Caffeine's Effect on Appraisal and Mental Arithmetic Performance: A Cognitive Modeling Approach Tells Us More

Sue E. Kase (skase@ist.psu.edu), Frank E. Ritter (frank.ritter@psu.edu)
College of Information Sciences and Technology, Pennsylvania State University
University Park, PA 16802 USA

Michael Schoelles (schoem@rpi.edu)
Cognitive Science Department, Rensselaer Polytechnic Institute
Troy, NY 12180 USA

Abstract

A human subject experiment was conducted to investigate caffeine's effect on appraisal and performance of a mental serial subtraction task. Serial subtraction performance data was collected from three treatment groups: placebo, 200 mg caffeine, and 400 mg caffeine. Data were analyzed by average across treatment group and by challenge and threat task appraisal conditions. A cognitive model of the serial subtraction task was developed and fit to the human performance data. How the model's parameters change to fit the data suggest how cognition changes across treatments and due to appraisal. Overall, the cognitive modeling and optimization results suggest that the speed of vocalization is changed the most along with some changes to declarative memory. This approach promises to offer fine-grained knowledge about the effects of moderators on task performance.

Keywords: Caffeine, stress, task appraisal, cognitive arithmetic

Introduction

Caffeine is widely consumed throughout the world in beverages, foods, and as a drug for a variety of reasons, including its stimulant-like effects on mood and cognitive performance (for review see Fredholm et al., 1999). Its positive effects on performance, notably sustained vigilance and related cognitive functions, are well documented when administered to rested volunteers in doses equivalent to single servings of beverages (Amendola et al., 1998; Smith et al., 1999). Additionally, its consumption in moderate doses is associated with few, if any, adverse effects (Nawrot et al., 2003). Therefore, caffeine has been a strategy examined for its usefulness to military personnel (Lieberman & Tharion, 2002; McLellan et al., 2007).

The majority of caffeine research is conducted through human experimentation with analysis of the collected performance data. Few studies have attempted to *model* the effects of caffeine. One such study by Benitez et al. (2009) presented a biomathematical model for describing performance during extended wakefulness with the effect of caffeine as a stimulant.

Likewise, this study takes a modeling approach employing cognitive modeling and optimization techniques to investigate the effects of caffeine on cognitive performance. In particular, we examined the effects of caffeine and task appraisal during the arithmetic portion of the Trier Social Stress Test (TSST), a mental serial subtraction task. Based on human subject observations, self-reported appraisal, and performance data, we then developed a cognitive model in the ACT-R cognitive architecture of the serial subtraction task. Parametric solution sets resulting from optimizing the serial subtraction cognitive model to data from three treatment groups (placebo, 200 mg, 400 mg) and two task appraisal conditions (challenge and threat) provided the first cognitive modeling-derived insights on the cognitive effects of caffeine.

Method

This section begins with an overview of the human subject experiment where performance and task appraisal data were collected and later utilized in the development and optimization of a cognitive model. A detailed description of the cognitive task follows, as well as, the formulation of the self-reported appraisal conditions. Lastly, results and interpretations of the human performance data are suggested.

As part of a larger project, human subject data was collected to study the effects of stress and caffeine on cardiovascular health. The authors collaborated with Dr. Laura Klein and her lab in the Biobehavioral Health Department at Penn State University. A mixed experimental design was conducted with 45 healthy men 18-30 years of age (Klein, Whetzel, Bennett, Ritter, & Granger, 2006). (Men are typically used in these types of studies because we also took additional physiological measures and their systems are simpler.)

All subjects were asked to perform a series of three cognitive tasks. Subjects individually performed a simple reaction time (RT) and a working memory (WM) task taking 15 minutes to complete. Then subjects were administered one of three doses of caffeine: none (placebo), 200 mg caffeine (equivalent to 1-2, 8 oz cups of coffee), or 400 mg caffeine (equivalent to 3-4, 8 oz cups of coffee). After allowing absorption time, a 20-minute stress session of the mental arithmetic portion of the TSST was performed. Following completion of this stressor, subjects again were asked to complete the RT and WM tasks. Cognitive performance was determined by calculating accuracy and response time scores.

This paper focuses on one portion of the experiment—the TSST. The TSST protocol has been used for investigating psychobiological stress responses in a laboratory setting since the 1960s (Kirschbaum, Pirke, & Hellhammer, 1993). TSST traditionally consists of an anticipation period and a test period in which subjects have to deliver a free speech and perform mental arithmetic in front of an audience. The mental arithmetic portion of the TSST is a mental serial subtraction task.

Serial Subtraction Task

The serial subtraction task utilized in the experiment consisted of four 4-minute blocks of mentally subtracting by 7s and 13s from 4-digit starting numbers. Figure 1 illustrates the serial subtraction task. These were the four starting numbers used to begin the four blocks of subtraction during the experiment.

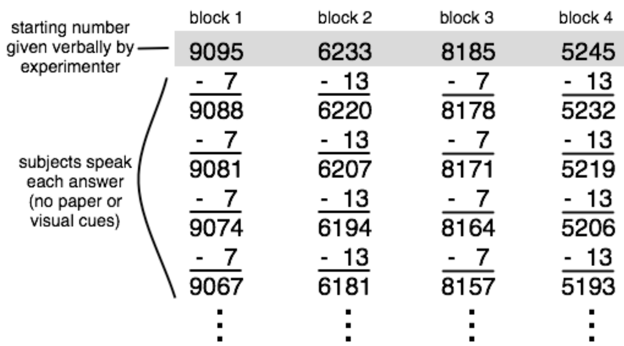

Figure 1: An illustration of the four blocks of the serial subtraction task as in the experiment.

Before the task begins the experimenter explains that the subject's performance is going to be voice recorded and reviewed by a panel of psychologists for comparison with the other subjects participating in the experiment. The task is performed mentally with no visual or paper clues. After the task is explained to the subject, a task appraisal questionnaire is completed, and the subject begins performing the task. It is thought that this anticipation period, for some subjects, increases anxiety and worry about poor performance on the upcoming task.

Subjects sit in a chair directly in front and near the experimenter who is holding a time keeping device and clipboard of the correct subtraction answers that she checks off as the subject performs the task. Before the task begins the experimenter emphasizes that the task should be preformed as quickly and as accurately as possible. An experimenter tells the subject the starting number; from then on, the subject speaks the answer to each subtraction problem. When an incorrect answer was given, the subject was told to "Start over at <the last correct number>". At two minutes into each 4-minute session, subjects were told that "two minutes remain, you need to go faster". This prompt enhances the time-pressure component of the task.

Task Appraisal

Before and after the serial subtraction stress session, subjects completed pre- and post-task appraisals based on Lazarus and Folkman's (1984) theory of stress and coping. Each subject was asked five questions orally: two focused on the subject's resources or reserves to deal with the serial subtraction task and three focused on the subject's perception as to how stressful the task would be.

For all questions the scale was from 1 to 5 with a value of 3 indicating that the subject is neither challenged nor threatened by the task. After correcting for the imbalance in questions, a ratio of perceived stress to perceived coping resources was created. For example, if a subject's total appraisal score was 1.5 or less, their perceived stress was less than or equal to their perceived ability to cope, which equated to a *challenge condition*. If a subject's appraisal score was greater than 1.5, their perceived stress was greater than their perceived ability to cope, which equated to a *threat condition*.

Each treatment group was composed of 15 subjects. The placebo group had approximately the same number of subjects in each appraisal condition (7 challenge, 8 threat). The 200 mg caffeine group had twice as many challenged subjects as threatened subjects (10 challenge, 5 threat). The 400 mg caffeine group contained only 2 challenged subjects with the remainder (13) subjects reporting a threatening appraisal.

Results and Discussion

For this investigation, the serial subtraction performance data from the placebo group (PLAC), the 200 mg caffeine group (LoCAF), and the 400 mg caffeine group (HiCAF), were analyzed by average across treatment group and by appraisal condition. The performance statistics of primary interest were number of attempted subtraction problems and a percentage correct score. The data are shown in Table 1 where each pair of values represents number of attempts and percent correct. The results discussed in this paper apply to data from the first block of subtracting by 7s.

Table 1: Human performance (average number of attempts and percent correct) by treatment group (each N=15) and appraisal condition (challenge, threat).

Treatment	Average	Challenge	Threat
PLAC	47.3, 81.5	50.7, 83.3	40.4, 77.9
LoCAF	59.1, 86.5	62.4, 88.3	37.5, 74.8
HiCAF	45.7, 79.2	51.6, 82.8	38.9, 75.1

For all treatment groups the challenge condition showed the best performance in both number of attempts and percent correct over the average across treatment and the threat condition. The threat condition showed the worst performance. Performance differences between the challenge and threat conditions were most pronounced in the LoCAF group with an impressive increase of nearly 25 more attempted subtraction problems and a 13.5% increase in subtraction accuracy by challenged subjects over threatened subjects. For the HiCAF group the challenge and threat condition differences were less than LoCAF but still substantial: 13 more attempted problems and a 7.7% increase in subtraction accuracy. Differences between the challenge and threat condition were least visible in the PLAC group, 10 more attempted problems and only a 5.4% increase in accuracy.

Figure 2 better illustrates these performance differences with the treatment groups labeled along the x-axis and the plot subdivided into three sections: averages across treatment groups (not by appraisal condition) in the leftmost section, and averages across treatment groups subdivided by appraisal condition in the center (challenge) and rightmost sections (threat).

The plot visualizes several interesting trends; some supported by existing caffeine and cognition research and others not. In the average across treatments plot (leftmost section), the performance of the HiCAF group drops below that of PLAC for both performance statistics. This supports findings that large doses of caffeine are occasionally associated with anxiety and disrupt performance (Haishman, & Henningfield, 1992; Wesensten, Belenky, & Kautz, 2002). Whether a 400 mg dose is considered 'large' may be in question as some studies administered up to 800 mg doses (McLellian et al., 2007). Generally, 100 to 300 mg doses are categorized as 'low' dosages because 50-300 mg of caffeine is available in a number of forms including tablets, chewing gum, a wide variety of beverages and some food products.

In the challenge condition (middle section), HiCAF performance does not drop below PLAC, but is approximately equivalent or slightly higher. In both the average across treatments and the challenge condition, LoCAF performance is well above that of PLAC. This is also supported in previous research that low doses of caffeine tend to increase performance (Amendola et al., 1998; Smith et al., 1999). In both these cases, the across treatments and challenge plots, the effects of caffeine take on characteristics related to level of arousal studies (i.e., Anderson & Revelle, 1982) and appear to follow the Yerkes-Dodson (1908) law that postulates that the relationship between arousal and performance follows an inverted U-shape curve.

There is no supporting research for the performance trends visible under the threat condition (right section). Threatened subjects self-reported stress and lack of coping skills to adequately perform the serial subtraction task. The threat plot shows performance decreases from PLAC to LoCAF (instead of increases as observed in the other sections of the plot) with HiCAF only very slightly higher than LoCAF (+1.4 attempts, and +0.3% correct). In this case, the U-shape is not inverted, but actually very slightly U-shaped.

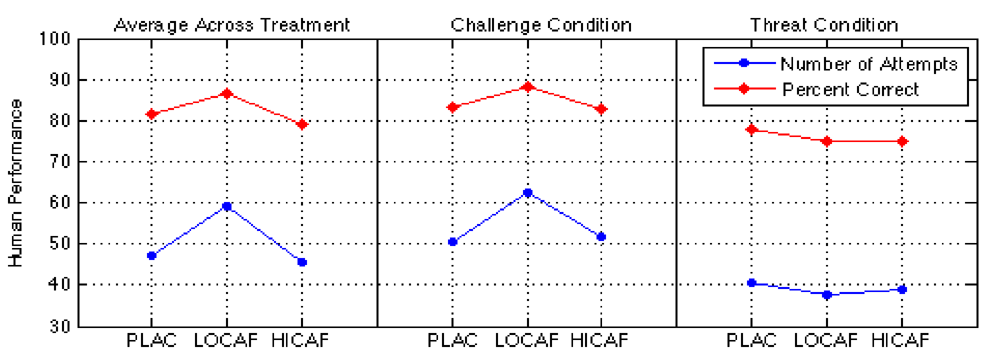

Figure 2: Comparing human performance differences in number of attempts and percent correct by treatment group (x-axis) and appraisal condition: treatment groups not accounting for appraisal (leftmost section), and averages across treatment groups divided by appraisal condition, challenge (middle section) and threat (rightmost section).

More can be discussed about the human performance data by way of analysis and interpretation of caffeine's effect on appraisal and serial subtraction. However, a more important question remains: Can these effects be modeled using a cognitive architecture and what might be learned from the parameters and values generating best fits during optimization of the model?

Modeling Serial Subtraction

Theory about how mental arithmetic is performed combined with observations gathered during the human subjects' performance of serial subtraction laid the foundation for the development of a cognitive model of the serial subtraction task. The ACT-R cognitive architecture (Anderson, 2007) was chosen to model the serial subtraction task for several reasons: it provides a parameter-driven subsymbolic level of processing; it permits the parallel execution of the verbal system with the control and memory systems, and it has been used for other models of addition and subtraction developed by other researchers.

The serial subtraction model performs a block of subtracting by 7s or 13s in a similar manner to that of the human subjects. The model's declarative knowledge consists of arithmetic facts and goal-related information. The model's procedural knowledge is production rules that allow for retrieval of subtraction and comparison facts necessary to produce an appropriate answer. The model performs subtractions by column-by-column.

The model runs under ACT-R 6.0 and utilizes the imaginal module and buffer. The imaginal buffer implements a problem representation capability. In the serial subtraction model the imaginal buffer holds the current 4-digit number being operated on (the minuend) and the number being subtracted (the subtrahend). The goal module and buffer implement control of task execution by manipulation of a state slot. ACT-R's vocal module and buffer verbalize the answer to each subtraction problem as the subjects do.

The model starts with the main goal to perform a subtraction and a borrow goal to perform the borrow operation when needed. Both types of goal chunks contain a state slot, the current column indicator, and the current subtrahend. The current problem is maintained in the imaginal buffer. This buffer is updated as the subtraction problem is being solved. The model begins with an integer minuend of 4-digits. All numbers in the model are chunks of type integer with a slot that holds the number. The model also contains subtraction and addition fact chunks whose slots are the integer chunks described above. This representation of the integers and arithmetic facts has been used in other ACT-R arithmetic models.

The model determines if a borrow operation is required by trying to retrieve a comparison fact that has two slots, a greater slot containing the minuend and a lesser slot containing the subtrahend. If the fact is successfully retrieved then no borrow is necessary, otherwise a borrow subgoal is created and executed. Borrowing is performed by retrieving the addition fact that represents adding ten to the minuend. The subtraction fact with the larger minuend is retrieved. The model then moves right one column by retrieving a next-column fact using the current column value as a cue. If this retrieval fails, there are no more columns so the borrow and the subgoal return back to the main task goal. If there is a next column and its value is not 0 than 1 is subtracted from it by retrieval of a subtraction fact. If the value is 0 then the problem is rewritten in the imaginal buffer with a 9 and the model moves to the next column and repeats the steps discussed above, returning to the main task when there are no more columns.

The model outputs the answer by speaking the 4-digit result. The model has two output strategies. For this paper the data reported are for the calc-and-speak strategy where the model speaks the answer in parallel with the calculation described above. If the answer is incorrect, the problem is reset to the last correct answer. If the answer is correct, the main problem task is rewritten in the imaginal buffer.

After the model has performed a block of subtractions the number of attempted subtraction problems and percent correct, are recorded. The model's performance can be adjusted by varying the values of architectural parameters associated with specific modules and buffers, and subsymbolic processes within the architecture.

Optimizing to Human Data

How does cognition change under stress and caffeine? We can explore this question by adjusting theoretically motivated parameters in architecture. The parameters that lead to better correspondences suggest how cognition changes. This section begins by discussing the architectural parameters selected for adjusting the model's performance to simulate the human data. This process of *fitting* the cognitive model to human data is a form of optimization. The optimization approach to fit the model is briefly described in the second part of the section. The optimization results, accompanied by interpretations of best fitting parameter values, is discussed at the end of the section.

Architectural Parameters

Three ACT-R architectural parameters appeared important in performing serial subtraction and were selected for adjusting the model's performance: seconds-per-syllable, base level constant, and activation noise. The rate the model speaks is controlled by the seconds-per-syllable parameter (SYL). The ACT-R default timing for speech is 0.15 seconds per assumed syllable based on the length of the text string to speak. There is a default of three characters per syllable controlled by the characters-per-syllable parameter. The seconds-per-syllable and characters-per-syllable parameters control subsymbolic processes in ACT-R's vocal module. The vocal module gives ACT-R a rudimentary ability to speak. It is not designed to provide a sophisticated simulation of human speech production, but to allow ACT-R to speak words and short phrases for simulating verbal responses in experiments such as the answers to the subtraction problems.

The other two parameters affect declarative knowledge access: the base level constant (BLC), and the activation noise parameter (ANS). The BLC parameter and a decay parameter affect declarative memory retrieval and retrieval time. The ANS value affects variance in retrieving declarative information and error rate for retrievals in the model. This instantaneous noise value can also represent variance from trial to trial. Other parameters, such as base level learning, decay, and the characters-per-syllable parameters were built into the model as modifiable but were left fixed at their default values for this study. The search space for the model optimization was defined by the parameter value boundaries: ANS and SYL 0.1 to 0.9, and BLC 0.1 to 3.0.

Optimization Approach

Because the search space was large and assumed to be rather complex a departure from the cognitive modeling community's traditional manual optimization technique was initiated (Kase, 2008). A new front-end function for the cognitive model was developed for execution in a parallel processing environment and the ACT-R parameter values (ANS, BLC, and SYL) were passed to multiple instances of running models from a parallel genetic algorithm (PGA). The SYL parameter was chosen for optimization because vocalization of the answer is the most time consuming aspect of this task. The BLC and ANS parameters were chosen because the task is memory intensive. Other memory parameters could have been chosen and ongoing work is exploring the fitting of other parameters. Normally, the parameter values are set within the model code before runtime. Using the PGA to search the parameter space for promising parameter value sets generating best fits between the model and human data saved a substantial amount of modeler time and computational resources. Model-to-data fit was determined by an objective function, or fitness function, defined as the discrepancy between model performance (number of attempts and percent correct) and the corresponding human performance (e.g., 47.3 – 48.1). The fitness is in terms of error (or cost) with a fitness value of 0 representing perfect correspondence between the model predictions and the human data.

Employing this type of 'automated' optimization approach allowed for 20,000 different sets of parameter value to be tested in a directed manner each time the PGA was executed. Using the approach, the model was optimized to nine sets of human performance data (see Table 2).

Results and Discussion

Table 2 shows the resulting model performance compared to the human performance data using parameter value solution sets identified by the PGA that produced the best fits (fitness values less than 1.0) to the human performance, and suggest how cognition changed. Several trends can be observed within the parameter values producing best fits. The parameter values shown in the table are averaged; denoted by the numeric value in parentheses after the parameter set values (i.e., '(3)' in the first row means that the PGA found 3 parameter sets producing fitness less than 1.0, and that these values were averaged). Each parameter set included in the average was run 200 times (i.e., 200 model runs per parameter set).

Beginning with the seconds per syllable parameter, SYL is shown in the last column and last value in the triple of Table 2. The model predictions indicate that challenged subjects speak a syllable more quickly than threatened subjects. This is true for all treatment groups. LoCAF shows the greatest difference in speech rate with challenge SYL at 0.31 (also lowest SYL overall) and threat SYL at nearly two times slower (0.61). HiCAF differences in SYL are less: challenge 0.40 compared to threat 0.57, a difference of 0.17. PLAC shows a slightly less SYL difference of 0.14. Challenge subjects self-report less stress and are generally confident that they can perform the serial subtraction task well. With less stress and a low dose of caffeine more fluid speech appears to result, or possibly the speech rate acts as a window into the cognitive processes required to complete the subtractions (i.e., fact retrieval, working memory and place-keeping operations, and concatenation of subsolutions).

Overall across treatments, the activation noise parameter values (ANS, first value in triple) are high as compared to what would be manually assigned to the model in the ACT-R modeling community. This could be because the nature of the task is stressful (i.e., purposively used to elicited a stress response). The ANS value range in Table 2 is narrow from the lowest ANS of 0.67 to the highest ANS of 0.78, a difference of only 0.11. This hints at the fact that caffeine may not effect this parameter's role in the model's performance of serial subtraction. ANS values are basically equivalent for the PLAC and LoCAF groups for challenge (0.68) and threat (0.71). In this case, the slightly higher ANS in predicting threatened subjects corresponds to the lower performance (less attempts and lower accuracy), and the self-reports where subjects do not believe they will perform well. Worrying or embarrassment about their poor performance is a distraction and may interfere with working memory processes and verbalizing solutions. The greatest variability in ANS values is found in HiCAF. Surprisingly, the trend reverses with HiCAF challenge predictions yielding a higher ANS value (0.75) than threat predictions (0.67).

The base level constant parameter values (BLC, middle value in triple) show a trend of nearly equivalent higher values for LoCAF and HiCAF challenge conditions (2.65 and 2.69) then threat conditions (2.48 and 2.35), and also for all BLC values under PLAC (2.49, 2.48 and 2.53). In this case, caffeine may be causing a 'boost' in the base level activation value of facts in declarative memory promoting higher probability of selection in response to a retrieval request and quicker fact retrieval time.

Table 2: Optimization results for three treatment groups (PLAC, LoCAF, HiCAF) and appraisal conditions (CH=challenge, TH=threat) comparing human performance and model predictions in number attempts and percent correct (both rounded), and fitness value associated with average (over N) of best fitting (less than 1.0) ACT-R parameter values (ANS, BLC, SYL).

	Human Performance	Model Prediction	Fitness Value	ACT-R parameters ANS, BLC, SYL (N)
PLAC (no caffeine)				
ALL	47.3, 81.5	48.1, 81.4	0.83	0.70, 2.49, 0.44 (3)
CH	50.7, 83.3	50.4, 83.0	0.47	0.68, 2.48, 0.41 (6)
TH	40.4, 77.9	40.3, 77.4	0.36	0.71, 2.53, 0.55 (5)
LoCAF (200 mg caffeine)				
ALL	59.1, 86.5	59.1, 86.7	0.12	0.72, 2.64, 0.33 (4)
CH	62.4, 88.3	62.7, 88.4	0.42	0.69, 2.65, 0.31 (3)
TH	37.5, 74.8	37.2, 74.9	0.58	0.71, 2.48, 0.61 (6)
HiCAF (400 mg caffeine)				
ALL	45.7, 79.2	44.7, 80.4	0.50	0.78, 2.65, 0.47 (4)
CH	51.6, 82.8	46.1, 87.7	0.53	0.75, 2.69, 0.40 (3)
TH	38.9, 75.1	50.4, 92.3	0.53	0.67, 2.35, 0.57 (4)

Conclusion

A cognitive model of the serial subtraction task was developed and fit to the human performance data from three caffeine treatments and by challenge and threat appraisal. This fit suggests that there are systematic changes in cognition due to caffeine and appraisal. Most notable is the speaking rate, but declarative memory retrievals are also affected.

These results show that using a cognitive model and parametric optimization approach can further our understanding of caffeine beyond a human experimentation approach. Overall, the cognitive modeling and optimization approach was successful. The preliminary modeling results and interpretations offer insight into the effects of caffeine on task appraisal and subsequent performance of the task, and promise an improved methodology for the study of other behavioral moderators and other cognitive tasks. At this point in our investigation more analysis is needed and additional parameter sets should be examined, along with continued refinement of the serial subtraction model for predicting the effects of caffeine on cognition.

Acknowledgments

This project is partially supported by ONR grant N000140310248. Computational resources were provided by TeraGrid DAC TG-IRI070000T and run on the NCSA clusters. The authors would like to thank Laura Klein and her lab and Jeanette Bennett at the Department of Biobehavioral Health, Penn State University, for collection of the human performance data and data analysis assistance.

References

Amendola, C. A., Gabrieli, J. D. E., & Lieberman, H. R. (1998). Caffeine’s effects on performance and mood are independent of age and gender. *Nutritional Neuroscience*, *1*, 269-280.

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* New York, NY: Oxford University Press.

Anderson, K. J., & Revelle, W. (1982). Impulsivity, caffeine, and proofreading: A test of the Easterbrook hypothesis. *Journal of Experimental Psychology: Human Perception and Performance*, *8,* 614-624.

Benitez, P. L., Kamimori, G. H., Balkin, T. J., Greene, A., & Johnson, M. L. (2009). Modeling fatigue over sleep deprivation, circadian rhythm, and caffeine with a minimal performance inhibitor model. *Methods in Enzymology*, *454*, 405-419.

Fredholm, B. B., Battig, L., Homen, J., Nehlig, A., & Zvarlaw, 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.

Haishman, S. J., & Henningfield, J. E. (1992). Stimulus functions of caffeine in humans: Relation to dependence potential. *Neuroscience & Biobehavioral Reviews, 16,* 273-287.

Kase, S. E. (2008). *HPC and PGA optimization of a cognitive model: Investigating performance on a math stressor task.* Unpublished PhD thesis, College of IST, Penn State University, University Park, PA.

Klein, L. C., Whetzel, C. A., Bennett, J. M., Ritter, F. E., & Granger, D. A. (2006). Effects of caffeine and stress on salivary alpha-amylase in young men: A salivary biomarker of sympathetic activity. *Psychosomatic Medicine*, *68*(1), A-4.

Kirschbaum, C., Pirke, K. M., & Hellhammer, D. H. (1993). The Trier Social Stress Test—A tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology*, *28*, 76-81.

Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal and coping*. New York, NY: Springer Publishing.

Lieberman, H. R., & Tharion, W. J. (2002). Effects of caffeine, sleep loss, and stress on cognitive performance and mood during U. S. Navy seal training. *Psychopharmacology*, *164*, 250-261.

McLellan, T. M., Kamimori, G. H., Voss, D. M., Tate, C., & Smith, S. J. R. (2007). Caffeine effects on physical and cognitive performance during sustained operations. *Aviation, Space, and Environmental Medicine*, *78*(9), 871-877.

Nawrot, P., Jordan, S., & Eastwood, J. (2003). Effects of caffeine on human health. *Food Additives and Contaminants*, *20,* 1-30.

Smith, A. P., Clark, R., & Gallagher, J. (1999). Breakfast cereal and caffeinated coffee: Effects on working memory, attention, mood and cardiovascular function. *Physiology & Behavior*, *67,* 9-17.

Wesensten, N. J., Belenky, G., & Kautz, M. (2002). Maintaining alertness and performance during sleep deprivation: Modafinil versus caffeine. *Psychopharmacology*, *159,* 238-47.

Yerkes, F. M., & Dodson, J. D. (1908). The relationship of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology, 18*, 459-482.