A Sex Difference in the Effect of Low Levels of Caffeine on the Stroop Task

Jeremy C. Owens
John Broida*
University of Southern Maine

Low levels of caffeine may decrease performance on complex cognitive tasks. To test this hypothesis 63 psychology students participated in a study examining the effects of 45 mg of caffeine on the Stroop task. The participants consumed either a 12-oz (355-ml) can of Coca-Cola or a 12-oz (355-ml) can of caffeine-free Coca-Cola. A short questionnaire was then answered by the participants, after which they completed a computerized version of the Stroop task. We observed that women had slower reaction times than men (p < .001), and that the effects of caffeine were sex specific, having a significant effect only on men (p < .01). We also found a significant main effect of display (p < .001), and an interaction between sex and display (p < .05), which occurred only in the conflict condition of the Stroop task. These sex differences may reflect the lipid-soluble nature of caffeine. The decreased performance in the caffeine group can be explained by the Yerkes-Dodson law.

Caffeine is the most widely used behaviorally active drug in the world, being found in coffee, tea, cola, and cocoa (Julien, 1995). The daily consumption of caffeine in the United States is 200 mg or more per person (Palfai & Jankiewicz, 1997). These high levels of caffeine consumption have produced interest in the effects of caffeine on various behaviors, including cognition.

Caffeine may have positive effects on cognitive performance (Hasenfratz & Baettig, 1994). For example, Jarvis (1993) showed that as caffeine consumption increased, reaction times and choice reaction times decreased. Jarvis also found that the number of recalled incidental verbal memory items increased and accuracy on visuo-spatial reasoning increased as caffeine consumption increased. This study used questionnaires to measure daily caffeine consumption of the participants rather than actually manipulating caffeine levels.

Smith, Brockman, Flynn, Maben, and Thomas (1993) administered 3 mg of caffeine per kg of body weight and examined its effects on alertness ratings, a variable four-period simple reaction time task, a five-choice serial response task, a semantic processing test, and a logical reasoning task. They found that caffeine (a) increased alertness ratings and the number of correct responses on the five-choice serial response task and (b) decreased reaction times on the four-period simple reaction time task and the semantic processing task. Lorist, Snel, Kok, and Mulder (1994) also found that 250 mg of caffeine reduced reaction times on low memory load and high memory load selective attention tasks.

In contrast, other studies suggest that caffeine may impair performance. Terry and Phifer (1986) found that 100 mg of caffeine caused participants to recall fewer words on the Auditory Verbal Learning Test (AVLT). Similarly, Jacobson, Winter-Roberts, and Gemmell (1991) found that 5 mg of caffeine per kg of body weight significantly increased reaction times to a tweezer pin placement dexterity task.

These contradictory research findings may be explained by the Yerkes-Dodson model of arousal. The Yerkes-Dodson law states that performance is an inverted-U function of arousal and that there is a negative association between optimal arousal and task complexity (Anderson, 1994). This model indicates that complex tasks have lower optimal arousal levels for peak performance than simple tasks. In addition to dose differences, the difference in the complexity of the tasks used to measure performance may explain these apparently contradictory results.
Other studies have also shown sex differences in the effects of caffeine (James, 1990). Smith, Davidson, and Green (1993) found that women were less affected by caffeine, recalling more items than men on a recall task after consuming 45 mg to 300 mg of caffeine. James (1990) found that 6 mg of caffeine per kg of body weight caused more hand tremor in men than women on a hand steadiness task. These sex differences may be explained by the lipid-soluble nature of caffeine. Women on the average have higher levels of body fat than men; therefore, the effect of caffeine would be expected to be more pronounced in men and more sustained in women. This differential is expected because body fat absorbs lipid-soluble drugs, reducing levels that get to the brain. Furthermore, as lipid-soluble drugs are eliminated from the body, these chemicals gradually leach back into the blood, resulting in a more sustained effect.

The Stroop task is a measure of cognitive performance in which words are presented in various colors and participants must determine the color the word is written in rather than reading the word itself. The task consists of three different conditions: conflict, congruent, and neutral. In the neutral condition nonsense words are written in different colors. This condition is used to determine baseline reaction times. In the congruent condition colored words are presented in the same colors. For example, the word red is presented in the color red. This condition has been shown to facilitate performance. In the conflict condition color words are written in different colors. For example, the word red is presented in the color green. This condition has been shown to decrease performance, a result referred to as the Stroop effect. Performance may be measured by the number of correct responses or by using reaction times.

Because the Stroop task has been used as an index of selective attention, performance on this task should be affected by caffeine. Using this index, Hasenfratz and Baettig (1992) found that 250 mg of caffeine reduced reaction times to the Stroop task in women. The present study was conducted to examine the effects of more common doses of caffeine on performance of this task, and to determine if such low doses would result in a sex difference.

**Method**

**Participants**

Participants were 63 undergraduate psychology students. Fifty-six percent of the participants were women and 51% were between ages 17 and 21. Most were well rested in that 41% indicated that it had been only 1 to 4 hr since they had last slept, and 81% of the participants indicated that they had slept 5 to 9 hr the last time they slept. In the 4 hr prior to the study 70% of the participants reported that they had eaten and 65% had consumed caffeine. Seventy-five percent of the participants consumed caffeine on a daily basis, with 57% consuming 1 to 2 cups of caffeinated beverage daily. In the 12 hr prior to the study 18% of the participants used drugs, including alcohol, prescription, nonprescription, and illicit drugs. All participants were informed of the potential health risks caused by caffeine consumption. Students who agreed to participate received extra credit in their psychology course.

**Materials**

The basic method of Stroop effect presentation used was developed by Chen and Johnson (1991). They used a voice-relay device to measure participants’ vocal response times, whereas we used color-coded keypads to measure the participants manual response times. Five IBM 386 computers with color monitors were used. Each computer had red, yellow, blue, and green tags placed over the four arrow keys. All computers were loaded with a program originally devised by Chen and Johnson that displayed colored stimuli and recorded reaction times. The colored stimuli consisted of words three to six characters long that were designed as congruent, neutral, or conflict stimuli.

A 12-oz (355-ml) can of Coca-Cola containing 45 mg of caffeine was given to participants in the experimental group. Each participant in the control group received a 12-oz (355-ml) can of caffeine-free Coca-Cola. Both the Coca-Cola and the caffeine-free Coca-Cola contained 26 mg of sugar.

**Procedure**

The study was conducted between 9:00 a.m. and 4:00 p.m. throughout the spring and summer semesters. Upon arrival participants were told they were free to leave at anytime. All participants were then instructed to drink a 12-oz (355-ml) soda with a covered label to mask whether it was Coca-Cola or caffeine-free Coca-Cola. Coca-Cola and caffeine-free Coca-Cola were randomly distributed to students in each experimental session. The participants were given 20 min to consume the beverage, delayed for 10 min, and then given 10 min to complete a questionnaire designed to determine various demographic variables, including sex, age, hours of sleep per night, food consumption, caffeine consumption, and drug use. The 10-min delay and the questionnaire completion interval were employed to allow the caffeine ample time to become diffused throughout the central nervous system.
The students were then taken to another room containing the computers. All participants were instructed to sit approximately the same distance from the computer. The monitors were checked to insure the color settings were the same on all machines. The participants were instructed to watch the screen and determine what color stimuli appeared on the screen. They were told to ignore the meaning of the stimuli, focus only on the color, and strike the red, blue, yellow, or green key that corresponded to the color on the screen. The participants were informed that they had only 3 s to determine the color before the next color appeared and that any response not made within 3 s would be scored as incorrect. Participants were then told to press any key to begin.

The task began with a blank screen with only a cursor on it; the cursor remained for 1 s after the program began. A congruent, neutral, or conflict stimulus then appeared. After a stimulus appeared, participants had 3 s to make a response before the computer returned to the cursor screen for the next trial. The 120 trials administered to each participant were a random mixture of congruent, neutral, and conflict stimuli. After the 120 trials were completed, participants were thanked for their participation and all questions were answered by the experimenter.

Procedural constraints prevented the experimenter from being blind to the condition (caffeinated or caffeine-free and sex of the participant) in which the participants were placed. For this reason the experimenter had a script that was precisely followed when interacting with each participant in order to limit the impact of experimenter bias. To further limit experimenter involvement, all data was collected and recorded by machine.

Results

A cross-tabs analysis was conducted on each survey question to determine if there were any significant demographic differences between groups prior to the study. Surprisingly, participants in the caffeine group indicated they had been awake fewer hours than those in the caffeine-free group, $\chi^2(1, N= 63) = 12.29, p < .05$. Men were significantly younger, $\chi^2(1, N=63) = 15.47, p < .05$, had been awake longer prior to the study, $\chi^2(1, N=63) = 22.55, p < .001$, and had slept fewer hours prior to the study, $\chi^2(1, N=63) = 8.62, p < .05$, than women. The differences that existed may be explained by a demographic difference in the population at the institution where the study was conducted. Women students are, on average, older than men. This age disparity may explain the sex differences that existed prior to the study. There was no significant difference in caffeine use between groups prior to the study.

We conducted a three-way repeated measures ANOVA: sex (male, female) by drug (caffeine, caffeine-free) by condition (congruent, conflict, neutral). As indicated in Figure 1, no significant main effect of caffeine was detected, $F(1, 58) = 2.91, p > .05$. A significant main effect of sex was obtained; men were quicker in their responses than women, $F(1, 58) = 22.78, p < .001$. The analysis also yielded a significant interaction between caffeine and sex, $F(1, 58) = 7.50, p < .01$. Post hoc tests (Sheffeé, $p < .05$) indicated that men were more affected by this drug than women.

There was a significant main effect of display, $F(2, 116) = 47.29, p < .001$. Post hoc analysis (Sheffeé, $p < .05$) indicated reaction times were longer in the conflict condition, showing the Stroop effect. We also found a significant interaction between sex and display, $F(2, 116) = 12.02, p < .01$. A post hoc analysis (Sheffeé, $p < .05$) determined women were significantly slower than men in the conflict condition (see Figure 1). We found no significant interaction between caffeine and display, $F(2, 116) = 30, p > .05$. The analysis also indicated the sex by caffeine by display interaction was not significant, $F(2, 116) = .01, p > .05$.

Discussion

The results of this study show low levels of caffeine decrease performance on the Stroop task of selective attention for men. This finding is consistent with other studies examining the effects of caffeine on performance (Jacobson et al., 1991). This sex difference in the effect of caffeine has been found in several other studies (James, 1990; Smith, Davidson, et al., 1993). The present study extends this sex difference to another cognitive task: the Stroop task of selective attention.

These results are congruent with sex differences concerning caffeine absorption into the body. Caffeine is a lipid-soluble drug that readily crosses the blood-brain barrier with significant brain and blood levels reached within 5 min of ingestion (Palfai & Jankiewicz, 1997). However, because it is lipid soluble it is also absorbed into the body fat. Because women on the average have higher levels of body fat than do men, the effect of caffeine would be expected to be more pronounced in men.

The present study also found a significant sex difference in the Stroop effect. The sex difference was pronounced only in the conflict condition, where women had slower reaction times than men. Von Kluge (1992) observed a similar sex difference when using the Stroop task as a measure.
The Yerkes-Dodson law appears to offer a viable explanation for the results. Because the Stroop task is considered a complex task, it would be expected to have a very low level of arousal for optimal performance. Because caffeine levels as low as 32 mg have been found to affect performance on vigilance and reaction time tasks (Lieberman et al., 1987, cited in Grilly, 1994), this level of arousal could be low enough to be exceeded by ingestion of 45 mg of caffeine.

The only results of this study that do not appear to adhere to the Yerkes-Dodson law are those between men in the caffeinated conflict and congruent conditions. As expected, caffeine improved performance of men in the congruent condition, because it was a simple task requiring a high level of arousal for peak performance. Because it is a complex task requiring a low arousal level for peak performance, it was expected that the conflict condition would show a significant decrease in performance. Surprisingly, the men in the conflict condition were not significantly slower than men in the congruent condition. This finding may be explained by the fact that the male sample was random and consisted of individuals of...
var
ing skill levels. Thus, for some men the conflict
condition may have been very complex, but for oth-
ers it may have been somewhat simple. If this was the
case, then the results would still be consistent with
the Yerkes-Dodson Law.

The implications of this study are apparent and
applicable to almost any area of professional life. Per-
haps the most immediately applicable aspect of this
study is to academic life. In test situations optimal
arousal levels may often be exceeded due to anxiety.
For men, as little as 45 mg of caffeine, or one can of
Coca-Cola, can raise them above their optimal level
of arousal in a nonanxious situation. In the test situ-
ation this increase of arousal due to caffeine could
have disastrous academic effects for the student.

In a country in which the average person con-
sumes 200 mg of caffeine daily there are obvious im-
lications. The average cup of coffee contains as
much or more caffeine than was used in this study.

References

Anderson, K. J. (1994). Impulsivity, caffeine, and task difficulty: A
within-subjects test of the Yerkes-Dodson law. Personality and
Individual Differences, 16, 813–829.

is more observable under a speed strategy than an accuracy
strategy. Perceptual and Motor Skills, 73, 67–76.

and Bacon.

and caffeine: Stroop effect, EEG, and peripheral physiology.
Pharmacology, Biochemistry and Behavior, 42, 155–161.

ships of caffeine and mental performance, EEG, cardiovascular
and subjective parameters. Psychopharmacology, 114, 281–
287.

Influence of caffeine on selected manual manipulation skills.
Perceptual and Motor Skills, 72, 1175–1181.

James, J. E. (1990). The influence of user status and anxious dis-
position on the hypertensive effects of caffeine. International

Jarvis, M. J. (1993). Does caffeine intake enhance absolute levels

W. H. Freeman.

of caffeine on selective attention in well-rested and fatigued sub-

Palfai, T., & Jankiewicz, H. (1997). Drugs and human behavior (2nd
ed.). Madison, WI: Brown and Benchmark.

Smith, A. P., Brockman, P., Flynn, R., Maben, A., & Thomas, M.
(1993). Investigation of the effects of coffee on alertness and
performance during the day and night. Neuropsychobiology, 27,
217–223.

Smith, B. D., Davidson, R. A., & Green, R. L. (1993). Effects of
caffeine and gender on physiology and performance: Further
tests of a biobehavioral model. Physiology and Behavior, 54, 415–
422.

Terry, W. S., & Phifer, B. (1986). Caffeine and memory perform-

von Kluge, S. (1992). Trading accuracy for speed: Gender differ-
ces on a Stroop task under mild performance anxiety.
Perceptual and Motor Skills, 73, 651–657.