into three preliminary groups based on their response to a final questionnaire item in which they were asked to rate, on a scale from 0 to 100%, how certain they were that they gum they chewed during the session was caffeinated. A response of 0% corresponded to "certain gum is not caffeinated", a response of 50% corresponded to "can't tell if gum is caffeinated", and a response of 100% corresponded to "certain gum is caffeinated". Because participants responding at or near the 50% mark were unlikely to have had a strong opinion as to whether their gum was caffeinated, those participants responding between 40% and 60% were classified as "neutral" and were excluded from the analyses on expectancy effects (n = 39, 16 from the caffeine group and 23 from the placebo group). Participants responding below 40% were classified as thinking that they had received non-caffeinated gum and participants responding above 60% were classified as thinking they had received caffeinated gum.

The two non-neutral groups (i.e., all participants responding below 40% or above 60% on this item) were further subdivided based on whether the participants had been in the caffeine or placebo condition, resulting in the following four groups: participants in

the caffeine condition who thought their gum was caffeinated (referred to as "CaffYes", n = 35), participants in the caffeine condition who thought their gum was non-caffeinated (referred to as "CaffNo", n = 7), participants in the placebo condition who thought their gum was caffeinated (referred to as "PlacYes", n = 17), and participants in the placebo condition who thought their gum was non-caffeinated (referred to as "PlacNo", n = 18). Please note that all analyses using this group classification scheme are based on an extreme-group post-hoc design, and as such should be cautiously interpreted.

Because of the unequal sample sizes for these four groups, the non-parametric equivalent of a one-way ANOVA, the Kruskal-Wallis test, was used in order to assess mean differences between groups on any variables of interest measured at time points after gum administration (because expectancy effects were thought to be driven by perceptions of the gum, it seemed unnecessary to include pre-gum administration measurements of the variables of interest in analyses of expectancy effects). For performance measures, the four groups did not differ significantly on average performance during Hour 2, Hour 3 or Hour 4. For the self-report measures, there were only two measurement points at which significant differences occurred between the groups: for PA at the post-Hour 2 measurement, χ^2 (3, N = 76) = 8.427, $p < .05$, and for fatigue at the post-Hour 2 measurement, χ^2 (3, N = 76) = 12.27, $p < .05$. Please see Table 7 for group means and Kruskal-Wallis analyses (please note that Chi-square values are reported for these analyses).

The graphs in Figures 8 and 9 give some indication as to where these between-group differences might be occurring. For self-reported fatigue at the post-Hour 2 measurement (Figure 8), it looks like there is at least a slight expectancy effect at work:

Table 7.

Means and non-parametric Kruskal-Wallis analyses for performance and self-report variables for the following groups: PlacYes, placebo group participants who were more than 60% sure their gum contained caffeine; PlacNo, placebo group participants who were less than 40% sure their gum contained caffeine; CaffYes, caffeine group participants who were more than 60% sure their gum contained caffeine; CaffNo, caffeine group participants who were less than 40% sure their gum contained caffeine

	PlacYes	**PlacNo**	**CaffYes**	**CaffNo**	**Chi-square**	**sig**	**df**
n	17	18	35	7	n/a	n/a	n/a
AvgHr1Perf	64.95	59.51	63.69	63.90	n/a	n/a	n/a
AvgHr2Perf	66.10	61.46	67.82	68.41	2.11	0.55	3
AvgHr3Perf	65.31	61.08	68.41	67.40	2.25	0.52	3
AvgHr4Perf	66.41	62.17	67.68	68.18	2.11	0.55	3
BaseFat	3.17	3.11	2.85	3.38	n/a	n/a	n/a
PostHr1Fat	2.86	3.03	2.61	3.18	2.49	0.48	3
PostHr2Fat	3.21	3.49	2.56	3.34	12.27	0.01	3
PostHr3Fat	3.51	3.76	3.05	3.60	6.54	0.09	3
PostHr4Fat	3.46	3.59	3.15	3.68	2.92	0.40	3
BasePA	1.83	1.91	1.88	1.86	n/a	n/a	n/a
PostHr1PA	2.04	1.80	2.05	1.82	1.87	0.60	3
PostHr2PA	1.79	1.44	2.06	1.61	8.43	0.04	3
PostHr3PA	1.60	1.28	1.64	1.35	3.97	0.26	3
PostHr4PA	1.84	1.56	1.76	1.51	2.28	0.52	3
BaseNA	1.47	1.60	1.40	1.52	n/a	n/a	n/a
PostHr1NA	1.30	1.70	1.38	1.58	7.17	0.07	3
PostHr2NA	1.51	1.78	1.45	1.59	5.90	0.12	3
PostHr3NA	1.58	1.87	1.49	1.68	6.00	0.11	3
PostHr4NA	1.59	1.80	1.55	1.80	1.49	0.68	3

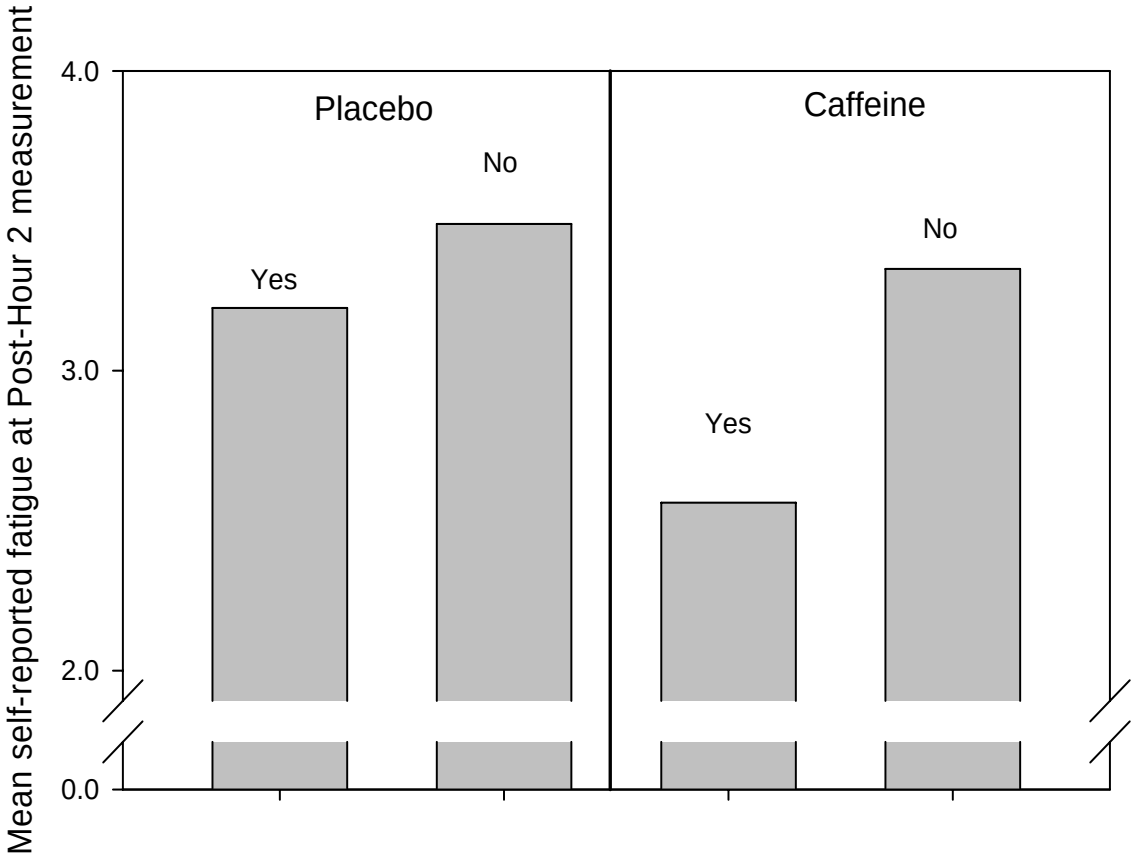

Figure 8.

Self-reported fatigue levels at the post-Hour 2 measurement for participants in the following groups: placebo group participants who were more than 60% sure their gum contained caffeine ("Yes"); placebo group participants who were less than 40% sure their gum contained caffeine ("No"); caffeine group participants who were more than 60% sure their gum contained caffeine ("Yes"); caffeine group participants who were less than 40% sure their gum contained caffeine ("No").

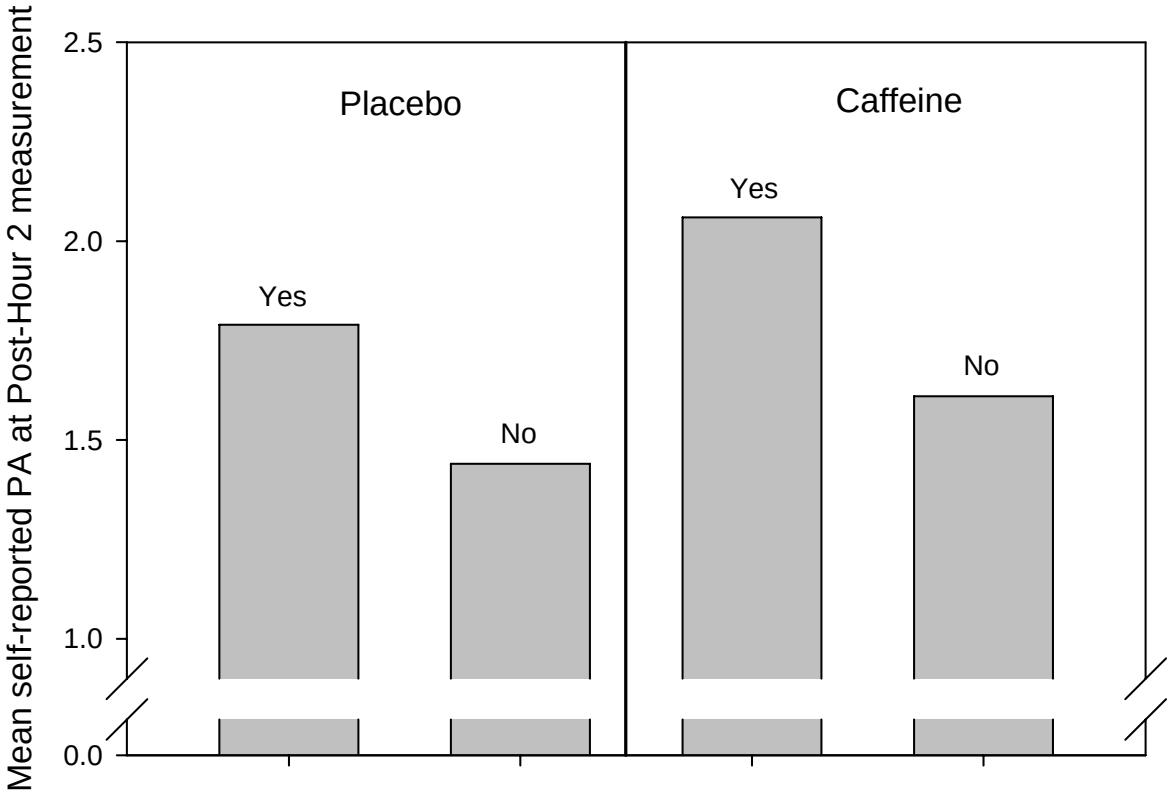

Figure 9.

Self-reported positive affect (PA) levels at the post-Hour 2 measurement for participants in the following groups: placebo group participants who were more than 60% sure their gum contained caffeine ("Yes"); placebo group participants who were less than 40% sure their gum contained caffeine ("No"); caffeine group participants who were more than 60% sure their gum contained caffeine ("Yes"); caffeine group participants who were less than 40% sure their gum contained caffeine ("No").

participants who thought they had caffeinated gum ("PlacYes" and "CaffYes" groups), whether they actually did or not, appear to have reported slightly lower fatigue levels than those participants who thought they had non-caffeinated gum ("PlacNo" and "CaffNo" groups). A similar effect can be seen in the self-reported PA levels at the post-Hour 2 measurement (Figure 9), as participants who thought they had caffeinated gum ("PlacYes" and "CaffYes" groups) appear to have reported slightly higher PA levels than those participants who thought they had non-caffeinated gum ("PlacNo" and "CaffNo" groups), regardless of what type of gum they actually had. If there were no expectancy effects occurring in this data, we would expect that both caffeine groups would be lower on fatigue and higher on PA than both placebo groups, regardless of what type of gum they thought they had.

The lack of significant between-group mean differences on the performance measure with simultaneous significant between-group differences for two of the self-report variables at one measurement point each suggests that subjective feelings may be slightly more vulnerable to expectancy effects than performance measures. Furthermore, the influence of expectancy effects in this sample does not seem to have been pervasive, as the majority of the analyses of self-report and performance variables did not indicate significant differences between these four groups.

3.4 Analysis of TPQ Items

The task perceptions questionnaire included similar fatigue, PA and NA self-report items as the prior interim questionnaires as well as approximately 25 items on caffeine consumption habits, strategy use during the session, perceptions of the gum chewed during the session, and plans immediately following the session. Chi-square tests

were used to analyze the response patterns on these items in order to identify differences between the caffeine and placebo conditions. Data from all participants was included in these analyses, unless otherwise specified.

For the item pertaining to the participants' strategy use, participants were asked whether they increased effort, kept effort constant, decreased effort, or first increased then later decreased effort. These responses correspond to the overactivity, normal, underactivity, and mixed groups, respectively. The caffeine and placebo groups did not differ significantly in their patterns of responding. For this and all further analyses of TPQ items, please see Table 8 for the item wording, response frequencies or means, and Chi-square or Mann-Whitney analyses. Participants in the two groups also did not differ significantly in their pattern of responses on the items regarding how fatiguing they found the session to be and what activity they planned to do first upon completing the session.

The two groups did differ significantly in their response pattern for the item asking when they thought they performed at their peak level χ^2 (1, N = 115) = 13.43, $p <$ .01. All participants were included in this analysis, except for one caffeine group participant who was eliminated due to giving multiple responses to the item. Based on the values presented in Table 9, it appears that the placebo group's responses were fairly evenly distributed among the four hours whereas the caffeine group's responses were concentrated in Hours 2 and 3. The caffeine group may have felt a boost after caffeine consumption that led them to believe their performance was highest in Hours 2 and 3. The two groups also differed significantly on the item asking whether they would probably consume a caffeinated beverage after leaving the session χ^2 (1, N = 115) = 4.20, $p < .05$. All participants were included in this analysis, except for one caffeine group

Table 8. Analysis of TPQ items

Multiple choice TPQ items, frequencies, and Chi-square analyses						
	CountCaff	**%Caff**	**CountPlacebo**	**%Placebo**	**Chi-square**	**Significance**
During the session, I…..					3.446	p > .05
a) increased my effort	9	15.52%	12	20.69%		
b) kept my effort at a constant level	19	32.76%	20	34.48%		
c) decreased my effort	3	5.17%	7	12.07%		
d) first increased my effort, then later decreased my effort	27	46.55%	19	32.76%		
I think that I performed at my peak level….					13.431	p < .01
a) during the first hour	5	8.77%	13	22.41%		
b) during the second hour	29	50.88%	18	31.03%		
c) during the third hour	20	35.09%	14	24.14%		
d) during the fourth hour	3	5.26%	13	22.41%		
Completing these tests was…					0.353	p > .05
a) less fatiguing than taking a course final exam	20	34.48%	19	32.76%		
b) about the same level of fatigue as taking a course final exam	18	31.03%	16	27.59%		
c) more fatiguing than taking a course final exam	20	34.48%	23	39.66%		
When I finish the session today, I will probably first…					8.798	p > .05
a) go home and take a nap or rest	13	22.41%	20	34.48%		
b) eat a meal	30	51.72%	34	58.62%		
c) study	5	8.62%	2	3.45%		
d) exercise	3	5.17%	0	0.00%		
e) have some leisure activity	7	12.07%	2	3.45%		

Table 8, continued.

	CountCaff	%Caff	CountPlacebo	%Placebo	Chi-square	Significance
When I finish this session, I will probably have a caffeinated beverage (e.g., coffee, a soft drink, etc.)					4.195	p < .05
a) No	24	42.11%	14	24.14%		
b)Yes	33	57.89%	44	75.86%		
Did you feel different after you chewed the gum during the session today?					11.293	p < .01
No	17	29.31%	35	60.34%		
Yes	41	70.69%	23	39.66%		
Did you feel the effects start to wear off at any time during the session?					0.085	p > .05
Note: this item includes only those participants who indicated that they felt different after chewing the gum, n = 64						
Yes	35	85.40%	19	82.60%		
No	6	14.60%	4	17.40%		
Response scale TPQ items, means, SDs, and Mann-Whitney analyses						
	CaffMean	**CaffSD**	**PlaceboMean**	**PlaceboSD**	**Mann-Whitney *U* value**	**Significance**
How certain are you that the gum you were given during the session contained caffeine?						
- response scale from 0 to 100% with 0% = "Certain Gum IS NOT Caffeinated" and 100% = "Certain Gum IS Caffeinated"	63.79	20.42	48.16	25.61	1055.00	p < .01

Table 8, continued.

	CaffMean	CaffSD	PlaceboMean	PlaceboSD	Mann-Whitney *U*	Significance
Rate the taste of the gum that you chewed during the session.						
- response scale from 1 to 6, with 1 = "Extremely Unpleasant" and 6 = "Extremely Pleasant"	2.26	1.12	4.57	1.05	297.00	$p < .01$
Rate the sensation/feelings of chewing the gum in terms of how different it was from chewing regular gum.						
- response scale from 1 to 5, with 1 = "No Difference" and 5 = "Extreme Difference"	3.50	1.03	1.98	1.12	583.00	$p < .01$
How soon after chewing the gum did you start to feel effects?						
Note: this item includes only those participants who indicated that they felt different after chewing the gum, n = 64						
- participants asked to respond with a number in minutes	11.10	12.50	15.59	15.79	391.00	$p > .05$
Were the effects less intense, more intense, or about the same level of intensity as compared to the feelings you get from consuming a typical dose of caffeine?						
Note: this item includes only those participants who indicated that they felt different after chewing the gum, n = 64						
- response scale from 1 to 5, with 1 = "Much Less Intense" and 5 = "Much More Intense"	2.66	0.965	2.46	0.94	411.50	$p > .05$

Table 9.

Percentage of caffeine and placebo conditions endorsing each response choice for the question "I think that I performed at my peak level during the…"

	1st hour	**2nd hour**	**3rd hour**	**4th hour**
Caffeine	8.77%	50.88%	35.09%	5.26%
Placebo	22.41%	31.03%	24.14%	22.41%

participant who was eliminated due to having given no response. Among caffeine group participants, 42.11% responded “No” and 57.89% responded “Yes”; among placebo group participants, 24.14% responded “No” and 75.86% responded “Yes”. These values suggest that the placebo group participants were more likely to be planning on getting a caffeinated beverage after the session than were the caffeine group participants.

As a way to see whether participants were able to tell if they had been given the caffeine or placebo gum, participants were asked to provide a percentage ranging from 0 to 100 indicating how certain they were that the gum they chewed during the session contained caffeine. The caffeine group participants gave a mean response of 63.79 (SD = 20.42) while the placebo group participants gave a mean response of 48.16 (SD = 25.61). A Mann-Whitney U test suggested that these means are significantly different, Mann-Whitney U = 1055.00, $p < .01$. So the groups did differ in their belief that they were chewing caffeinated gum, and in the direction of making a correct prediction (i.e., the caffeine group participants were more certain that they had chewed caffeine gum whereas the placebo group participants were less so).

It is unclear what is driving this difference, though. Did participants perceive a difference in the gum that they chewed, or did they feel different after chewing the gum? The answer to both of these questions is yes. Participants were asked two questions about the experience of chewing the gum they were given during the session. The first item pertained to the taste of the gum; participants were asked to rate the taste of the gum on a scale from one to six, with one being extremely unpleasant and six being extremely pleasant. Participants in the caffeine group responded with a mean of 2.26 (SD = 1.12), which corresponds to slightly higher than “Moderately Unpleasant”. Participants in the

placebo group rated the taste of the gum more favorably, with a mean of 4.57 (*SD* = 1.05). A Mann-Whitney U test indicated that these mean responses are significantly different, Mann-Whitney $U = 297.00$, $p < .01$.

Participants were also asked to indicate how different the experience of chewing the gum they received during the session was from chewing regular gum. The caffeine group participants gave a mean response of 3.50 (*SD* = 1.03), corresponding to a point midway between the scale markers of "Moderate Difference" and "Large Difference". The placebo group participants provided a mean response of 1.98 (*SD* = 1.12), which is just below the scale marker "Slight Difference". A Mann-Whitney U test indicated that these mean responses are significantly different, Mann-Whitney $U = 583.00$, $p < .01$. So it seems that the caffeine group participants experienced their gum differently than did the placebo group participants; specifically, they rated the gum as having a more unpleasant taste and as being more different from regular chewing gum than did the placebo group participants.

Participants were also asked whether they felt different after chewing the gum during the session. A chi-square test indicated that the groups differed significantly in their response patterns on this item, χ^2 (1, N = 116) = 11.293, $p < .01$. Among caffeine group participants, 29.31% responded "No" and 70.69% responded "Yes"; among placebo group participants, 60.34% responded "No" and 39.66% responded "Yes". These values suggest that caffeine group participants were more likely to respond yes to this item than placebo group participants. So the difference between the two conditions in reported levels of confidence that the gum was caffeinated seems to be driven both by

differences in perceptions of the gum (in terms of taste and comparison to regular chewing gum) as well as differences in subjective feelings after chewing the gum.

A set of follow-up items was presented immediately after the question concerning whether the participants felt different after chewing the gum; participants were instructed to respond to these three items only if they had answered that they did feel differently after chewing the gum. The analyses for these three items only include the 64 participants who responded that they did feel differently after chewing the gum. Participants in the two conditions did not differ significantly on their responses to any of these three follow-up items, which pertain to how soon after chewing the gum they started to feel the effects, whether they felt these effects begin to wear off at any time during the session, and how the effects of the gum compared to the effects associated with consumption of the participant's typical caffeine dose. These findings suggest that for those who perceived effects of the gum, there were no significant differences between those who chewed the caffeine gum and those who chewed the placebo gum on these three aspects of their perceptions of the gum.

CHAPTER 4

DISCUSSION

4.1 Performance effects

Consistent with prior findings (e.g., Ackerman et al., 2008), performance on the Cloze task was expected to remain relatively stable over time at the aggregate level with increasing time-on-task in the no caffeine condition. The lack of a significant effect of hour in the repeated-measures ANOVA for the placebo group suggests that this was the case; the high power (*power* = .99 for this analysis) allows for statements about the truth of the null hypothesis (i.e., that there were no significant differences between mean Cloze task performance for the four hours). This is because the high power to detect an effect if one had been present allows for a reasonably high level of confidence that an effect was not found because there truly was no effect, rather than no effect being found due to low power. Some caution must be taken in making such an interpretation even with this level of power, but it seems reasonable to draw some conclusions about truth of the null hypothesis in this case. The placebo group showed stable aggregate level performance over the course of the 4.5 hour testing session, a finding in line with both predictions for this study and prior research.

Individuals were expected to show varying reactions to increasing time-on-task, with their performance effects falling into one of the four groups theorized by Davis (1946). Using a classification scheme based on the presence or absence of performance increments and decrements within each hour of testing, participants were classified into one of four groups based on their overall performance. Caffeine administration was

expected to decrease occurrences of the reactions associated with increasing time-on-task among placebo group participants (drops and spurts in performance). This prediction was not supported, as there were no significant differences in the number of drops or spurts exhibited by the two groups after caffeine administration.

An increase in the aggregate level of performance was also expected to be associated with caffeine administration. This prediction was supported somewhat, as performance of the caffeine group participants was significantly higher than that of the placebo group participants within Hour 3, with a corresponding effect size close to the medium threshold. The higher mean performance levels of the caffeine group as compared to the placebo group within Hours 2 and 4 failed to reach significance, but the effect sizes were between the small and medium effect size cutoffs (d = .35 for Hour 2 and Hour 4). Additionally, the participants within the caffeine group showed an increase in performance with a large effect size (d = 1.33) from Hour 1 to Hour 2; furthermore, this increase does not appear to have dissipated over the course of the testing session (see Figure 3). The placebo group showed no such improvement, exhibiting fairly stable performance over the course of the testing session.

In summary, both between group and within groups comparisons support the notion of at least slightly improved complex task performance being associated with caffeine administration. Analysis of the t-tests between the group's mean performance levels for Hours 2, 3, and 4 and the corresponding effect size estimates suggest elevated performance for the caffeine group as compared to the placebo group. The t-test for Hour 1 performance reveals that the two groups started off at equivalent performance levels. The caffeine group then experienced a performance increase that remained in place for

the remainder of the session, while the placebo group continued to perform near their own baseline level. While these results are somewhat modest, they do illustrate the notion that caffeine carries benefits which may include improved performance on a complex, real-world relevant cognitive task.

4.2 Self-reports

As time-on-task increased, self-reports of fatigue and NA were expected to increase and self-reports of PA were expected to decrease. The expected increases in fatigue with increasing time-on-task among the placebo group participants were supported by the results. The increases in self-reported fatigue during the intervals between both the post-Hour 1 and post-Hour 2 measurements and the post-Hour 2 and post-Hour 3 measurements had large effect sizes (d = 1.29 and d = 1.08, respectively). So it appears that during the middle of the testing session, placebo group participants felt increasing amounts of fatigue. The effect sizes for the placebo group's decrease from the baseline to the post-Hour one measurement and for the decrease from the post-Hour 3 to the post-test measurement were both small (d = .37 and d = .36, respectively), indicating that the changes over these two intervals were less important than those changes occurring during the middle portion of testing. So increased fatigue among placebo group participants did occur as hypothesized during the 2nd and 3rd hours of testing.

The hypothesis of decreasing PA with increasing time-on-task in the placebo group was also supported. The decrease in PA from the post-Hour 1 to the post-Hour 2 measurement had a large effect size (d = 1.44), and the decrease in the following interval (post-Hour 2 to post-Hour 3) had a medium-sized effect approaching the large effect cutoff (d = .77). There was virtually no change in PA levels from baseline to the post-

Hour 1 measurement (d = .01), but there was a substantial increase in PA during the last hour of the session (d = .80 for the increase in PA from the post-Hour 3 to the post-test measurement). This is likely a result of positive feelings associated with the participants' awareness that they had nearly reached the end of the session. As was the case with changes in fatigue among the placebo group participants, the predicted decreases in PA did occur but were isolated to the middle portion of testing. Furthermore, a substantial increase in PA was seen during the last hour of testing; this increase was not predicted but is in line with findings from previous research (e.g., Ackerman et al., 2008) and lends itself to the notion that participants felt positively about nearing the end of the session.

As with the other two self-report variables, the results lend support to the predicted increases in NA with increasing time-on-task among the placebo group participants. They showed almost no change in the interval from baseline to the post-Hour 1 measurement (d = .03), then exhibited a large increase from the post-Hour 1 to the post-Hour 2 measurement (d = .80) and a slightly smaller increase from the post-Hour 2 to the post-Hour 3 measurement (d = .75). Similar to the increase in PA observed during the last hour of testing, these participants also showed a decrease in NA during the last hour of testing, but this effect was smaller than was seen with PA (d = .43). So the predicted increases in NA were seen during the 2nd and 3rd hours of testing, and positive attitudes thought to be associated with nearing the end of the session were evident in the modest decrease in NA during the final hour of the session. In general, these results for the three self-report variables are in line with our expectations, and also with prior research related to self-reported fatigue, PA, NA and increasing time-on-task.

Administration of caffeine was expected to attenuate these effects, and to lead to increased self-reports of PA and decreased self-reports of fatigue and NA among caffeine group participants, relative to the placebo group. Overall, the support for this prediction was fairly strong. Evidence is provided by both the between and within group analyses; both of these will be compared for the caffeine and placebo group's self-reports of fatigue, PA, and NA.

For fatigue, the between group comparisons (independent samples t-tests) suggest significantly higher fatigue levels for placebo group participants as compared to caffeine group participants at the post-Hour 2, post-Hour 3, and post-test measurements. These differences have effect sizes of $d = .71$, $d = .64$, and $d = .45$, respectively. These results indicate strongly that caffeine administration was associated with reduced feelings of fatigue. The within group comparisons (separate one-way repeated-measures ANOVAs for each treatment group, along with estimates of effect sizes for the differences between subsequent measurement points in each treatment group) indicate that caffeine may have changed the time course of feelings of fatigue. Both groups exhibited large increases in fatigue, but these large increases were seen an hour earlier in the placebo group as compared to the caffeine group. The placebo group showed increases with large effect sizes during the 2nd and 3rd hours of testing while the caffeine group did not have an increase with a large effect size until the 3rd hour. So caffeine did not erase feelings of fatigue, but did serve to delay them. In a real-world setting, repeated and/or caffeine larger doses may work to extend this delay, allowing consumers a longer reprieve from fatigue feelings, but these results suggest that caffeine delays feelings of fatigue that inevitably will occur, rather than eliminating them altogether. Both between and within

group comparisons suggest that the caffeine group fared better with regard to feeling fatigued: they felt less fatigued than their non-caffeinated counterparts, and they also felt the onset of fatigue approximately one hour later.

Turning to PA, the between group comparisons indicate significantly higher PA levels in the caffeine group as compared to the placebo group within Hours 2 and 3; the effect sizes are medium for both (d = .63 and d = .47 for Hours 2 and 3, respectively). Significant differences would not be expected at the baseline or post-Hour 1 measurements, and the results support these expectations. It was somewhat unexpected, however, to find that the groups had reached equivalent levels of PA at the post-test measurement. A possible explanation for this is that positive feelings about reaching the end of the session would have been felt equally by both groups, and may have been strong enough to override the remaining effects exerted by caffeine on PA, acting as a sort of neutralizer of existing differences between the two treatment groups. The within group comparisons suggest that a similar delaying mechanism may have occurred with caffeine administration. A large decrease in PA was experienced by the placebo group within the 2nd hour of testing (d = 1.44 for the interval between the post-Hour 1 and the post-Hour 2 measurement); the caffeine group did not experience a similarly large increase until the 3rd hour of testing (d = .95 for the interval between the post-Hour 2 and the post-Hour 3 measurement). As with fatigue, the PA results suggest that caffeine allowed participants both to experience higher PA levels during the middle portion of testing and to hold off on experiencing a large drop in PA for an hour. Further investigation into the effects of caffeine on the time course of fatigue and PA would be

interesting. A study in which participants were given caffeine at different points in time and tested for a longer period of time would allow for further exploration of these effects.

For NA, the between-group comparisons reveal fewer significant differences than were seen with fatigue and PA. The only significant difference between mean NA levels in the two groups occurred during Hour 3 (d = .45), where the caffeine group participants reported a lower mean NA level than the placebo group participants. So NA was experienced more similarly by the two groups than were the other two self-report variables.

Where caffeine delayed the onset of increased fatigue and decreased PA, the NA results tell a slightly different story. The caffeine group had no hourly changes with large effect sizes; the largest effect size goes with the change from the post-Hour 1 to the post-Hour 2 measurement (d = .58). Conversely, the placebo group had a large change during this interval (d = .80 for the change from the post-Hour 1 to the post-Hour 2 measurement) followed by a slightly smaller change during the next interval (d = .75 for the change from the post-Hour 2 to the post-Hour 3 measurement). The smaller within hour changes seen with the caffeine group suggest that caffeine may have led to a smoother, more stable experience of NA. Further investigation of this finding could prove to be fruitful; experience of negative emotions is problematic in many settings, particularly during important events at work or tests in school, and if caffeine were found to stabilize feelings of negative emotions this could be important information for a wide range of individuals.

4.3 Implications & future research directions

These results are important for two main reasons. First, they provide support for improved performance on a complex cognitive task associated with a single 170 mg caffeine dose. Most prior caffeine research has failed to address complex task performance; the performance effects most often studied with caffeine are isolated to lower-level tasks like reaction time or vigilance tasks. This study used a task that much more closely resembles something that would be encountered in school or in the workplace, arenas in which people frequently use caffeine in an attempt to improve their performance. These results give support to the numerous students and employees who flock to Starbucks, vending machines, and the office coffee pot: caffeine may do more than just make one feel better. It might actually improve performance on every-day, real-world relevant tasks. The performance effects of caffeine in this study represent a novel and unique finding, one that has been largely missing from prior caffeine research.

The caffeine dosage amount is a critical element of the importance of the finding of improved performance in this study. Some caffeine research has been criticized for using abnormally high caffeine doses, such that the amount of caffeine administered would not be reasonably consumed by most caffeine consumers within a single day. The dose used in this study is below the average caffeine consumption estimates discussed in the introduction, and as such represents a dose that is comparable to what regular caffeine consumers can be reasonably expected to consume. Due to both the reasonableness of the dose and the complexity of the task, these results represent some of the most real-world relevant findings regarding caffeine usage in the literature to date.

Secondly, these findings extend the work done by Smith (2009) in suggesting that caffeine gum elicits similar effects to caffeine delivered in more traditional vehicles. The findings of increased PA and decreased fatigue associated with caffeine intake are in line with those seen with caffeine delivered in beverage and capsule form. This is an important finding because countless research hours could be saved if caffeine gum became the delivery vehicle of choice in caffeine research, due to its much lower time requirement to become active after ingestion. Also, as caffeine gum becomes readily available in the market, it is important that researchers understand how its effects are similar to, or differ from, those associated with caffeinated beverages.

One weakness of this study is that it did not provide for examination of the extent to which the withdrawal hypothesis could be largely responsible for these results. The withdrawal hypothesis suggests that there are no net mood or performance benefits associated with caffeine consumption; rather, when these effects are seen in experimental studies, it is usually a result of alleviation of caffeine withdrawal symptoms (James & Rogers, 2005). This hypothesis is rooted in the idea that most caffeine study participants are regular caffeine consumers who have been made to be in a state of caffeine withdrawal by being instructed to eliminate caffeine consumption for some interval prior to the start of the study. Then these participants are split into caffeine and placebo groups, assessed at baseline, then assessed again after the caffeine group participants are given caffeine. Proponents of the withdrawal hypothesis would argue that all participants are experiencing caffeine withdrawal symptoms which exert negative effects on mood and performance. The caffeine participants then experience alleviation of these symptoms upon caffeine intake, resulting in mood and performance improvements relative to the

still-in-withdrawal placebo participants. So the apparent benefits of caffeine consumption merely represent a return to baseline after negative effects of withdrawal, rather than a benefit to the consumer.

There is much debate in the literature regarding the validity of this hypothesis as an explanatory mechanism for all positive caffeine findings, and various methodologies have been employed in an attempt to rule out or provide support for this hypothesis. Because my methodology fits exactly the criteria needed for the withdrawal hypothesis to be in effect (i.e., using regular caffeine consumers who are caffeine withdrawn as participants), I am not able to definitively state that my results could not be due to alleviation of withdrawal symptoms rather than actual benefits of caffeine consumption. One way to investigate the applicability of the withdrawal hypothesis to a study like this would be to include a third group, one whose participants were instructed to consume their regular caffeine doses in the morning and then were treated the same as the placebo group participants in the study. If this group started out at a higher level of mood and/or performance than the two groups whose participants had abstained from caffeine, and then the caffeine group's mood and/or performance levels shifted up to match the levels of this non-withdrawn group after caffeine administration, this would provide strong support for the withdrawal hypothesis. However, if this non-withdrawn group started off at a comparable level to the other two groups and followed a similar pattern to the placebo group participants, the withdrawal hypothesis could effectively be eliminated as an explanation for these results. Adding this third withdrawal-hypothesis testing group would represent a more substantial contribution to the extant literature than the present study, as it could offer contributions to the withdrawal hypothesis debate.

The current study investigates the presence or absence of an enhancing effect of caffeine on complex cognitive task performance, but does not attempt to explain why this performance enhancement may occur. A possible explanatory mechanism for the performance effects of caffeine is provided in an article by Ataka and colleagues (Ataka et al., 2008). In this study, caffeine consumption was regulated for eight days prior to the study and during the experimental session. Participants in the caffeine group, who were given 200 mg caffeine per day for eight days prior to the session and 100 mg caffeine during the experimental session, performed significantly higher than placebo group participants, who had received placebo capsules for eight days prior to and during the experimental session, on an advanced trail-making test. This performance effect of caffeine, however, may be a reward with a future cost.

Blood tests for plasma branched-chain amino acid (BCAA) levels, which represent a biomarker for mental fatigue, were lower among caffeine group participants as compared to placebo group participants at the conclusion of the testing session (lower BCAA levels represent greater mental fatigue). The authors suggest that caffeine leads to a temporary performance increase by increasing activation of the central nervous system, but that this current increased activation is associated with later mental fatigue, evidenced by the higher BCAA levels in caffeine group participants at the end of testing. These participants did not report higher feelings of fatigue than the placebo group participants, nor did they suffer any performance decrements (conversely, they performed at a higher level than the placebo group participants), but biomarkers indicate that their brains were more fatigued (Ataka et al., 2008).

Perhaps if the study had been carried out for a longer period of time, these biological indicators of fatigue would have manifested in impaired performance and/or increased self-reported fatigue. The current study demonstrates that feelings of mental fatigue occur in the absence of performance decrements; in both the caffeine and placebo groups, self-reported fatigue increased over the course of the session but performance either improved or remained relatively stable (for the caffeine and placebo groups, respectively). It would have been interesting to see how the time course of a biomarker for fatigue such as BCAA levels may have changed over time. Use of physiological measures like this would enhance mental fatigue research and represents an important area for future study. A finding that caffeine consumption provides only temporary benefits with high future costs in terms of fatigue might cause people to rethink their consumption habits. However, it is unclear that these biomarkers translate into outcomes that are relevant for peoples' daily lives, and further investigation is needed to understand their practical implications.

In general, the findings of this study are in line with those of prior research, with some additional contributions. First, the modest support for improved complex cognitive task performance represents an important addition to the literature as most studies have focused on lower-level tasks. Second, the use of caffeine gum is somewhat novel, as prior research using caffeine gum is limited to the military arena, with the exception of one study. Potential improvements to this study include a larger sample size to increase power both to detect effects and to allow for interpretation of the truth of a null hypothesis, implementation of a design which would allow for investigation of the extent to which

the results can or cannot be explained by the withdrawal hypothesis, and addition of one or more physiological measures of fatigue.

Appendix A

Sources for Cloze passages:

College Board SAT. (2002). *10 Real SATs*. College Entrance Examination Board, New York.

College Board SAT. (2004). *The Official SAT Study Guide: For the new SAT*. College Entrance Examination Board, New York.

College Board SAT. (2006). *The Official SAT Study Guide.* College Entrance Examination Board, New York.

Gleitman, H., Fridlund, A.J., & Reisberg, D. (1999). Psychology, 5th edition. W.W. Norton & Company, Inc. New York, NY.

Henretta, J.A., Brody, D., & Dumenil, L. (2002). America: A concise history, 2nd edition. Volume 1: To 1877. Bedford/St. Martin's. Boston, MA.

Kaplan Test Prep and Admissions. (2004). *The New SAT Critical Reading Workbook*. Simon & Schuster, New York.

Kaplan Test Prep and Admissions. (2007). *12 Practice Tests for the SAT.* Kaplan Publishing, New York, New York.

Oxtoby, D.W., Freeman, W.A., & Block. T.F. (1998). Chemistry: Science of Change, 3rd edition. Saunders College Publishing.

The Princeton Review. (2005). *11 Practice Tests for the NEW SAT & PSAT.* Random House, Inc. New York.

REFERENCES

Ackerman, P.L., Beier, M.E., & Bowen, K.R. (2000). Explorations of crystallized intelligence: Completion tests, cloze tests, and knowledge. *Learning and Individual Differences, 12,* 105-121.

Ackerman, P. L., & Kanfer, R. (2006). Test length and cognitive fatigue: Phase 1. Survey of the theoretical and empirical literature. Final report to the College Board. Author.

Ackerman, P.L., Kanfer, R., Wolman, S.D., & Haag, S. (2008). *An empirical investigation of cognitive fatigue during testing.* Manuscript under review.

Alsene, K., Deckert, J., Sand, P. & de Wit, H. (2003). Association between A_{2a} receptor gene polymorphisms and caffeine-induced anxiety. *Neuropsychopharmacology, 28*, 1694 – 1702.

Arai, T. (1912). *Mental Fatigue*. Teachers College, Columbia University: New York City.

Associated Press. (2007). Energy drink banned at Colorado school. *CNN.com, 2/23/07.*

Ataka, S., Tanaka, M., Nozaki, S., Mizuma, H., Mizuno, K., Tahara, T. et al. (2008). Effects of oral administration of caffeine and D-ribose on mental fatigue. *Nutrition, 24*, 233- 238.

Attwood, A.S., Higgs, S., & Terry, P. (2007). Differential responsiveness to caffeine and perceived effects of caffeine in moderate and high regular caffeine consumers. *Psychopharmacology, 190,* 169 – 477.

Barone, J.J., & Roberts, H.R. (1996). Caffeine Consumption. *Food & Chemical Toxicology, 34(1), 119 – 129.*

Benjamin, L.T., Rogers, A.M., & Rosenbaum, A. (1991). Coca-Cola, caffeine, and mental deficiency: Harry Hollingworth and the Chattanooga trial of 1911. *Journal of the History of the Behavioral Sciences, 27*, 42-55.

Brice, C.F., & Smith, A.P. (2002). Effects of caffeine on mood and performance: a study of realistic consumption. *Psychopharmacology, 164*, 188-92.

Cameron, C. (1973). A theory of fatigue. *Ergonomics, 16(5),* 633 – 648.

Childs, E. & de Wit, H. (2008). Enhanced mood and psychomotor performance by a caffeine-containing energy capsule in fatigued individuals. *Experimental and Clinical Psychopharmacology, 16*(1), 13 – 21.

Christian, M.S., & Brent, R.L. (2001). Teratogen update: Evaluation of the reproductive and developmental risks of caffeine. *Teratology, 64,* 51-78.

Christopher, M. (2006). Taking on the new SAT. *Colleges and Careers*, *Fall 2006*, p. 46.

Christopher, G., Sutherland, D., & Smith, A. (2005). Effects of caffeine in non-withdrawn volunteers. *Human Psychopharmacology, 20,* 47 – 53.

Committee on Military Nutrition Research, Food and Nutrition Board. (2001). *Caffeine for the sustainment of mental task performance: Formulations for military operations*. Washington, DC: National Academy Press.

Davis, D.R. (1946). The disorganization of behaviour in fatigue. *Journal of Neurology, Neurosurgery, and Psychiatry, 9,* 23 – 29.

Frary, C.D., Johnson, R.K., Wang, M.Q. (2005). Food sources and intakes of caffeine in the diets of persons in the United States. *Journal of the American Dietetic Association, Jan. 2005*, 110 – 113.

Fredholm, B.B., Battig, K., Holmen, J., Nehlig, A., & Zvartau, E.E. (1999). Actions of caffeine in the brain with special reference to factors that contribute to its widespread use. *Pharmacological Reviews, 51*, 83 – 133.

Haskell, C.F., Kennedy, D.O., Wesnes, K.A., & Scholey, A.B. (2005). Cognitive and mood improvements of caffeine in habitual consumers and habitual non-consumers of caffeine. *Psychopharmacology, 179(4),* 813 – 25.

Hendrick, B. (2006). Caffeine abuse among young is making docs jittery. *Atlanta Journal Constitution, 11/25/06*.

Hewlett, P. & Smith, A. (2007). Effects of repeated doses of caffeine on performance and alertness: New data and secondary analyses. *Human Psychopharmacology Clinical & Experimental, 22,* 339 – 350.

Hindmarch, I., Quinlan, P.T., Moore, K.L., & Parkin, C. (1998). The effects of black tea and other beverages on aspects of cognition and psychomotor performance. *Psychopharmacology, 139(3)*, 230 – 8.

Hindmarch, I., Rigney, U., Stanley, N., Quinlan, P., Rycroft, J., & Lane, J. (2000). A naturalistic investigation of the effects of day-long consumption of tea, coffee and water on alertness, sleep onset and sleep quality. *Psychopharmacology, 149(3),* 203 – 16.

Hollingworth, H.L. (1912). The influence of caffeine on mental and motor efficiency. *Archives of Psychology, 22,* 1 – 166.

James, J.E. (1991). Human reproduction. In J.E. James (Ed.), *Caffeine and Health* (pp. 19 – 244). London: Academic Press.

Jarvis, M. (1993). Does caffeine intake enhance absolute levels of cognitive performance? *Psychopharmacology, 110,* 45-52.

Kamimori, G.H., Johnson, D., Thorne, D., & Belenky, G. (2005). Multiple caffeine doses maintain vigilance during early morning operations. *Aviation, Space, and Environmental Medicine, 76(11)*, 1046-1050.

Kamimori, G.H., Karyekar, C.S., Otterstetter, R.O., Cox, D.S., Balkin, T.J., Belenky G.L., et al. (2002). The rate of absorption and relative bioavailability of caffeine administered in chewing gum versus capsules to normal healthy volunteers. *International Journal of Pharmaceutics, 234,* 159 – 167.

Klein, J. (2004). Planning middle school schedules for improved attention and achievement. *Scandinavian Journal of Educational Research, 48(4)*, 441-450.

Kohler, M., Pavy, A., & Van Den Heuvel, C. (2006). The effects of chewing versus caffeine on alertness, cognitive performance and cardiac autonomic activity during sleep deprivation. *Journal of Sleep Research, 15,* 358 – 368.

Lieberman, H.R. (1992). Caffeine. In: A.P Smith & D.M. Jones (Eds.), *Handbook of Human Performance, vol. 2*. (pp. 49 – 72). London: Academic Press.

Lieberman, H.R., Tharion, W.J., Shukitt-Hale, B., Speckman. K.L., & Tulley, R. (2002). Effects of caffeine, sleep loss, and stress on cognitive performance and mood during U.S. Navy SEAL training. Sea-Air-Land. *Psychopharmacology, 164(3),* 250 – 61.

Liu, J., Allspach, J.R., Feigenbaum, M., Oh, H., & Burton, N. (2004). *A study of fatigue effects from the New SAT*® (College Board Re. No. 2004-5). New York, NY: Educational Testing Service.

Lorist, M.M., & Tops, M. (2003). Caffeine, fatigue, and cognition. *Brain & Cognition, 53(1)*, 82 – 95.

Martin, P.Y., Hamilton, V.E., McKimmie, B.M., Terry, D.J., & Martin, R. (2007). Effects of caffeine on persuasion and attitude change: The role of secondary tasks in manipulating systematic message processing. *European Journal of Social Psychology, 37,* 320 – 338.

Martyn, G.W. (1913). A study of mental fatigue. *British Journal of Psychology, 5,* 427 – 446.

McCusker, R.R., Goldberger, B.A., & Cone, E.J. (2006). Caffeine content of energy drinks, carbonated sodas, and other beverages. *Journal of Analytical Toxicology, 30*, 112 – 114.

McCusker, R.R., Goldberger, B.A., & Cone, E.J. (2003). Caffeine content of specialty coffees. *Journal of Analytical Toxicology, 27*, 520 – 522.

McLellan, T.M., Bell, D.G., Lieberman, H.R., & Kamimori, G.H. (2003-2004). The impact of caffeine on cognitive and physical performance and marksmanship during sustained operations. *Canadian Military Journal, Winter 2003-2004*, 47-54.

McLellan, T.M., Kamimori, G.H., Voss, D.M., Bell, D.G., Cole, K.G., & Johnson, D. (2005). Caffeine maintains vigilance and improves run times during night operations for special forces. *Aviation, Space, and Environmental Medicine, 76(7)*, 647-654.

McNair, D.M., Lorr, M., & Droppleman, L.F. (1992). *Profile of Mood States*. (Revised). San Diego, CA: EdITS/Educational and Industrial Testing Service.

Melman, S., Little, S.G., & Akin-Little, K.A. (2007). Adolescent overscheduling: The relationship between levels of participation in scheduled activities and self-reported clinical symptomology. *The High School Journal, Feb/Mar 2007,* 18 – 30.

Myers, C.S. (1937). Conceptions of mental fatigue. *American Journal of Psychology, 50,* 296 – 306.

Nawrot, P., Jordan, S., Eastwood, J., Rotstein, J., Hughenholts, A., & Feeley, M. (2003). Effects of caffeine on human health. *Food Additives and Contaminants, 20,* 1 – 30.

Patat, A., Rosenzweig, P., Enslen, M., Trocherie, S., Miget, N., Bozon, M., et al. (2000). Effects of a new slow release formulation of caffeine on EEG, psychomotor and cognitive functions in sleep-deprived subjects. *Human Psychopharmacology, 15*, 153 – 170.

Peeling, P. & Dawson, R. (2007). Influence of caffeine ingestion on perceived mood states, concentration, and arousal levels during a 75-min university lecture. *Advances in Physiological Education, 31,* 332 – 335.

Reissig, C.J., Strain, E.C., & Griffiths, R.R. (2009). Caffeinated energy drinks – A growing problem. *Drug and Alcohol Dependence, 99,* 1 – 10.

Revelle, W., Amaral, P., & Turriff, S. (1976). Introversion/extroversion, time stress, and caffeine: Effect on verbal performance. *Science, 192,* 149 – 150.

Revelle, W., Humphreys, M.S., Simon, L., & Gilliland, K. (1980). The interactive effect of personality, time of day, and caffeine: A test of the arousal model. *Journal of Experimental Psychology: General*, *109(1)*, 1 – 31.

Rogers, P.J. (2007). Caffeine, mood and mental performance in everyday life. *British Nutrition Foundation Nutrition Bulletin*, *32,* 84 – 89.

Rogers, P.J., Richardson, N.J., Dernoncourt, C. (1995). Caffeine use: is there a net benefit for mood and psychomotor performance? *Neuropsychobiology, 31*, 195 – 199.

Ruxton, C.H.S. (2008). The impact of caffeine on mood, cognitive function, performance and hydration: A review of benefits and risks. *British Nutrition Foundation Nutrition Bulletin*, *33,* 15 – 25.

Scholey, A.B., & Kennedy, D.O. (2004). Cognitive and physiological effects of an "energy drink": an evaluation of the whole drink and of glucose, caffeine and herbal flavouring fractions. *Psychopharmacology, 176 (3-4),* 320-30.

Smit, H.J., & Rogers, P.J. (2000). Effects of low doses of caffeine on cognitive performance, mood and thirst in low and higher caffeine consumers. *Psychopharmacology, 152,* 167 – 173.

Smith, A. (2009). Effects of caffeine in chewing gum on mood and attention. *Human Psychopharmacology, 24,* 239 – 247.

Smith, A. (2002). Effects of caffeine on human behavior. *Food and Chemical Toxicology, 40*, 1243 – 1255.

Smith, A., Kendrick, A., Maben, A., & Salmon, J. (1994). Effects of breakfast and caffeine on cognitive performance, mood and cardiovascular functioning. *Appetite, 22,* 39 – 55.

Smith, A., Sutherland, D., & Christopher, G. (2005). Effects of repeated doses of caffeine on mood and performance of alert and fatigued volunteers. *Journal of Psychopharmacology, 19(6),* 620 – 6.

Stine, M.M., O'Connor, R.J., Yatko, B.R., Grunberg, N.E., & Klein, L.C. (2002). Evidence for a relationship between daily caffeine consumption and accuracy of time estimation. *Human Psychopharmacology, 17(7)*, 361 – 7.

Syed, S.A., Kamimori, G.H., Kelly, W., & Eddington, N.D. (2005). Multiple dose pharmacokinetics of caffeine administered in chewing gum to normal healthy volunteers. *Biopharmaceutics & Drug Disposition, 26*, 403-409.

Taylor, W.L. (1953). Cloze procedure: A new tool for measuring readability. *Journalism Quarterly, 30*, 415 – 433.

Thayer, R.E. (1978). Factor analytic and reliability studies on the activation-deactivation adjective check list. *Psychological Reports, 42,* 747 - 756.

Thorndike, E.L. (1900). Mental fatigue I. *Psychological Review, 7,* 466 – 482.

Thorndike, E.L. (1912). The curve of work. *Psychological Review, 19(3)*, 165 – 194.

Troyer, H. (1988). Review of cancer among 4 religious sects: Evidence that life-styles are distinctinve sets of risk factors. *Social Science Medicine, 26*(10), 1007 – 1017.

van der Linden, D., Frese, M., & Sonnentag, S. (2003). The impact of mental fatigue on exploration in a complex computer task: rigidity and loss of systematic strategies. *Human Factors, 45(3)*, 483 – 494.

Voelkle, M.C., Ackerman, P.L., & Wittmann, W.W. (2007). Effect sizes and *F* ratios < 1.0: Sense or nonsense? *Methodology*, *3*(1), 35 – 46.

Wahlstrom, K. (2003). Later high-school start times still working. *The Education Digest, February 2003,* 49-53.

Yeomans, M.R., Ripley, T., Davies, L.H., Rusted, J.M., & Rogers, P.J. (2002). Effects of caffeine on performance and mood depend on the level of caffeine abstinence. *Psychopharmacology, 164*, 241-249.

Zijlstra, F.R.H. (1993). *Efficiency in work behavior: A design approach for modern tools*. Delft: Delft University Press.