Fitting a Model to Behavior Tells Us What Changes Cognitively when under Stress and with Caffeine

Frank E. Ritter[1], Sue E. Kase[1], Laura Cousino Klein[2], Jeanette Bennett[2], and Michael Schoelles[3]

[1]College of Information Sciences and Technology, Pennsylvania State University
[2]Department of Biobehavioral Health, Pennsylvania State University
[1,2]University Park, PA 16802, USA
[3]Cognitive Science Department, Rensselaer Polytechnic Institute, Troy, NY 12180, USA
frank.ritter@psu.edu, skase@ist.psu.edu, lxk18@psu.edu, jmb814@psu.edu, schoem@rpi.edu

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, and 400 mg caffeine. The data were analyzed by caffeine treatment group and how subjects appraised the task (as challenging or threatening). A cognitive model of the serial subtraction task was developed. The model was fit to the human performance data using a parallel genetic algorithm. How the model's parameters change to fit the data suggest how cognition changes due to caffeine and appraisal. Overall, the cognitive modeling and optimization results suggest that the speed of vocalization varies the most along with changes to declarative memory. This approach provides a way to compute how cognitive mechanisms change due to moderators.

Introduction

How is cognition preformed? Cognitive architectures are an approach to answer this question. How does cognition change with moderators like stress? This is a more difficult question that has kept many scientists and engineers busy.

In this paper we present a large-scale approach that starts to provide a solution to the question of how cognition changes. This approach uses methods from physiological psychology, cognitive architectures, and parallel genetic algorithms. We are able to provide an initial answer to how cognition changes due to stress in a task and due to caffeine consumed as a potential mitigator of stress.

We first describe the task that the subjects and model performed, which is a simple subtraction task. We then describe the model of this task, followed by the study methodology, the results, and a brief discussion of the results. We then describe how we fit the model to the data by varying three parameters of the cognitive architecture. How the parameters are modified give some indication of how performance was modified by stress and by caffeine. This approach has flaws and further opportunities. These are taken up after all the parts are explained.

Serial Subtraction Task

The task we used to study stress is the serial subtraction task. A brief summary of it is shown in Figure 1. It is part of the Trier Social Stressor Task (TSST), which starts with a public speaking task about an embarrassing episode or interviewing for a job (Kirschbaum, Pirke, and Hellhammer 1993). This task has been used quite often in physiology studies to cause stress in subjects, which can then be measured in a variety of ways. Typically, the subjects' performance on the task is not recorded—the task is solely used to cause changes in physiology not to give insights about cognition and stress.

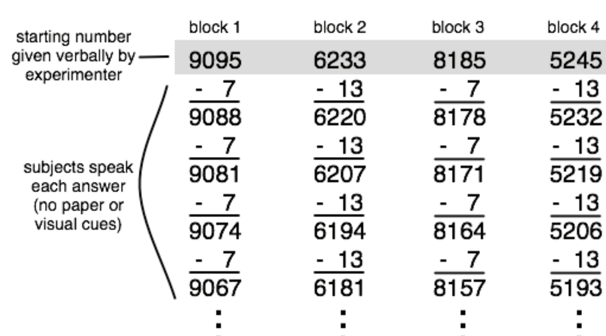

Figure 1. The serial subtraction task and the starting numbers.

The task is designed to cause physiological stress, and it routinely does as measured by changes in heart rate, blood pressure, and stress hormone levels. Before the task begins the experimenter explains that the subject's performance is going to be rated by a review panel during the lab session or is going to be recorded and analyzed. After the task is explained to the subject, a task appraisal questionnaire is completed, and the subject begins performing the task with no visual or paper clues. 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. Often, and in our case, the experimenter wears a white lab coat to increase stress. 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 is told to "Start over at <the last correct number>". At two minutes into each 4-minute session, subjects are told that "two minutes remain, you need to go faster". This prompt enhances the time-pressure component of the task.

With this task in hand, we now describe a model to perform this task.

Modeling Serial Subtraction

We created a simple model of this task. The model provides a description of how the task is performed, and will provide us with a theory of how cognition and cognitive mechanisms change to give rise to performance. In this model, theory about how mental arithmetic is performed was combined with observations gathered during a previous serial subtraction study (Ritter, Bennett, and Klein 2006) to create a cognitive model of the serial subtraction task.

The ACT-R cognitive architecture (J. R. Anderson 2007) was chosen 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. Figure 2 shows the components of the ACT-R architecture that are used.

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 approximately 650 arithmetic facts and goal-related information. The model's procedural knowledge is made up of 24 production rules that allow for retrieval of subtraction and comparison facts necessary to produce an appropriate answer. The model performs subtractions using a column-by-column strategy.

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 goal chunks types contain a state slot, the current column indicator, and the current subtrahend. The imaginal buffer maintains the current problem. This buffer is updated as the subtraction is performed. 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 left 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.

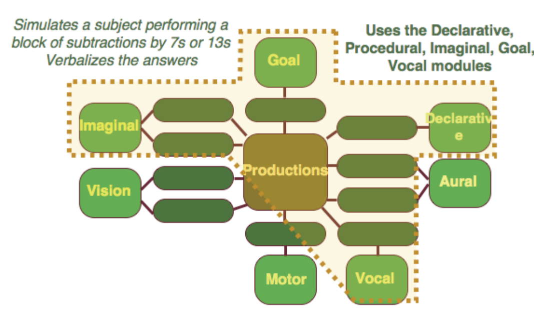

Figure 2. The components of the ACT-R architecture used by the serial subtraction model.

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.

Experimental Method

Subjects

As part of a larger project, human subject data was collected to study the effects of stress and caffeine on cardiovascular health. A mixed experimental design was conducted with 45 healthy men 18-30 years of age (Klein et al. 2006). (Men are typically used in these types of studies because we also took additional physiological measures and male systems are simpler.)

Design and Procedure

The full protocol is shown in Figure 3. After obtaining informed consent, and confirming that subjects did not have any conditions that would interact with stress and caffeine, all subjects filled out some questionnaires and were asked to perform a series of three cognitive tasks. A baseline was taken for several physiological measures (hormones from saliva, heart rate, and blood pressure). Preliminary results from these measures are reported elsewhere (Bennett et al. 2006; Klein et al. 2006; Whetzel, Ritter, and Klein 2006).

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 (serial subtraction) 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.

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 noted the four starting numbers used to begin the four blocks of subtraction during the experiment.

Task Appraisal Analysis

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. We use here the post-task appraisals because this group did not find the task threatening before they experienced the task. (Previous work typically uses pre-task appraisals, and can split subjects because the full task includes a public speaking component that is more threatening. Also, we used raw scores here rather than simply taking a median split to group subjects into two groups.)

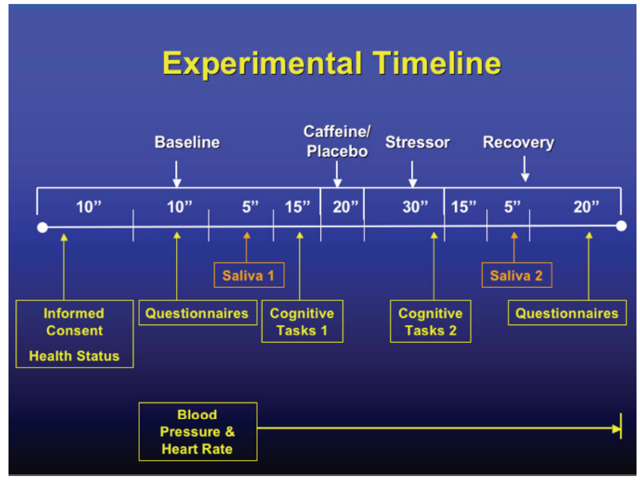

Figure 3. An illustration of the four blocks of the serial subtraction task as in the experiment.

After correcting for the imbalance in questions, a ratio of perceived stress to perceived coping resources was created (total task requirements score / total coping ability score). For example, if a subject's total appraisal score was 1 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, their perceived stress was greater than their perceived ability to cope, which equated to a *threat condition*.

Each caffeine treatment group had 15 subjects. Table 1 shows the distribution of subjects into appraisal groups. 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. This is consistent with previous results that has consistently shown an increase in self-appraised alertness (e.g., Yu et al. 1991).

Table 1. Subjects' appraisals by caffeine condition.

Caffeine Treatment		1-2 8 oz cups of coffee	3-4 8 oz cups of coffee
Number of subjects	**Placebo**	**200 mg**	**400 mg**
Challenge	7	10	2
Threat	8	5	13

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 percentage correct. The data are shown in Table 2 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 2. Human performance (average number of attempts and percent correct) by caffeine treatment group (each N=15) and appraisal condition (challenge, threat).

Treatment	Average	Challenge	Threat
PLAC	47.3, 81%	50.7, 83%	40.4, 78%
LoCAF	59.1, 86%	62.4, 88%	37.5, 75%
HiCAF	45.7, 79%	51.6, 83%	38.9, 75%

For all treatment groups the challenge condition showed the best performance in both number of attempts and percent correct across caffeine treatments. The threat condition showed the worst performance. Previous work has only found that there are fewer attempts when threatened, not that there is also lower percent correct (Tomaka, Blascovich, Kelsey, and Leitten 1993).

Performance differences between the challenge and threat conditions were most pronounced in the LoCAF group with an 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 4 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 (e.g., Haishman and Henningfield 1992; Wesensten, Belenky, and Kautz 2002). Whether a 400 mg dose is considered 'large' may be in question as some studies administered up to 800 mg doses (McLellan 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, Gabrieli, and Lieberman 1998; Smith, Clark, and Gallagher 1999). In both these cases, the across treatments and challenge plots, the effects of caffeine take on characteristics related to level of arousal studies (e.g., K. J. Anderson and 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.

Thus, we see that task appraisal correlates with performance. This might not be surprising given that the appraisal was taken after performance, but similar appraisal measures taken before also correlate, including in this task, and we know that self appraisal scores are often pretty generous in general (Dunning, Johnson, Ehrlinger, and Kruger 2003).

We also can see that caffeine dose generally provided an inverted U-shaped curve, with moderate caffeine providing the greatest number of attempts and the highest percent correct. This result was not obtained for subjects making a threatened appraisal.

These results provide differences that are interesting. The next step is to find out what changes to cognition could give rise to such differences.

Optimizing to Human Data

To understand how cognition changes for these groups, we can adjust theoretically motivated parameters in the architecture, and treat the adjustments as a description of how cognition changed. If a pattern of parameter changes that lead to better correspondences are found, they suggest how cognition changes. This process in other areas is sometimes called docking, which is an alignment procedure for comparing models (Burton, 1998; Louie et al. 2003).

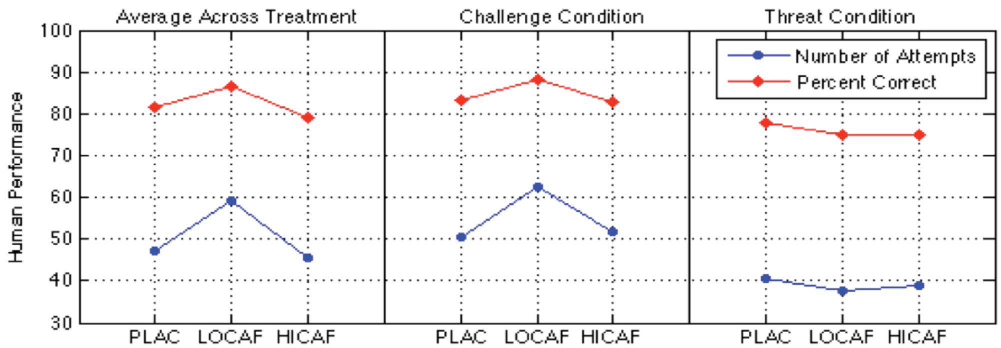

Figure 4. Comparing human performance differences in number of attempts and percent correct by treatment group (x-axis) and appraisal condition: treatment groups not accounting for appraisal (leftmost section), and averages across treatment groups divided by appraisal condition, challenge (middle section) and threat (rightmost section).

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 model fitting approach is briefly described in the second part of the section. The fitting results, accompanied by interpretations of best fitting parameter values, is discussed at the end of the section.

Architectural Parameters

Several architectural parameters in ACT-R appeared important in performing serial subtraction. We chose what we thought were the first three to explore. 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. We would, of course, like to explore a wider set. The parameters used in this study were: seconds-per-syllable, the declarative knowledge' base level constant, and the declarative memory's 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, 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

The search space for just these three parameters is large and is rather complex. Recent work with ACT-R has also shown that this fitting is to a noisy, multidimensional, non-linear, multi-parameter function. It is not an appropriate task to do by hand—a recent PhD thesis has shown that simple hill climbing does poorly (Kase 2008).

A parallel genetic algorithm was used to fit the ACT-R model to the number of attempts and percent correct data for the nine sets of data, performing a type of regression, fitting a multi-variable non-linear stochastic function (ACT-R) to multivariate data. This is a departure from the cognitive modeling community's traditional manual optimization technique. We have also fit the subjects individually, and obtained similar results (Kase 2008).

Model-to-data fit was determined by an objective function, or fitness function, defined as the sum of squared discrepancies between model performance (number of attempts and percent correct) and the corresponding human performance (e.g., $(47.3\text{-}48.1)^2 + (81.5\text{-}81.4)^2$). 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 values 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 3).

Results and Discussion

Table 3 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 over 200 runs each).

Beginning with the seconds per syllable parameter, SYL is shown in the last column and last value in the triple of Table 3. 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, concatenation of subsolutions).

Across treatments, the activation noise parameter values (ANS, first value in triple) are high 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 3 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 lower fact retrieval time.

Table 3. Optimization results for three treatment groups (PLAC, LoCAF, HiCAF) and appraisal groups (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 Perform-ance	Avg. Model Prediction	Avg. 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)

Conclusions

We have started to explore a more complete approach to studying how cognition changes under moderators like stress and caffeine. We used a fairly complex study to gather data on a task that was stressful. We fit a model to the three different caffeine treatments and appraisal groups. The fits were very close. The changes to the model to fit these data told us—based on this model and this task and this population—what changes were necessary to the cognitive mechanisms to lead to these differences in behavior.

The results suggest that there are systematic changes in cognition due to caffeine and appraisal. Most notable is the speaking rate going up when challenged and down when threatened, but declarative memory retrievals are also affected in a more complex pattern.

These changes represent the changes to cognition using this architecture as it currently exists and is commonly used. A future extension to this work would be to include a more explicit appraisal process, such as implemented by Gratch and Marsella (2004)

These results show that using a cognitive model and parametric optimization approach can further our understanding of caffeine beyond a strictly human experimentation approach. Overall, the cognitive modeling and optimization approach was successful. The preliminary modeling results and interpretations offer insight on 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 was supported by ONR grant N000140310248. Computational resources were provided by TeraGrid DAC TG-IRI070000T and run on the NCSA clusters. A more preliminary version of this report was presented at the 2009 International Conference on Cognitive modeling. Comments from Sowmyalatha Srinivasmurthy improved the presentation.

References

Amendola, C. A., Gabrieli, J. D. E., and 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 exist in the physical universe?* New York, NY: Oxford.

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

Bennett, J. M., Whetzel, C. A., Ritter, F. E., Reifers, A., and Klein, L. C. (2006). Effects of caffeine and stress on cortisol and serial subtraction performance in young healthy men. [abstract of poster presented at the annual meeting of the American Psychosomatic Society, Denver, CO]. *Psychosomatic Medicine, 68*(1), A-62.

Burton, R. (1998). Validating and docking: An overview, summary and challenge. In M. Prietula, K. Carley and L. Gasser (Eds.), *Dynamics of organizations* (pp. 215-228). Menlo Park, CA: AAAI.

Dunning, D., Johnson, K., Ehrlinger, J., and Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science, 12*(3), 83-87.

Gratch, J., and Marsella, S. (2004). A domain-independent framework for modeling emotion. *Journal of Cognitive Systems Research, 5*(4), 269-306.

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

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

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

Klein, L. C., Whetzel, C. A., Bennett, J. M., Ritter, F. E., and Granger, D. A. (2006). Effects of caffeine and stress on salivary alpha-amylase in young men: A salivary biomarker of sympathetic activity. [abstract of talk presented at the annual meeting of the American Psychosomatic Society, Denver, CO.]. *Psychosomatic Medicine, 68*(1), A-4.

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

Louie, M., Carley, K. M., Haghshenass, L., Kunz, J. C., and Levitt, R. E. (2003). Model comparisons: Docking OrgAhead and SimVision. In *NAACSOS Conference 2003*, Day 3, Electronic Publication, Pittsburgh, PA.

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

Ritter, F. E., Bennett, J., and Klein, L. C. (2006). *Serial subtraction performance in the cycling study* (Tech. Report No. 2006-1): ACS Lab, College of IST, Penn State.

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

Tomaka, J., Blascovich, J., Kelsey, R. M., and Leitten, C. L. (1993). Subjective, physiological, and behavioral effects of threat and challenge appraisal. *Journal of Personality and Social Psychology, 65*(2), 248-260.

Wesensten, N. J., Belenky, G., and Kautz, M. (2002). Maintaining alertness and performance during sleep deprivation: Modafinil versus caffeine. *Psychopharmacology, 159*, 238-247.

Whetzel, C. A., Ritter, F. E., and Klein, L. C. (2006). DHEA-S and cortisol responses to stress and caffeine in healthy young men: Is DHEA-S a reliable marker for stress? [abstract of poster presented at the annual meeting of the American Psychosomatic Society, Denver, CO] *Psychosomatic Medicine, 68*(1), A-77.

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

Yu, G., Maskray, V., Jackson, S. H. D., Swift, C. G., and Tiplady, B. (1991). A comparison of the central nervous system effects of caffeine and theophylline in elderly subjects. *British Journal of Clinical Pharmacology, 32*, 341-345.