Pharmacology Biochemistry and Behavior, Vol. 54, No. 2, pp. 309–315, 1996

0091-3057/96 $15.00 + .00

ELSEVIER

SSDI 0091-3057(95)02116-7

Effects of Caffeine, Practice, and Mode of Presentation on Stroop Task Performance

STEPHEN EDWARDS,[1] CAROLYN BRICE, CATHERINE CRAIG AND REBECCA PENRI-JONES

Department of Psychology, University of Wales Swansea, Singleton Park, Swansea SA2 8PP, Wales, UK

Received 28 May 1995

EDWARDS, S., C. BRICE, C. CRAIG AND R. PENRI-JONES. *Effects of caffeine, practice, and mode of presentation on Stroop task performance.* PHARMACOL BIOCHEM BEHAV 54(2) 309–315, 1996.—A series of experiments were designed to investigate the effects of 125 and 250 mg caffeine, or placebo, on performance of the Stroop task. Caffeine had no effect on performance of either the classic colour-word version or a numerical version of the task, either using computerised presentation of the stimuli or a traditional card version. However, significant practice effects were found using a within-subjects design with the card version of the task, and differences were found between performance in the card and computerised versions of the task. It is concluded that at these doses, caffeine does not significantly affect Stroop performance. It is also suggested that practice effects in studies using within-subject designs may be a problem when attempting to detect subtle effects of drugs on cognitive performance, and that the computerised version of the Stroop task may not be an exact analogue of the traditional card version.

Stroop task   Caffeine   Cognitive performance   Attention   Computer

ALTHOUGH the effects of caffeine on cognitive performance have been the subject of numerous experiments and reports [e.g., (1,4) for reviews], there are many contradictions and inconsistencies in the literature. It it not unexpected, then, that it has been concluded that caffeine has yet to be shown to have consistent effects on cognitive function (4). However, one aspect of cognitive performance that has attracted little attention with reference to caffeine intake is assessed using the Stroop paradigm (7), which has been in use for over half a century [see (5) for review], and has been described as "the gold standard of attentional measures" (6). Because this paradigm is essentially a test of selective attention, a clear demonstration that caffeine influences Stroop performance would, for the first time, reveal an unambiguous and consistent cognitive effect of the drug, and it is perhaps surprising that relatively little work has been done in this area. Nevertheless, two recent papers have addressed the issue. In the first one, Foreman et al. (2) gave subjects 125 or 250 mg caffeine, or placebo, and reported that the higher dose of caffeine significantly increased the magnitude of the Stroop effect in a numerical version of the task, compared to placebo. It was suggested that this apparently clear finding, in contrast to the variability seen in much of the previous work on the cognitive effects of caffeine, may be explained by the demanding nature of the Stroop task, involving the processing of ambiguous or confusing stimuli, which may make it particularly sensitive to high levels of caffeine. Unfortunately, however, because only difference scores for performance in the neutral condition and the confusing Stroop condition were given in this paper, it is impossible to evaluate whether the observed differences in Stroop effect magnitude were due primarily to differences in performance in the confusing Stroop condition or in the neutral condition. If it was the neutral condition that was primarily affected, the authors' explanation of the observed effect would be considerably weakened, whereas if it were the confusing Stroop condition that was primarily affected, the explanation would be strengthened, and would represent an important step in the understanding of the cognitive effects of caffeine.

In the second paper, however, Hasenfratz and Bättig (3) contradicted the findings of Foreman et al. (2). They included four groups in their experiment—250 mg caffeine only, smoking only, caffeine plus smoking, and control—in a numerical version of the task. They reported that when the interval between the presentation of successive Stroop stimuli was 1 s, the treatments, when compared to control, had no effect on the general Stroop performance improvements observed between pretreatment and posttreatment testing. When the inter-

[1] To whom requests for reprints should be addressed.

stimulus interval was 0 s, the improvements observed in posttreatment testing compared to pretreatment testing were significantly greater in subjects in the smoking-only and caffeine-only groups than in the control group, although there was no difference between control and the smoking plus caffeine group. However, a major confounding factor in this study was that subjects in both the control and caffeine conditions were regular smokers who were in a state of nicotine deprivation. It is not possible to deduce from the paper the precise length of nicotine deprivation, but if testing started at 0900 h, then the subjects would have been nicotine-deprived for 13 h. The importance of this factor is underlined by their finding that the smoking-only group, but not the caffeine plus smoking group, displayed a reduced Stroop effect compared to control. They did not directly compare the smoking-only group with the caffeine plus smoking group (which would represent a caffeine vs. noncaffeine comparison in nicotine-replete smokers), but inspection of their data shows that the Stroop performance of the smoking-only group improved between pretreatement and posttreatment testing, whereas the performance of the smoking plus caffeine group deteriorated. Thus, from these data, although caffeine may reduce the magnitude of the Stroop effect in nicotine-deprived smokers, it is unlikely that it would so in nicotine-replete smokers. One alternative reasonable interpretation of their data is that administration of a stimulant (rather than caffeine per se) improves Stroop performance in nicotine-deprived smokers. It may also be possible, at least in principle, to interpret their findings in terms of the influence of stimulants on practice effects, although the lack of information on the training procedure given in the paper, together with the confound of nicotine deprivation, make such an analysis impossible on the basis of the data presented. Whichever explanation is correct, these data do not unambiguously support those of Foreman et al. (2).

Given these inconsistencies, we decided to investigate this issue further, and designed a series of experiments to explore the nature of the effects of caffeine on performance of the Stroop task. Initially, it was decided to examine the effects of caffeine on both the numerical version of the task used by Foreman et al. (2) and the classic colour-word version (7), using the same doses as used in the papers described before.

EXPERIMENT 1

Method

Subjects. We used 20 subjects (10 male and 10 female), ages 18–23. On recruitment, a caffeine questionnaire was administered. Only subjects who regularly consumed the caffeine equivalent of at least two cups of coffee per day were included in the study. All subjects were requested to avoid consuming any caffeine-containing food or beverages on the morning of the study, and to eat breakfast before 0830 h. Subjects who smoked were not asked to refrain from smoking. Subjects with high blood pressure or taking oral contraceptives were excluded from the study.

Drugs. Caffeine in the form of ground Pro-Plus tablets (Roche Nicolas Consumer Healthcare, Welwyn Garden City, UK) was administered in two white, opaque, size-0 gelatin capsules (Eli Lilley, Basingstoke, UK). Corn flour was used as a packing agent and as the placebo, administered similarly in two capsules. Caffeine doses of 125 and 250 mg were used, spread evenly across the two capsules. Drug administration was performed with the investigators double-blinded.

Apparatus and procedure. Subjects attended the laboratory for testing at 1000 h. They were given the appropriate caffeine dosage with 30 ml of mineral water, and were asked not to eat or drink anything until the testing procedure was completed. One hour later, they returned to the laboratory to complete the Stroop tasks.

In the classic colour-word Stroop task, subjects were asked to identify the colours in which a sequence of words were printed, ignoring the semantic content of the printed word. In the neutral condition, the printed words were window, cook, hat, and shade. In the colour-word condition, the words were yellow, red, green, and blue. In both conditions, the words were printed on a card consisting of 50 words (in five columns of 10 words each) in yellow, red, green, and blue ink colours. In the colour-word condition, the colour of the ink and the colour-word printed were always incongruent. In both conditions, the same word or ink colour never appeared twice in succession, and the frequency of occurrence of words and colours were counterbalanced across columns as far as possible. Subjects were asked to work down each column in turn, naming the ink colour of each word in turn as quickly and accurately as possible. The order of presentation of the neutral and colour-word cards was counterbalanced, and the times taken to complete each card was noted, together with the number of errors made.

The apparatus and procedure for the numerical Stroop task was the same as for the colour-word Stroop task, except that the stimuli were symbols and digits (printed in black ink). In the neutral condition, each stimulus consisted of two to five asterisks. In the digit condition, each stimulus consisted of two to five digits (e.g., 2222). The digits used were 2, 3, 4, and 5, and, in each case, the printed digits were always incongruous with the number of digits printed. The subjects were asked, for each stimulus, to name the number of digits or symbols printed.

In each case, the subjects were given a set of 10 practice stimuli before completion of the relevant condition. The order of completion of the colour-word and numerical Stroop tasks was counterbalanced. Each subject attended testing on three occasions, with the order of administration of the three caffeine conditions randomised across the three sessions.

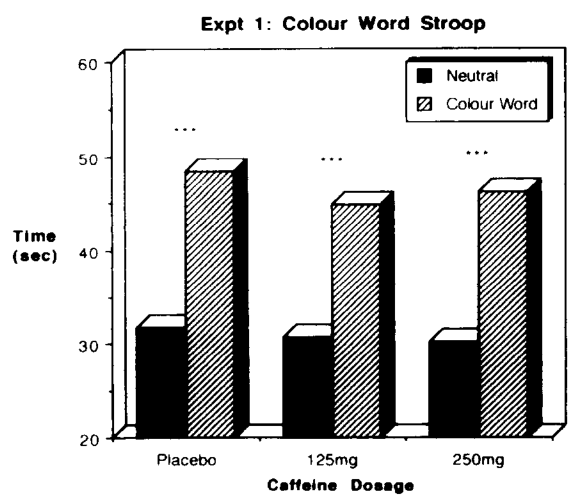

FIG. 1. Effect of caffeine on performance of the classic colour-word Stroop task in Experiment 1. $^{***}p < 0.001$.

Results

The data for the classic colour-word Stroop task were analysed using a two-way within-subjects analysis of variance (ANOVA), with caffeine dosage and Stroop condition as factors (Fig. 1). A significant main effect of Stroop condition was found [$F(1, 19) = 108.05, p < 0.0001$], but not caffeine dosage [$F(2, 38) = 1.53, p > 0.05$]. No significant interaction effect was found ($F < 1, p > 0.05$). Nevertheless, analysis of simple main effects revealed that a significant Stroop effect was observed in all three caffeine conditions [smallest $F(1, 19) = 35.76, p < 0.001$]. The error rate was very low (<4%); therefore, no analysis was performed on these data (this was true in all versions of the Stroop task in all three experiments).

The data for the numerical Stroop task were analysed similarly (Fig. 2). A significant main effect of Stroop condition was found [$F(1, 19) = 110.57, p < 0.0001$], but no significant main effect of caffeine was found ($F < 1, p > 0.05$). No significant interaction effect was found ($F < 1, p > 0.05$). Nevertheless, an analysis of simple main effects revealed that a significant Stroop effect was observed in all three caffeine conditions [smallest $F(1, 19) = 65.91, p < 0.001$].

Thus, caffeine had no detectable effect on Stroop performance. However, an inherent problem with using a within-subjects design is the potential for practice effects. Because caffeine had no effect, the data were recast to test for practice effects, with the first, second, and third sessions (for each subject) as a factor. Thus, the data for the classic colour-word Stroop task were analysed using a two-way within-subjects ANOVA, with session and Stroop condition as factors (Fig. 3). Not surprisingly, a significant main effect of Stroop condition was found [$F(1, 19) = 103.26, p < 0.0001$]. However, a significant main effect of session was also found [$F(2, 38) = 13.65, p < 0.0001$]. In addition, a significant interaction effect was found [$F(2, 38) = 3.33, p < 0.05$]. An analysis of simple main effects showed that there was a significant practice effect across time for both the neutral [$F(2, 38) = 24.45, p < 0.001$] and colour-word [$F(2, 38) = 8.26, p = 0.001$] conditions. Post-hoc Tukey testing showed that a significant practice effect occurred between sessions 1 and 2 for the colour-word ($p < 0.01$) but not the neutral word condition.

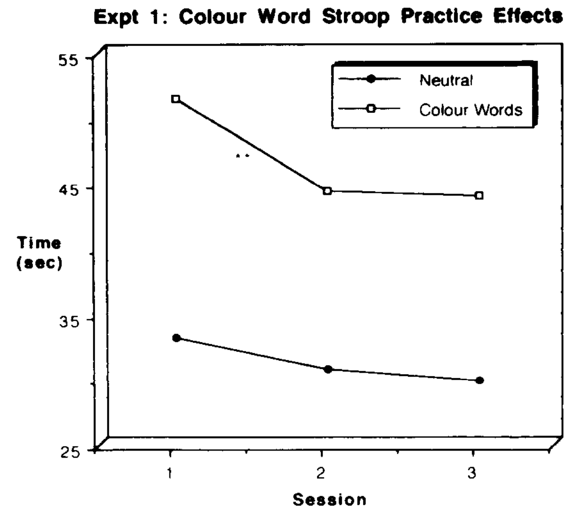

FIG. 3. Performance across sessions of the classic colour-word Stroop task in Experiment 1. $^{**}p < 0.01$.

There was no significant difference between sessions 2 and 3 in either Stroop condition. Thus, the source of the significant interaction was clearly the greater practice effect in the colour-word condition, which was particularly strong between sessions 1 and 2.

A similar analysis was carried out for the data from the numerical Stroop test (Fig. 4). Again, a significant main effect of Stroop condition was found [$F(1, 19) = 110.56, p < 0.0001$]. However, a significant main effect of session was also found [$F(2, 38) = 21.55, p < 0.0001$], as well as a significant interaction effect [$F(2, 38) = 4.65, p < 0.05$]. An analysis of simple main effects showed that there was a significant practice effect across time for both the neutral [$F(2, 38) = 14.24, p < 0.001$] and digit [$F(2, 38) = 18.34, p < 0.001$] conditions. Post-hoc Tukey testing showed that a significant practice effect occurred between sessions 1 and 2 for the colour-

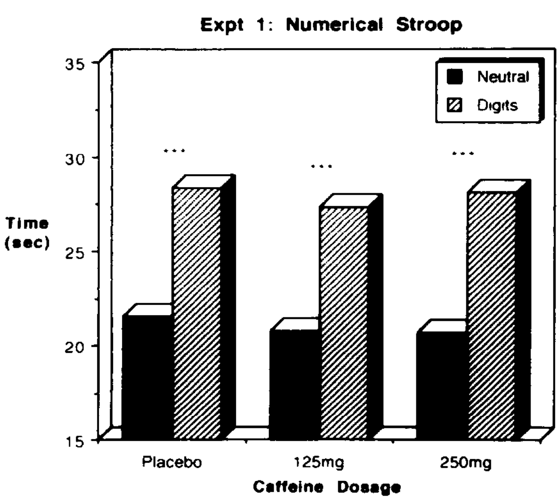

FIG. 2. Effect of caffeine on performance of the numerical Stroop task in Experiment 1. $^{***}p < 0.001$.

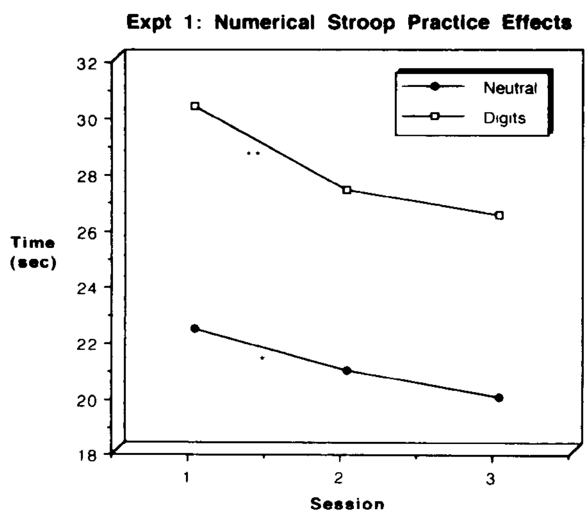

FIG. 4. Performance across sessions of the numerical Stroop task in Experiment 1. $^{**}p < 0.01$; $^{*}p < 0.05$.

word ($p < 0.01$) and the neutral word condition ($p < 0.05$). There was no significant difference between sessions 2 and 3 in either Stroop condition. Thus, the source of the significant interaction was again the greater practice effect in the colour-word condition between sessions 1 and 2.

Although there was no a priori reason to expect sex differences, the analyses were repeated with sex as an additional factor. No significant sex differences were found.

Discussion

The data clearly indicate that although the expected Stroop effect was observed in all conditions for both tests, caffeine had no effect on either the magnitude of the Stroop effect or the speed at which the task was completed. These findings are inconsistent with previous reports (2,3). However, there are two possible reasons for this. First, in both of our tests practice effects were found, which were greater in the colour word than the neutral conditions. There is a danger, therefore, that any effect caffeine may have had on Stroop performance was swamped by noise introduced into the data by practice effects. Second, the previous papers both used a computerised mode of presentation, incorporating a locomotor response, in contrast to the standard card presentation and vocal response procedure used here. Thus, it is possible that effects of caffeine on Stroop performance are either specific to single stimulus presentation (as occurs in the computerised version of the tasks) or are a reflection of effects on locomotor performance rather than Stroop performance per se. Thus, we decided to repeat this experiment using a between-subjects design, and incorporating computerised as well as standard card presentation.

EXPERIMENT 2

Method

Subjects. We used 60 female subjects, aged 19–41. Conditions were as for Experiment 1 above.

Apparatus and procedure. The general procedure was exactly the same as that used for Experiment 1, except that

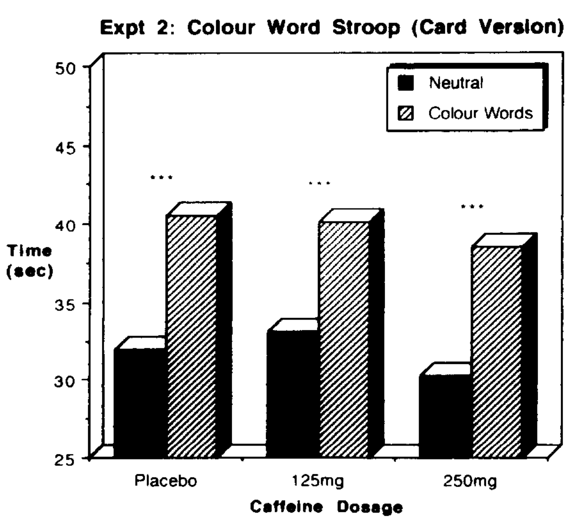

FIG. 5. Effect of caffeine on performance of the card version of the classic colour-word Stroop task in Experiment 2. $^{***}p < 0.001$.

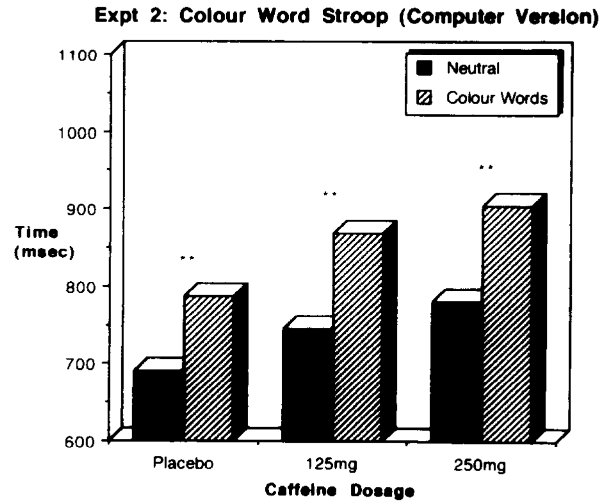

FIG. 6. Effect of caffeine on performance of the computerised version of the classic colour-word Stroop task in Experiment 2. $^{**}p < 0.005$.

computerised versions of the Stroop tasks were added to the test battery, and the design was between- rather than within-subjects. Thus, each subject attended on only one occasion. An IBM-compatible PC was used to display the stimuli and record response times. Stimuli were administered singly on screen, and responses were made via a push-button apparatus. This contained four response keys and a home key. Subjects were asked to use the index finger of their preferred hand to make responses. Successive stimuli appeared on screen only after the previous response had been made, and the index finger replaced on the home key. Thus, subjects were able to work through the stimuli at their own pace. The order of presentation of the four versions of the Stroop task was counterbalanced, and for each version the order of presentation of the neutral and colour-word/digit conditions was counterbalanced. In this experiment, subjects attended for testing on only one occasion, and were randomly allocated to one of the three caffeine conditions.

Results

The data for the card version of the classic colour-word Stroop task were analysed using a two-way between-subjects ANOVA, with caffeine dosage and Stroop condition as factors (Fig. 5). A significant main effect of Stroop condition was found [$F(1, 57) = 93.07$, $p < 0.0001$], but not caffeine dosage ($F < 1$, $p > 0.05$). No significant interaction effect was found ($F < 1$, $p > 0.05$). Nevertheless, an analysis of simple main effects revealed that a significant Stroop effect was observed in all three caffeine conditions [smallest $F(1, 57) = 24.09$, $p < 0.001$].

The data for the computerised version of the classic colour-word Stroop task were analysed similarly (Fig. 6). A significant main effect of Stroop condition was found [$F(1, 57) = 46.44$, $p < 0.0001$] but not caffeine dosage [$F(2, 57) = 2.95$, $p > 0.05$]. No significant interaction effect was found ($F < 1$, $p > 0.05$). Nevertheless, an analysis of simple main effects revealed that a significant Stroop effect was observed in all three caffeine conditions [smallest $F(1, 57) = 11.13$, $p < 0.005$].

The data for the card version of the numerical Stroop task were also analysed similarly (Fig. 7). A significant main effect of Stroop condition was found [$F(1, 57) = 188.67$, $p < 0.0001$] but not caffeine dosage ($F < 1, p > 0.05$). No significant interaction effect was found ($F < 1, p > 0.05$). Nevertheless, an analysis of simple main effects revealed that a significant Stroop effect was observed in all three caffeine conditions [smallest $F(1, 57) = 54.91, p < 0.001$].

Finally, the data for the computerised version of the numerical Stroop task were analysed similarly (Fig. 8). A significant main effect of Stroop condition was found [$F(1, 57) = 40.20, p < 0.0001$] but not caffeine dosage [$F(2, 57) = 1.96$, $p > 0.05$]. No significant interaction effect was found ($F < 1, p > 0.05$). Nevertheless, an analysis of simple main effects revealed that a significant Stroop effect was observed in all three caffeine conditions [smallest $F(1, 57) = 9.45$, $p < 0.005$].

Discussion

Once again, the data clearly indicated that although the expected Stroop effect was observed in all conditions for both tests in both card and computerised versions, caffeine had no effect on either the magnitude of the Stroop effect or the speed at which the task was completed. It would appear, therefore, that the failure to observe any effect of caffeine on Stroop performance in Experiment 1 was due to neither noise introduced into the data by practice effects nor mode of presentation. However, there is still one possible explanation for the discrepancy between the data reported here and previous data. Here, each part of the Stroop test was composed of 50 trials, which is a lower number than that used in the previous reports (where 110 were used). Thus, a final experiment was designed to test the possibility that any effects of caffeine on Stroop performance may not become evident until more than 50 trials have been completed.

EXPERIMENT 3

Method

Subjects. We used 27 subjects (nine male and 18 female). The mean age was 21.4 years. Conditions were as for Experiments 1 and 2.

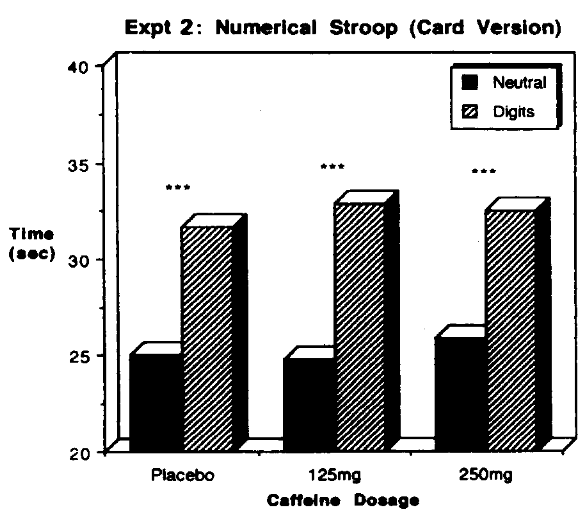

FIG. 7. Effect of caffeine on performance of the card version of the numerical Stroop task in Experiment 2. ***$p < 0.001$.

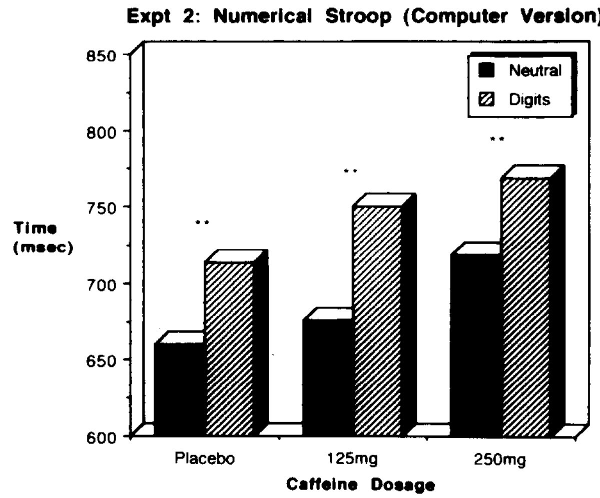

FIG. 8. Effect of caffeine on performance of the computerised version of the numerical Stroop task in Experiment 2. **$p < 0.005$.

Apparatus and procedure. The general procedure was exactly the same as that used for Experiment 2, except that only the colour-word version of the Stroop task was used, and that in each condition 200 stimuli were used in four blocks of 50. In the card version, four separate cards, each containing 50 stimuli, were used in each condition. In both versions, an interval of 1 min was left between blocks. In addition, the 125 mg caffeine condition was omitted.

Results

The data for the card version of the classic colour-word Stroop task were analysed using a three-way between-subjects ANOVA, with caffeine dosage, Stroop condition, and block as factors. A significant main effect of Stroop condition was found [$F(1, 25) = 101.04, p < 0.0001$] but not caffeine dosage [$F(1, 25) = 1.31, p > 0.05$]. However, a significant main effect of block was found [$F(3, 73) = 4.98, p < 0.005$]. No significant two-way interaction effects were found between caffeine dosage and block ($F < 1, p > 0.05$) or between caffeine dosage and Stroop condition [$F(1, 25) = 1.62, p > 0.05$]. However, a significant two-way interaction was found between block and Stroop condition [$F(3, 75) = 5.31, p < 0.005$] (Fig. 9). An analysis of simple main effects showed that there was no significant difference across blocks for the time taken to complete the colour-word condition [$F(3, 75) = 1.76, p > 0.05$]. However, such a difference did occur for the neutral word condition [$F(3, 75) = 16.36, p < 0.001$]. Post-hoc Tukey testing revealed that times for block 4 were significantly slower than those for block 2 ($p < 0.05$) and block 1 ($p < 0.01$). Thus, for the neutral word condition, but not the colour-word condition, subjects became slower across time. No significant three-way interaction was found [$F(3, 75) = 1.01, p > 0.05$].

The data for the computerised version of the classic colour-word Stroop task were analysed similarly. A significant main effect of Stroop condition was found [$F(1, 25) = 10.39, p < 0.005$], but not caffeine dosage ($F < 1, p > 0.05$). However, a significant main effect of block was found [$F(3, 73) = 7.38$, $p < 0.0005$]. No significant two-way interaction effects were found between caffeine dosage and block ($F < 1$, $p > 0.05$)

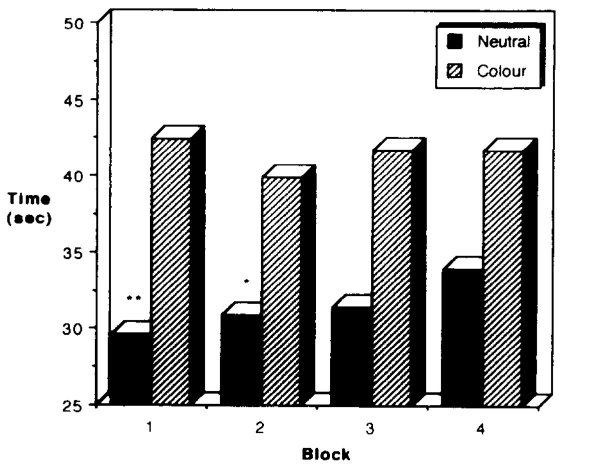

FIG. 9. Performance across blocks of the card version of the classic colour-word Stroop task in Experiment 3. Asterisks indicate significantly faster than block 4: $^{**}p < 0.01$; $^{*}p < 0.05$.

or between caffeine dosage and Stroop condition ($F < 1, p > 0.05$). However, the two-way interaction between block and Stroop condition almost achieved statistical significance [$F(3, 75) = 2.49$, $p = 0.07$] (Fig. 10). An analysis of simple main effects showed that there was a significant difference across blocks for the time taken to complete both the neutral word condition [$F(3, 75) = 3.34$, $p < 0.05$] and the colour-word condition [$F(3, 75) = 7.02$, $p < 0.001$]. For the neutral word condition, post-hoc Tukey testing revealed that times for block 4 were significantly faster than those for block 1 ($p < 0.05$). For the colour-word condition, times for block 2 were significantly faster than those for block 1 ($p < 0.05$), and times for blocks 3 and 4 were also significantly faster than those for block 1 ($p < 0.01$). Thus, for both word conditions, subjects became faster across time, although this was greater in the colour-word condition than for the neutral condition. No significant three-way interaction was found [$F(3, 75) = 1.27$, $p > 0.05$].

Discussion

It is again clear from the data that caffeine had no significant effect on Stroop performance. Thus, the failure to replicate previous reports (2,3) in Experiments 1 and 2 cannot be due to the lesser number of trials used. However, an interesting effect observed here was that Stroop performance changed across time, and that these changes were different in the card and computerised versions of the task. Using card presentations, subjects tended to get slower over time in the neutral word condition. However, using the computerised presentation, subjects tended to get faster across time in both neutral and colour-word conditions. Although in the case of computerised presentation the interaction between block and Stroop condition did not quite achieve statistical significance, the effect was considerably more pronounced in the colour-word condition. Thus, the results of this experiment suggest that the card and computerised versions of the Stroop task are not exact analogues.

GENERAL DISCUSSION

There are three main conclusions from this experiment. First, caffeine had no detectable effect on performance of the Stroop task. This was true for card and computerised versions, for classic colour-word and numerical versions, and, in the case of the colour-word version, for numbers of trials up to 200. These results contradict the findings of Foreman et al. (2). They also contradict the conclusions of Hasenfratz and Bättig (3), which were essentially based on data from nicotine-deprived subjects. Thus, the balance of the evidence to date suggests that at these doses, caffeine has no effect per se on performance of the Stroop task, except possibly in combination with other significant factors, such as nicotine depletion in smokers or extensive training in the Stroop paradigm.

Second, the use of a within-subjects design in Experiment 1 revealed large practice effects across test sessions. Although the noise introduced into the data by practice effects here could not account for the failure of caffeine to affect performance on the Stroop task, it is suggested that this phenomenon should be taken into account as a drawback of within-subjects designs when undertaking any attempt to detect subtle effects on Stroop performance in psychopharmacologic studies. The detection of practice effects across blocks in the computerised version of the task in Experiment 3 also suggests that this phenomenon may also become important in between-subjects designs involving large numbers of trials.

Third, the data from Experiment 3 suggest that the card and computerised versions of the Stroop task are not exact analogues. In the card version of the task, subjects became slower across time in the neutral word condition, but not the colour-word condition. In the computerised version, subjects became quicker across time in both conditions. This effect was more exaggerated in the colour-word condition, although the interaction did not quite achieve statistical significance. There are a number of possible explanations for the observed differences between the two versions of the task. First, they may be a function of response mode. Pressing a button may be a more difficult response mode initially, allowing greater scope for practice effects. Conversely, a vocal response may be easier initially, with greater scope for fatigue effects to occur. Second, they may be a function of fatigue in making saccadic eye

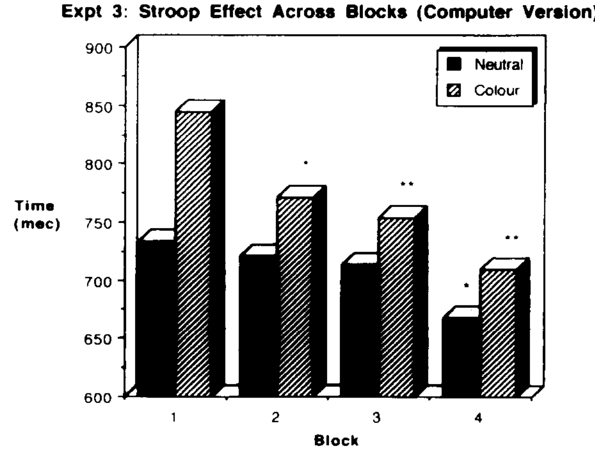

FIG. 10. Performance across blocks of the computerised version of the classic colour-word Stroop task in Experiment 3. Asterisks indicate significantly faster than block 1: $^{**}p < 0.01$; $^{*}p < 0.05$.

movements in the card version of the task. Such movements may be less important in the computerised version of the task, where stimuli appear singly on the screen. Third, they may be a function of the relative difficulty of developing strategies to enhance task performance in the two versions. Whichever explanation is correct, the data presented here suggest that the two versions of the task are not exactly analogous, and that it might be dangerous to generalise performance from one version of the task to another, especially when large numbers of trials are administered.

ACKNOWLEDGEMENTS

The authors thank Neil Carter, John Griffiths, Harry Olive, and Ken Williams for technical assistance.

REFERENCES

1. Bättig, K.; Welzl, H. Psychopharmacological profile of caffeine. In: Garattini, S., ed. Caffeine, coffee, and health. New York: Raven Press; 1993:213–253.
2. Foreman, N.; Barraclough, S.; Moore, C.; Mehta, A.; Madon, M. High doses of caffeine impair performance of the Stroop task in men. Pharmacol. Biochem. Behav. 32:399–403; 1989.
3. Hasenfratz, M.; Bättig, K. Action profiles of smoking and caffeine: Stroop effect, EEG, and peripheral physiology Pharmacol. Biochem. Behav. 42:155–161; 1992.
4. James, J. E. Caffeine and health. London: Academic Press; 1991.
5. MacLeod, C. M. Half a century of research on the Stroop effect: An integrative review. Psychol. Bull. 109:163–203; 1991.
6. MacLeod, C. M. The Stroop task: The "gold standard" of attentional measures. J. Exp. Psychol. Gen. 121:12–14; 1992.
7. Stroop, J. R. Studies of interference in serial verbal reactions. J. Exp. Psychol. 18:643–662; 1935.