Biologically Inspired Cognitive Architectures 22 (2017) 1–9

Contents lists available at ScienceDirect

Biologically Inspired Cognitive Architectures

journal homepage: www.elsevier.com/locate/bica

Research article

Fitting a model to behavior reveals what changes cognitively when under stress and with caffeine

Sue E. Kase[a,*], Frank E. Ritter[b], Jeanette M. Bennett[c], Laura Cousino Klein[d], Michael Schoelles[e]

[a] Computational Information Sciences Directorate, Army Research Laboratory, United States
[b] College of Information Sciences and Technology, The Pennsylvania State University, United States
[c] Department of Psychological Science, The University of North Carolina at Charlotte, United States
[d] College of Health and Human Development, The Pennsylvania State University, United States
[e] Cognitive Science Department, Rensselaer Polytechnic Institute, United States

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 individual human performance data using a parallel genetic algorithm (PGA). The best fitting parameters found by the PGA suggest how cognition changes due to caffeine and appraisal. Overall, the cognitive modeling and optimization results suggest that due to caffeine and task appraisal the speed of vocalization varies the most along with changes to declarative memory. This approach using a PGA provides a new method for computing how cognitive mechanisms change due to moderators or individual differences.

Introduction

How is cognition preformed? Cognitive architectures are an approach to answer this question (Anderson, 2007; Newell, 1990). How does cognition change with moderators such as stress? Cognitive architectures enable researchers to better understand and model human cognition, as well as extend such models to encompass cognition in stressful situations. Understanding human cognition under stress through cognitive architectures has importance implications for improving Soldier performance during modern asymmetric warfare operations (Morelli & Burton, 2009; Stetz et al., 2007). Stress is used to describe experiences that are challenging both emotionally and physiologically (Selye, 1956), as well as psychologically (Matthews, 2016). Today's network-centric battlefield environment is highly stressful and cognitively demanding. A better understanding of cognition while under stress can provide insight into how warfighters make decisions on the battlefield, especially under time-critical life-or-death situations (Kowalski-Trakofler, Vaught, & Scharf, 2003).

A large-scale computational approach is presented that begins to explore the question of how cognition changes with stress. 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 commonly used to study stress and due to caffeine consumed as a potential modulator of stress.

First, the task (a serial subtraction task), the subjects, and model are described. Then the model of the task is detailed, followed by the experiment methodology, and the results and discussion of the human study. Next, how the model was fit to the human data by varying three parameters of a cognitive architecture is explained. How the parameters varied gives some indication of how performance was modified by stress and by caffeine. This approach has flaws and further opportunities are discussed.

This section overviews the basic experimental method. The task is introduced, then the model used to predict performance on the task is described, and finally the methodological approach is explained.

Serial subtraction task

The task used to study stress is the serial subtraction task. A brief summary of the task is shown in Fig. 1. Serial subtraction is part of the Trier Social Stressor Task (TSST), which starts with a public speaking task about an embarrassing real-life episode or interviewing for a job

* Corresponding author.
E-mail addresses: sue.e.kase.civ@mail.mil (S.E. Kase), frank.ritter@psu.edu (F.E. Ritter), jbenne70@uncc.edu (J.M. Bennett), lxk18@psu.edu (L.C. Klein), schoem@rpi.edu (M. Schoelles).

http://dx.doi.org/10.1016/j.bica.2017.09.008
Received 23 September 2017; Received in revised form 24 September 2017; Accepted 24 September 2017
2212-683X/ Published by Elsevier B.V.

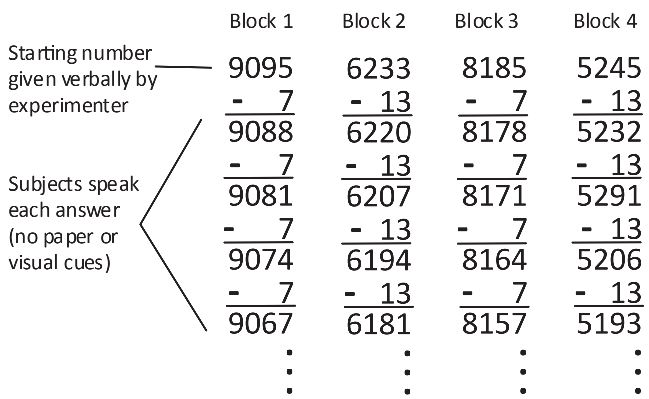

Fig. 1. An example of the serial subtraction task stimuli and the starting numbers for each block.

(Kirschbaum, Pirke, & Hellhammer, 1993). This task has been used hundreds of times in physiology studies to cause stress in subjects, which can then be measured in a variety of ways. The task is designed to cause psychosocial evaluative stress that results in physiological stress activation, and it does so routinely and reliably as measured by changes in heart rate, blood pressure, and stress hormone levels (e.g., Klein, Whetzel, Bennett, Ritter, & Granger, 2006; Kudielka, Hellhammer, Kirschbaum, Harmon-Jones, & Winkielman, 2007; Tomaka, Blascovich, Kelsey, & Leitten, 1993). Typically, the subjects' performance on the task is not recorded, and despite performance, all participants are told to improve their speed and accuracy. The task has solely been used to cause changes in physiology due to math anxiety and social comparison and has not been used to give insights about cognition and stress.

Before the task begins the experimenter explains that the subject's performance is going to be recorded and analyzed for accuracy (e.g., error type, percent correct, etc.). 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 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 of 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 the task should be performed as quickly and as accurately as possible. Often, and in our case, the experimenter wears a white lab coat to increase stress. The experimenter informs the subject of the starting number. From this point forward, the subject speaks the answer to each subtraction problem.

When an incorrect answer is given, the subject is directed to "Start over at < the last correct number > ". At two minutes into each four-minute session, subjects are told "two minutes remain, you need to go faster". This prompt enhances the time pressure component of the task. The next section describes the model developed to perform the serial subtraction task.

Modeling serial subtraction

A simple model of the serial subtraction task was developed to provide a description of how the task is performed, and contributes to a theory of how cognition and cognitive mechanisms change to give rise to performance. In the model, theory on mental arithmetic performance was combined with observations gathered during a previous serial subtraction study (Ritter, Bennett, & Klein, 2006) to create a cognitive model of the serial subtraction task.

The ACT-R cognitive architecture (Anderson, 2007) was chosen for several reasons: (1) it provides a symbolic structure in the form of a production system and a parameter-driven level of processing using a number of mathematical equations; (2) it permits the parallel execution of the verbal system with the control and memory systems; and (3) it has been used for other models of mathematical processing developed by other researchers (Anderson, 2005; Dehaene, Piazza, Pinel, & Cohen, 2003; Lebiere, 1999; Ravizza, Anderson, & Carter, 2008; Rosenberg-Lee, Lovett, & Anderson, 2009). According to Rosenberg-Lee et al. (2009), "The ACT-R cognitive architecture proposes that cognition is accomplished by the activity of independent modules that are co-ordinated by a production system. Modules represent various perceptual and motor modalities, such as vision and finger manipulation, and aspects of central cognition, such as retrieving memories, cognitive control, and the maintenance of internal representations." Fig. 2 shows the primary components of the ACT-R architecture. The dashed line represents the components used in the serial subtraction model.

The serial subtraction model performs a block of subtracting by 7 s or 13 s 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. Table 1 shows several example arithmetic facts and a production rule used in the serial subtraction model. The declarative memory of the serial subtraction model is loaded with integer, addition, subtraction, multiplication, and comparison facts. A 'chunk-type' defines the structure for each type of fact. The top section of Table 1 shows example chunk types and several integer, addition and subtraction facts. The bottom section of the table shows an example production rule which is part of the borrowing operation when performing subtraction. If the fields in the top part of the production (above the = = > symbol) match the current state of the model's buffers, the production executes the lower statements (below the = = > symbol) which can modify the buffers.

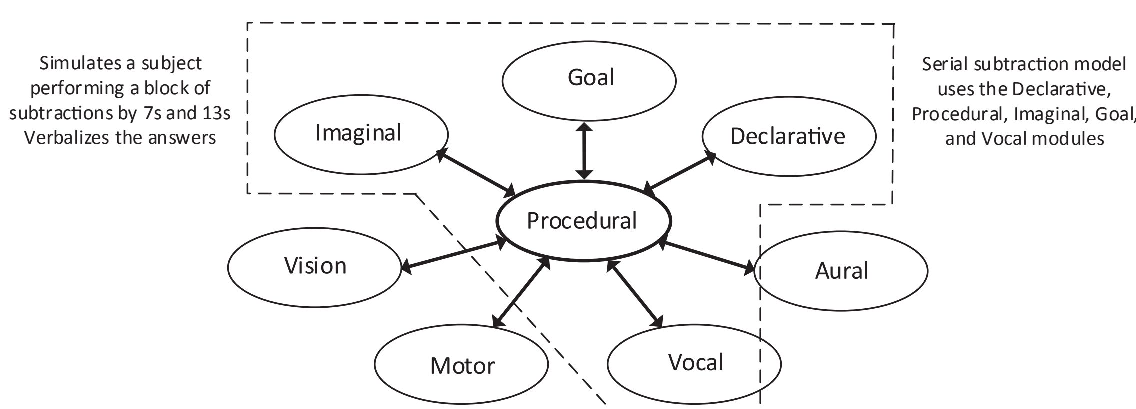

Fig. 2. The primary components of the ACT-R architecture. The components within the dashed line are used by the serial subtraction model.

Table 1
Example declarative memory chunk-types and facts (Top); example production rule from part of the borrowing operation in the serial subtraction model (Bottom).

Declarative memory consists of facts defined by chunk-types

```
(chunk-type integer value)
(chunk-type addition-fact arg1 arg2 sum)
(chunk-type subtraction-fact arg1 arg2 diff)
(ZERO ISA INTEGER VALUE 0)
(ONE ISA INTEGER VALUE 1)
(TWO ISA INTEGER VALUE 2)
(FOUR + SEVEN ISA ADDITION-FACT ARG1 FOUR ARG2 SEVEN SUM ELEVEN)
(FOUR + EIGHT ISA ADDITION-FACT ARG1 FOUR ARG2 EIGHT SUM TWELVE)
(FOUR + NINE ISA ADDITION-FACT ARG1 FOUR ARG2 NINE SUM THIRTEEN)
(NINE + FOUR ISA SUBTRACTION-FACT ARG1 NINE ARG2 FOUR DIFF FIVE)
(NINE + FIVE ISA SUBTRACTION-FACT ARG1 NINE ARG2 FIVE DIFF FOUR)
(NINE + SIX ISA SUBTRACTION-FACT ARG1 NINE ARG2 SIX DIFF THREE)
```

Procedural memory consists of productions representing knowledge

```
(p* borrow-move-right
"Positioned at new column, val is not zero, subtract one to borrow"
  =goal > isa borrow
    state move-right-to-borrow
  =retrieval > isa next-column
    next-col = new-col
  =imaginal > isa subtract-problem
    =new-col = val
  - =new-col zero
==>
 !output!(at new column = new-col need to borrow from = val)
  =goal > state reduce-minuend
    current-col = new-col
  +retrieval > isa subtraction-fact
    arg1 = val
    arg2 one)
```

The serial subtraction model executes in ACT-R 6.0 and uses 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 chunk 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 four-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 (Lebiere, 1999; Rosenberg-Lee et al., 2009).

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; therefore, the borrow and the subgoal return back to the main task goal. If there is a next column and its value is not 0, then 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 four-digit result. The model has two output strategies. In this investigation, 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.

Experimental method

Subjects

As part of a larger project, human subject data were collected to study the effects of stress and caffeine on cognition. A mixed experimental design was conducted with 45 healthy men 18–30 years of age who consumed caffeine daily (Klein et al., 2006). Due to known sex differences in caffeine metabolism and cortisol responses to stress (Benowitz, Lessov-Schlaggar, Swan, & Jacob, 2006; Kirschbaum, Kudielka, Gaab, Schommer, & Hellhammer, 1999), this initial investigation was limited to males.

Design and procedure

The full protocol is shown in Fig. 3. After obtaining informed consent, and confirming the subjects did not have any health or medical

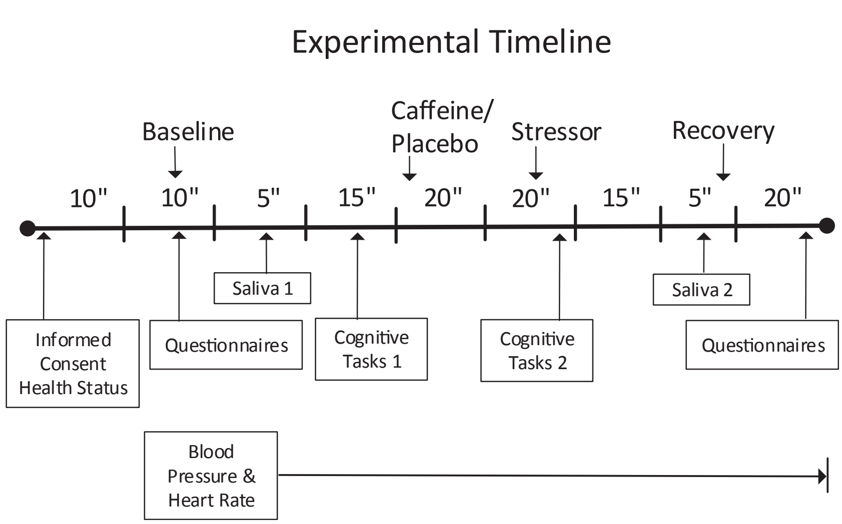

Fig. 3. The experimental protocol, including an illustration of the four blocks of the serial subtraction task labeled 'Stressor'.

conditions that would interact with stress and caffeine, all subjects completed several questionnaires and were asked to perform a series of three cognitive tasks. A baseline was taken for several physiological measures (i.e., hormones from saliva, heart rate [HR], and blood pressure [BP]). Preliminary results from these measures are reported elsewhere (Bennett, Whetzel, Ritter, Reifers, & Klein, 2006; Klein et al., 2006; Whetzel, Ritter, & Klein, 2006).

Subjects individually performed a simple reaction time (RT) and a working memory (WM) task, which required 15 min 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-min stress session of the mental arithmetic (e.g., serial subtraction) portion of the TSST was performed. Following completion of this stressor, subjects again were asked to complete the same RT and WM tasks. Cognitive performance was determined by calculating accuracy and response time scores.

The serial subtraction task used in the experiment consisted of four four-minute blocks of mentally subtracting by 7 s and 13 s from four-digit starting numbers. Fig. 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 answered five questions: 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. The post-task appraisals were analyzed in this case because subjects reported that the task was not threatening prior to completing the task.

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 less than or equal to one, 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 one, 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 2 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 two challenged subjects with the remaining (13) subjects reporting a threatening appraisal. These differences are reliable, χ^2 (1, 2) = 8.92, p = .012. This is consistent with previous results that show an increase in self-appraised alertness for moderate but not high caffeine levels (Brice & Smith, 2001; Yu, Maskray, Jackson, Swift, & Tiplady, 1991). It is understood that subjective alertness does not always provide an objective measure of alertness. A subjective increase alertness may result in the stressful task being appraised as a challenge, then as a threat. On the other hand, high caffeine levels may induce anxiety which would lead to more threatening appraisals (Brice & Smith, 2001; Winston, Hardwick, & Jaberi, 2005; Yu et al., 1991).

Table 2
Subjects' post-task 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

Table 3
Human performance (average number of attempts and percent correct) by caffeine treatment group (each N = 15) and appraisal condition (challenge PLAC N = 7, LoCAF N = 10, HiCAF N = 2; threat PLAC N = 8, LoCAF N = 5, HiCAF N = 13).

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%

Human experimental 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 3 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 7 s.

Across all treatment groups, the subjects in the challenge condition tended to demonstrate some of the best performance in both number of attempts and percent correct. Subjects who perceived the task as threatening, demonstrated some of the worst performance. Previous work has only found that subjects make fewer attempts when threatened, not that there is also lower percent correct (Tomaka et al., 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.

Fig. 4 visualizes 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). With limited sample sizes, what might appear as a large difference within the plotted data does not take into account some of the variance in the samples. The plotted data should be viewed from the perspective of a proof-of-concept which visualizes potential trends; however, a larger dataset is needed to test statistical significance.

Fig. 4 plot reveals 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 & 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 (McLellan, Kamimori, Voss, Tate, & Smith, 2007). Generally, 100–300 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, & Lieberman, 1998; Smith,

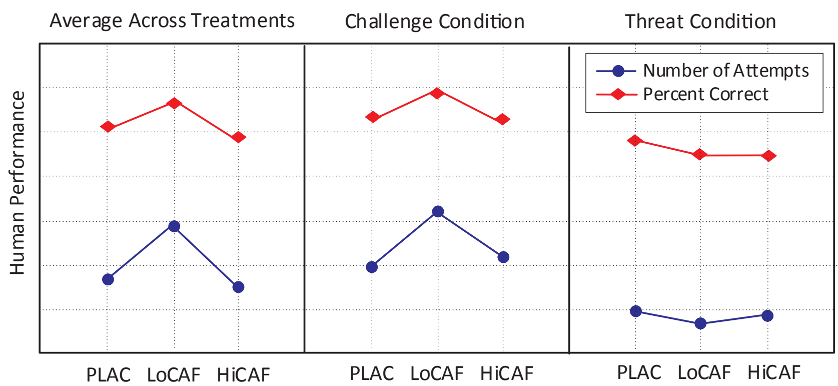

Fig. 4. Comparing human performance differences in number of attempts and percent correct by treatment group (x-axis) and appraisal condition: treatment groups not accounting for appraisal (leftmost section), and averages across treatment groups divided by appraisal condition, challenge (middle section)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

Clark, & Gallagher, 1999). In both these cases, across treatments and challenge plots, the effects of caffeine mirror trends previously published in arousal literature (e.g., Anderson & Revelle, 1982) and appear to support the Yerkes and Dodson (1908) law that postulates 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.

In summary, task appraisal appears to be associated 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 self-appraisal scores are often generous in general; that is, people tend to hold overinflated views of their skills (e.g., Dunning, Johnson, Ehrlinger, & Kruger, 2003).

Caffeine dose generally provided an inverted U-shaped curve, with moderate caffeine providing the greatest number of attempts and the highest percent correct for challenged subjects. Similar performance was not obtained for subjects appraising the task as threatening. Also, changes in post-task appraisal with the same shape occur with moderate caffeine more often with a challenging appraisal.

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

Understanding changes under stress

To understand how cognition changes for these groups, we can adjust theoretically motivated parameters in a cognitive architecture, and treat the adjustments as a description of how cognition changed. If patterns 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, Carley, Haghshenass, Kunz, & Levitt, 2003). 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 based on parallel generic algorithms that are described in the next several subsections. The fitting results, accompanied by interpretations of best fitting parameter values, are discussed at the end of this section.

Architectural parameters

Several architectural parameters in ACT-R appeared important in performing serial subtraction. We selected what we thought were the three most task-relevant parameters to explore out of more than 80 parameters available in the ACT-R architecture version 6.0. A speech parameter was chosen for optimization because vocalization of the answer is the most time-consuming aspect of this task. Two memory 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. One would, of course, like to explore a wider set, but to illustrate and start to explore this model fitting approach, three task specific parameters were selected. The parameters used in model fitting were: seconds-per-syllable, the declarative knowledge base level memory activation constant, and the declarative memory activation noise.

Because subjects perform the serial subtraction task by speaking the answer to each subtraction problem, the seconds-per-syllable (SYL) parameter was important to investigate. SYL controls the rate the serial subtraction model speaks. The ACT-R default timing for speech is 0.15 s per 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 the cognitive 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.

Experimentation involving mental arithmetic investigates the mental representation of numbers and arithmetic facts (counting, addition, subtraction, multiplication) and the processes that create, access, and manipulate them. The activation of chunks storing arithmetic facts in declarative memory is critical to ACT-R's performance in mental arithmetic (Lebiere, 1999). Two parameters affecting declarative knowledge access are 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 the error rate for retrievals in the model. It was thought a stressful task such as serial subtraction would show a large amount of variability in ANS. Also, the ANS value can represent subjects' variance from trial to trial. Other parameters, such as base level learning, decay, and the characters-per-syllable parameters were left fixed at their default values for this study.

The search space for the model optimization was defined by parameter value boundaries. The ACT-R website (http://act-r.psy.cmu.edu/) is an online repository of ACT-R studies where both the manuscripts and the models are available broadly categorized by type of task. Information was collected on models related to cognitive arithmetic and mathematical problem solving and which particular parameters and values were being used. The commonly found parameter value boundaries for ANS, SYL, and BLC where then expanded for the optimization to ensure all plausible combinations of values in performance

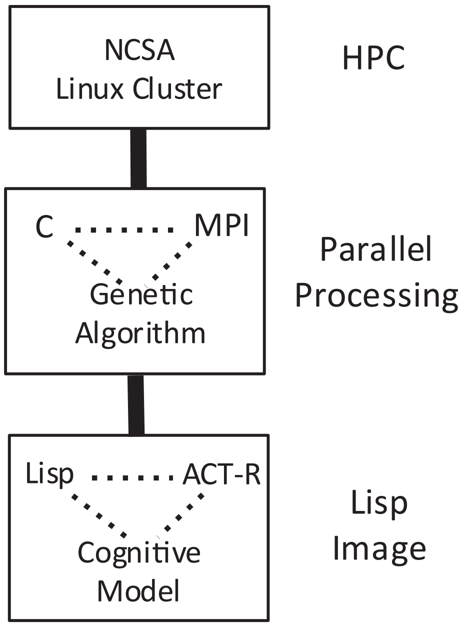

Fig. 5. Components of the optimization platform on a HPC Linux cluster.

of the serial subtraction task could be tested. The parameter value boundaries selected were: ANS and SYL 0.1–0.9, and BLC 0.1–3.0.

Optimization approach

The search space for these three parameters is large and rather complex. Recent work with ACT-R has also shown this fitting is to a noisy, multidimensional, non-linear, multi-parameter function. It is not an appropriate task for manual optimization. For example, Kase (2008) has shown that simple hill climbing performs poorly in optimization.

Genetic algorithms (GAs) have been used to optimize the fit of functions to data for a long time (e.g., Davis & Ritter, 1987; Goldberg, 1989). GAs use a set of genotypes, the genes for a solution, and compute the fitness of a phenotype, a solution created by the genotype. In cognitive modeling, a genotype could represent a set of parameters. The phenotype is the predictions of a model arising from those parameters. Better evaluated phenotypes are more likely to have their genes passed to a later generation.

The first model used for a similar purpose improved the speed of a neural network to find XOR (Ritter, 1991). Later work by Lane and Gobet (2005) has suggested GAs can be helpful in this area. Further work by Peebles (2016) has also shown this approach of using GAs to understand and improve model fit can be a productive approach to understanding models and their relationship to data.

In this study, a parallel genetic algorithm (PGA) was applied to overcome the combinatorial parameter search spaces and substantial computational and time resources associated with optimizing the ACT-R model to the human performance on the serial subtraction task. This is a stochastic global optimization problem where it is reasonable to assume multiple local optima exist. In general, the goal of global optimization is to find a point for which the object function obtains its smallest value, the global minimum. Genetic algorithms have the advantage of being less susceptible to getting stuck at local optima than a gradient search method.

The genotypes were composed of ACT-R parameters sets applied to the cognitive model. The population evolved to find better ‘solutions’ by selecting the most fit parameter sets (those that give the best match to the human data), and propagating these solutions to the next generation. In the algorithm, the fitness evaluation consisted of running the model in the ACT-R cognitive architecture, analyzing the model’s performance output, and calculating a fitness evaluation based on the match of the predictions to human behavior using multiple processors that reduced the time required to reach acceptable solutions.

The PGA used a fitness evaluation of the number of attempts and percent correct performance for the nine sets of data, executing 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 using this approach, and obtained similar results (Kase, 2008).

Parallel implementation

There are several classes of PGAs distinguished by their level of parallelization. In this study, the algorithm used is a master-slave global parallelization PGA. This type of PGA is characterized by a high computation to communication ratio. In a master-slave PGA, one master-processing node executes the GA-related functions, while the fitness evaluation is distributed among numerous slave processors. The slave processors evaluate the fitness of the genotypes that they receive from the master process, and then return the fitness results back to the master node.

The ACT-R architecture and cognitive model are written in Lisp. Generally, MPI (i.e., multiprocessor coordination libraries) is available on cluster computing resources in the form of C or Fortran libraries. To utilize parallel processing in the cognitive model optimization process, ACT-R and the cognitive model are packaged into an executable Lisp image or core file. This image file can be run by a system call from a C program on each processor in parallel while using MPI to communicate genotypes and fitness values among the processors. Fig. 5 illustrates the optimization platform on a high-performance computing (HPC) cluster.

The population of genotypes (ACT-R parameter sets), in the form of a matrix, are ‘scattered’ row-wise to the processors. Each processor executes the Lisp image file that runs the model within the ACT-R architecture. Each processor then calculates a fitness based on the model’s performance predictions and the human data statistics. In this case, sum of the squares error is calculated on both number of attempts and the percent correct from a block of serial subtractions by 7 s. The fitness values calculated by the processors are ‘gathered’ up by the master process, which then applies genetic functions to the population based on the fitness of the genotypes. This is repeated through a number of generations with the effect of evolving a set of candidate solutions.

Optimization setup

Nine PGAs were set to run 100 generations of 200 binary-encoded genotypes. Each PGA optimized the serial subtraction model to the subjects’ group performance data (treatment by appraisal) gathered from the experiment. The PGA used genotypes consisting of one 36-bit chromosome divided into three 12-bit substrings each representing the value of the three ACT-R parameters ANS, BLC, and SYL.

In the PGA code the selection probability (selection of the fittest) was set to 0.5, meaning half the genotype population is replaced each generation by offspring of the fittest genotypes. Random mutations alter a certain percentage of the bits in the list of chromosomes. This operation introduces new traits in the original population and keeps the PGA from converging too quickly before sampling the entire search space. The mutation rate was set at 0.15. The terminating condition was a specified number of generations (1 0 0), instead of proximity to each subject’s performance statistics.

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., $(47.3 - 48.1)^2 + (81.5 - 81.4)^2$). 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, and values less than 1.0 represent a fit less than 1

Table 4
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) in a four-minute block, and fitness value associated with average (over N) of best fitting (less than 1.0 fitness value) ACT-R parameter values. The ACT-R parameters are ANS, activation noise in declarative memory; BLC, base level constant activation of declarative memory; and SYL, seconds per syllable speaking rate).

	Human performance	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	0.58	0.67, 2.35, 0.57 (4)

difference in number of attempts and percent correct.

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 this approach, the model was optimized to the nine sets of human performance data in Table 3.

Results of the PGA

Table 4 is a summary of the PGA optimization results by caffeine treatment and appraisal condition. The first column denotes the appraisal condition with CH for challenge, TH for threat, and ALL for the average across challenge and threat. The next two columns, Human Performance and Avg. Model Prediction, list the number of attempts (first value) and percent correct (second value) for the human (second column) and the model (third column). The model's performance is an average across the number of best fitting parameter sets. For example, the (3) in the first row, last column means the PGA found three parameter sets producing fitness less than 1.0 and that these parameter values (last column), model predictions (third column), and fitness value (fourth column) are averaged across those three best PGA runs. The fitness value column shows the PGA optimizations were able to find good solutions for the three treatment groups and appraisal conditions within a fractional part of a subtraction problem. Considering the complexity of the serial subtraction task and the human performance variability, these are exceptional model to human data fits that suggest how cognition changed.

Several trends can be observed within the parameter values producing best fits. Beginning with the seconds-per-syllable parameter, SYL is shown in the last column and last value in the triple of Table 4. 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 s per syllable (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 surrogate for other 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 occurrence could be due to the nature of the task as more stressful than typically found in psychology experiments (i.e., purposively used to elicit a stress response). The ANS value range in Table 4 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) than 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.

Discussion and conclusion

This investigation started to explore a more complete approach for studying how cognition when moderators such as stress and caffeine are considered. A fairly complex experimental protocol was used to collect data on a task that has been used in previous stress research.

A cognitive model was fit to three different caffeine treatments and appraisal groups using a PGA optimization approach. The fits were very close revealing several patterns in the parameter values. The changes to the model to fit these data increased our understanding of the cognitive mechanisms that lead to these differences in behavior. The average fitness values in Table 4 suggest the model can be well fit by this approach of identifying parameters using a PGA. The results suggest there are systematic changes in cognition due to caffeine and appraisal—most notable talking faster when challenged and slower when threatened, and slightly more noise in declarative memory processes and less basic activation of declarative memory. We also are able to see an inverted U-shaped curve in these values as might be expected with these caffeine doses. These changes represent the changes to cognition using this architecture as it currently exists and is commonly used. It is possible that the changes revealed in this study may be surrogates for other changes, but the cognitive model and PGA optimization begin to illustrate how this approach can help summarize cognitions that may lead to a particular behavior when factoring in stress and task appraisal.

This investigation also shows 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.

Several extensions to this research can be envisioned. The current serial subtraction model simulates a stressed subject by manipulating ACT-R architectural parameters. An alternative representation of particular theories of stress or anxiety could be implemented using subsets of ACT-R production rules executed as the model runs. This would be considered more of a knowledge representation simulation of stress instead of a parametric overlay representation. A study by Miwa and Simon (1993) suggested a knowledge representation alternative to parametric modeling of individual differences by presenting a model that could make a distinction between common skills and individual differences. Miwa and Simon used the example that 70% of the subject's behavior could be explained by the common part of the model, and the remaining 30% by the individual part of the model.

The current serial subtraction model does not possess self-evaluative mechanisms. In other words, the model does not know whether it is performing poorly or not. Audio file transcriptions of subjects' performance confirmed when a subject made an error, especially consecutive errors, the subject would become discouraged with his performance and appear to intentionally slow down his mental step-wise calculations of the subtraction problem in order to answer correctly. Similar to Miwa and Simon's model representation, a production or set of productions representing appraisal feedback could be added to the serial subtraction model executing when the model detects it has made an error. These extra productions could 'waste time' in simulating the subject slowing down his mental calculations; or execute internal verbalizations about doing poorly and feeling embarrassed, thus, interfering with the vocal module's speaking of the subtraction answer. This would be supportive of what was observed in the pattern of SYL values.

Another extension to consider is the type of PGA used in the optimization process. This study implemented a master-slave GA with a single global population. The global population consisted of one large search space defined by the three ACT-R parameters (ANS, BLC, SYL) with minimum and maximum values of each parameter constraining the space. The next level up in complexity is a multiple-population GA which consists of several subpopulations that exchange individuals occasionally. The exchange of individuals is called migration and is controlled parametrically. Each sub-population can identify a different locally-optimal solution. The populations send emigrants that have the effect of attracting the other subpopulations to their solutions possibly crossing valleys of low fitness that would have remained unexplored otherwise. This additional exploration may discover even better solutions.

In the results, the patterns of ANS and BLC value pairs were not easily interpretable in reference to theories of stress and anxiety. Subpopulations of a multiple-population PGA could be set up to divide the ACT-R parameter space for searching. The divisions could be based on different hypotheses about cognitive performance on the task. A processor would be allocated for each genotype as in the global population GA, but the genetic operators would be applied to each subpopulation separately. After the GA terminates, the solutions from each subpopulation could be compared and evaluated in reference to theories of stress and anxiety. The search space defined by the best subpopulation could then be used in a global population PGA for a more fine-grained search.

In general, the multiple-population GA would make for more efficient theory development as one run of the GA could focus search on several different regions thought to be theoretically feasible. Other alternatives for use of a multiple-population GA include subpopulations based on the use of different fitness functions, or subpopulations each running a different version of the model. As discussed above, a base model with a primary set of productions, and then several other versions of the base model with different additional production sets. The 'family' of models could be optimized in one run of the PGA and then compared.

From an educational perspective, two extensions of this research come to mind; formulating a parameter pattern library, and developing a more flexible and reusable model optimization platform. Broadly speaking, the PGA optimization technique is efficient enough to be employed as an educational tool in learning about a cognitive architecture and how different combinations of parameters influence specific cognitive processes. The ACT-R architecture has approximately 80 different parameters available for manipulating a model's behavior. With the PGA optimization, any sized search space could be explored using many different sets of parameter combinations. The results of each search could be overlaid to interpret the effects of each combination of parameters.

With the search space and parameter combinations held constant, different models could be used during each search. If tasks being modeled were broadly classified into categories, libraries or repositories could be set up to collect patterns of parameter combinations with their resulting effects on specific cognitive processes, possibly ordered by architectural module. Patterns of predictions could be submitted by researchers in the cognitive modeling community to a central library located on the architecture's website forming a collective pattern language of architectural parameter combinations when applied to different classes of tasks.

This PGA optimization approach is available and extendable by other modelers. A large cluster computing resource was used for this project. Most medium to large sized universities have some type of cluster computing resource. With the exception of the serial subtraction model all other applications used in this project are open source (CMUCL, ACT-R) or commonly reside on a cluster (C, MPI). The optimization approach could be used to fit other models besides the serial subtraction model. In theory, any cognitive model written in the ACT-R architecture could be modified in the same way as the serial subtraction model to run in a parallel processing environment. The primary modifications required included reformatting the model's startup function to accept parameter values coming in as arguments, and rewriting the backend function to support the fitness criteria of the PGA.

Lastly, different types of search algorithms could be applied, even in combination, to find the best model-to-data fits. For this project, the majority of the search algorithm code was written separately from the cognitive model interfacing with the model only in the fitness function. For example, the basic code for the PGA was extracted from a textbook and then modified to incorporate the cognitive architecture and model. With the availability of open source applications and academic resources, and easy integration of a cognitive architecture and model, this optimization approach can be adopted for use by other cognitive modeling communities using different architectures.

In conclusion, several earlier studies mentioned in section 'Understanding changes under stress' (Lane & Gobet, 2005; Peebles, 2016; Ritter, 1991) formulated the ground work for this PGA-based global optimization approach for fitting parameterized cognitive models to human data. This line of research has not been actively pursued by the cognitive modeling community, but should have been as this study points out. The results reported here show our optimization approach was successful in producing excellent model to human data fits and shows promise for replacing the cognitive modeling community's traditional manual optimization technique—an iterative step-by-step process that encourages modeler bias in selecting parameters values that support a chosen hypothesis.

Acknowledgments

This project was supported by ONR grants N000140310248 and N00014-15-1-2275. Computational resources were provided by TeraGrid DAC TG-IRI070000T and run on the NCSA clusters. Partial results from this investigation were presented at the 2009 AAAI Fall Symposium Series on Biologically Inspired Cognitive Architectures.

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. (2005). Human symbol manipulation within an integrated cognitive architecture. *Cognitive Science, 29*, 313–341.
Anderson, J. R. (2007). *How can the human mind exist 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.
Bennett, J. M., Whetzel, C. A., Ritter, F. E., Reifers, A., & Klein, L. C. (2006). Effects of caffeine and stress on cortisol and serial subtraction performance in young healthy men. [Poster presented at the annual meeting of the American Psychosomatic Society, Denver, CO]. *Psychosomatic Medicine, 68*(1), A62.
Benowitz, N. L., Lessov-Schlaggar, C. N., Swan, G. E., & Jacob, P. (2006). Female sex and oral contraceptive use accelerate nicotine metabolism. *Clinical Pharmacology & Therapeutics, 79*(5), 480–488.
Brice, C., & Smith, A. P. (2001). The effects of caffeine on simulated driving, subjective alertness and sustained attention. *Human Psychopharmacology, 16*, 523–531.
Burton, R. (1998). Validating and docking: An overview, summary and challenge. In M. Prietula, K. Carley, & L. Gasser (Eds.). *Dynamics of organizations* (pp. 215–228). Menlo Park, CA: AAAI.
Davis, L. W., & Ritter, F. (1987). Schedule optimization with probabilistic search. *The third IEEE computer society conference on artificial intelligence applications* (pp. 231–236). Washington, DC: IEEE Press.
Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology, 20*, 487–506.
Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science, 12*(3), 83–87.
Goldberg, D. E. (1989). *Genetic algorithms in search, optimization, and machine learning.* Reading, MA: Addison-Wesley.
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) *Parallel genetic algorithm optimization of a cognitive model: Investigating group and individual performance on a math stressor task* (Unpublished doctoral thesis). College of Information Sciences and Technology, The Pennsylvania State University, University Park, PA.
Kirschbaum, C., Kudielka, B. M., Gaab, J., Schommer, N. C., & Hellhammer, D. H. (1999). Impact of gender, menstrual cycle phase, and oral contraceptives on the activity of the hypothalamus-pituitary-adrenal axis. *Psychosomatic Medicine, 61*(2), 154–162.
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.
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. [Abstract of talk presented at the annual meeting of the American Psychosomatic Society, Denver, CO.]. *Psychosomatic Medicine, 68*(1), A4.
Kowalski-Trakofler, K. M., Vaught, C., & Scharf, T. (2003). Judgment and decision making under stress: An overview for emergency managers. *International Journal of Emergency Management, 1*, 278–289.
Kudielka, B. M., Hellhammer, D. H., Kirschbaum, C., Harmon-Jones, E., & Winkielman, P. (2007). Ten years of research with the Trier Social Stress Test—Revisited. *Social Neuroscience: Integrating Biological and Psychological Explanations of Social Behavior*, 56–83.
Lane, P. C. R., & Gobet, F. (2005). Discovering predictive variables when evolving cognitive models. *Proceedings of the pattern recognition and data mining, third international conference on advances in pattern recognition, ICAPR 2005, bath, UK, August 22–25, 2005, part I* (pp. 108–117). Heidelberg, Germany: Springer.
Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal and coping.* New York: Springer Publishing.
Lebiere, C. (1999). The dynamics of cognition: An ACT-R model of cognitive arithmetic. *Kognitionswissenschaft, 8*(1), 5–19.
Louie, M., Carley, K. M., Haghshenass, L., Kunz, J. C., & Levitt, R. E. (2003). Model comparisons: Docking OrgAhead and SimVision. In *NAACSOS conference 2003, Day 3*. Pittsburgh, PA: Electronic Publication.
Matthews, G. (2016). Multidimensional profiling of task stress states for human factors: A brief review. *Human factors, 58*(6), 801–813.
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.
Miwa, K., & Simon, H. A. (1993). Production system modeling to represent individual differences: tradeoff between simplicity and accuracy in simulation of human behavior. In A. Sloman, D. Hogg, G. Humphreys, A. Ramsay, & D. Partidge (Eds.). *Prospects for artificial intelligence* (pp. 158–167). Amsterdam, Netherlands: Kaigai Publications.
Morelli, F., & Burton, P. A. (2009). The impact of induced stress upon selective attention in multiple object tracking. *Military Psychology, 21*, 81–97.
Newell, A. (1990). *Unified theories of cognition.* Cambridge, MA: Harvard University Press.
Peebles, D. (2016). Two methods for optimizing cognitive model parameters. In *Presentation at the ACT-R post-graduate summer school.*
Ravizza, S. M., Anderson, J. R., & Carter, C. S. (2008). Errors of mathematical processing: The relationship of accuracy to neural regions associated with retrieval or representation of the problem state. *Brian Research, 2138*, 118–126.
Ritter, F. E. (1991). Towards fair comparisons of connectionist algorithms through automatically generated parameter sets. *Proceedings of the 13th annual conference of the cognitive science society* (pp. 877–881). Hillsdale, NJ: Erlbaum.
Ritter, F. E., Bennett, J., & Klein, L. C. (2006). Serial subtraction performance in the cycling study (Tech. Report No. 2006–1): Applied Cognitive Science Lab, College of Information Sciences and Technology, Penn State.
Rosenberg-Lee, M., Lovett, M. C., & Anderson, J. R. (2009). Neural correlates of arithmetic calculation strategies. *Cognitive, Affective, & Behavioral Neuroscience, 9*(3), 270–285.
Selye, H. (1956). *The stress of life.* New York, NY: McGraw-Hill.
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.
Stetz, M. C., Thomas, M. L., Russo, M. B., Stetz, T. A., Wildzunas, R. M., McDonald, J. J., ... Romano, J. A. (2007). Stress, mental health, and cognition: A brief review of relationships and countermeasures. *Aviation, Space, and Environmental Medicine, 78*(5), B252–B260.
Tomaka, J., Blascovich, J., Kelsey, R. M., & 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., & Kautz, M. (2002). Maintaining alertness and performance during sleep deprivation: Modafinil versus caffeine. *Psychopharmacology (Berl), 159*, 238–247.
Whetzel, C. A., Ritter, F. E., & 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? [Poster presented at the annual meeting of the American Psychosomatic Society, Denver, CO]. *Psychosomatic Medicine, 68*(1), A77.
Winston, A. P., Hardwick, E., & Jaberi, N. (2005). Neuropsychiatric effects of caffeine. *Advances in Psychiatric Treatment, 11*, 432–439.
Yerkes, R. M., & 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., & Tiplady, B. (1991). A comparison of the central nervous system effects of caffeine and theophylline in elderly subjects. *British Journal of Clinical Pharmacology, 32*(32), 341–345.