Perceptual and Motor Skills, 1991, 73, 67-76.

THE STROOP CONGRUENCY EFFECT IS MORE OBSERVABLE UNDER A SPEED STRATEGY THAN AN ACCURACY STRATEGY [1,2]

JENN-YEU CHEN AND MARCIA K. JOHNSON

New York State Psychiatric Institute *Princeton University*

Summary.—Naming the ink color of an incongruent color word (e.g., RED printed in green) usually takes longer than naming the ink color of a color bar. However, when the ink matches the word (e.g., RED printed in red), naming tends to be faster. These phenomena are known as the Stroop interference effect and the Stroop congruency effect, respectively. Although the interference effect has been robust and reliable across studies, the congruency effect tends to be elusive. It was hypothesized that this variation in outcomes might be related to subjects' response strategy. The experiment conducted to test this hypothesis induced either a speed or an accuracy strategy in two separate groups of subjects. Significant interference effects were found for both groups and the magnitudes did not differ. At the same time, the congruency effect was observed in the speed group but not in the accuracy group. These results suggest that researchers who wish to observe and study the Stroop congruency and interference effects should place special emphasis on speed. Implications of the study for a model of the Stroop effect are also discussed.

In a Stroop color-naming task, a stimulus is printed in some color and subjects are asked to name the color. Three kinds of stimuli are usually included, neutral (e.g., a red color bar or xxxx printed in red), congruent (e.g., the word RED printed in red), and conflict (e.g., the word GREEN printed in red ink). The neutral condition serves as a control for assessing subjects' response speed and accuracy in the other conditions. The typical findings are an increase in both response time and number of errors in the conflict condition and a decrease in both in the congruent condition. These phenomena have been referred to as the Stroop interference effect and congruency effect, respectively (Stroop, 1935; Dyer, 1973).

Traditionally, the interference effect has been the focus of most studies because it bears directly on theories of selective attention (Broadbent, 1958; Deutsch & Deutsch, 1963; Norman, 1968). In particular, the effect has been thought to indicate how and where attention fails to select relevant information from the environment (Hock & Egeth, 1970; Morton & Chambers, 1973). The interference effect has also been very reliable, making it a favorite for study.

The story for the congruency effect has been different, however. First,

[1]This research was supported in part by National Science Foundation Grant BNS-8510633. We gratefully acknowledge the technical assistance of the staff of the Social Sciences Electronic Shop at the State University of New York, Stony Brook, as well as Tina Loose, who helped in recruiting and testing subjects at Princeton University.
[2]Address correspondence to Jenn-Yeu Chen, Ph.D., New York State Psychiatric Institute, Box 78, 722 West 168th Street, New York, NY 10032.

facilitation has not been as reliable a phenomenon as interference; although some investigators (e.g., Dalrymple-Alford, 1972; Hintzman, Carre, Eskridge, Owens, Shaff, & Sparks, 1972) have found it, others have not (e.g., Seymour, 1977; Stirling, 1979; Stirling & Coltheart, 1977). Second, for those studies claiming a congruency effect, there is some doubt about the adequacy of the control condition used to derive the effect. For example, Dalrymple-Alford (1972) used noncolor words as the control rather than color bars or crosses. Because Klein (1964) has shown that noncolor words themselves also produced some interference (i.e., naming time was longer as compared with that for color bars), the use of noncolor words as the control might have boosted a perhaps nonexistent congruency effect into existence. The same possibility can be found in Hintzman, *et al.*'s study (1972). They used single-solution anagrams constructed from color words (e.g., RDE, BEUL) as the control. The mean response times for these anagram stimuli (696 msec.) and for the noncolor word stimuli (694 msec.) also included in their study were very similar. In both studies, if color bars or crosses had been used as the control, there might not have been a congruency effect.

As a result of its elusive status, few attempts have been made to explain the congruency effect. Moreover, since there has been some dispute over what type of stimuli should be used as the control for measuring the interference and congruency effects (Zajano, Hoyceanyls, & Quellette, 1981), some researchers have suggested we forget about the congruency effect and only use the congruent condition as the control for measuring the interference effect (e.g., Virzi & Egeth, 1985; McClain, 1983; W. T. Neill, 1977, personal communication). However, such a practice could lead to a spurious interference effect, or, at the least, inflate its magnitude should a congruency effect become established as a phenomenon in its own right. Besides, if there is a theoretical reason to believe that the congruency effect should exist, the effect should be given more attention.

In fact, a congruency effect is consistent with any model in which activation (or response tendencies) from various sources can accumulate. For example, a spreading activation network model of human memory (e.g., Collins & Loftus, 1975; Anderson, 1983; Cohen, Dunbar, & McClelland, 1990) would predict a congruency effect. The reasoning is that, when a color word and a color both activate their corresponding nodes in memory, activations spread, meet, and sum to exceed the response threshold faster than when only the color node is activated. The congruency effect, then, presumably reflects a savings in processing time when two different stimulus dimensions converge on the same response compared with when only one does.

The congruency effect, should it become a reliable phenomenon, would also present a challenge to the existing theories of the Stroop effect, which are primarily theories of interference. The perceptual competition theory was discounted partly because it could not properly deal with the congruency effect (Hintzman, *et al.*, 1972). Other theories must address the effect as well to achieve generality and validity. However, to expect this to happen, one needs a way of observing the congruency effect reliably.

The purpose of the present study is to identify a factor that might contribute to the elusiveness of the congruency effect, which, if controlled, might produce a more stable phenomenon for study. Aside from the use of different control conditions in deriving the effect, a possible reason for the elusiveness of the effect is variation in subjects' response strategies. Subjects who focus on being accurate might tend to be overcautious and not trust a response that came to mind first because it could be an incorrect one. This seems plausible given that conflict trials usually are mixed with neutral and congruent trials. As a result, any savings in processing time for the congruent trials might be canceled out by a double-check procedure imposed on every response. On the other hand, if subjects focused on responding quickly and were not too concerned about making errors, they might be more likely to benefit from the congruent trials and show a congruency effect.

To test the hypothesis that response strategy affects the congruency effect, we encouraged an accuracy strategy in one group of subjects and a speed strategy in another group. If response strategies do influence the congruency effect, the effect should be more likely to occur in the speed group than in the accuracy group because subjects should be less likely to double-check in the speed group. The interference effect, however, was expected not to be affected by the strategy because double-checking a response is always necessary on a conflict trial, assuming that subjects would not generally want to err on too many such trials.

Method

Subjects

Twenty-four college students (11 men and 13 women) were recruited from an introductory psychology course at the State University of New York, Stony Brook. An additional 23 students (4 men and 19 women) were recruited from Princeton University. All the subjects reported normal color vision. This was further verified by the experimenter during the practice trials, where the subjects actually named the colors used in the experiment.

Apparatus and Materials

An Apple IIe microcomputer was programmed to present the stimuli on an Amdek-II color monitor. A voice-relay device and a John Bell parallel interface card with timer (Model No. 79295) were connected with the computer to measure subjects' vocal response times. The timer measured response times to milliseconds. Presentation of a stimulus started the timer, which was stopped by subjects' vocal responses detected by the voice-relay device.

Four colors and their names were used: RED, BLUE, GREEN, and YELLOW. There were three types of 40 trials each. Each of the neutral trials consisted of an array of three to six symbols randomly arranged and printed in one of the four colors (e.g., *@&%#$ printed in blue). The advantage of using these symbols was that they controlled for homogeneity of form. Color bars and crosses have been criticized as inadequate controls because they contain homogeneous forms across trials while color-words contain nonhomogeneous forms (Zajano, *et al.*, 1981). Each of the congruent trials consisted of a color-word printed in its own color (e.g., RED printed in red). Each of the conflict trials consisted of a color-word printed in one of the other three colors than its own (e.g., RED printed in blue).

The three types of trials were randomly ordered into a block of 120 trials, with the restriction that the same color or color word did not appear twice in a row. These stimuli were used in the naming task. Another block of 120 trials was prepared in a similar way, with the only difference in the neutral trials. Here, the neutral trials were color words printed in screen white color. These stimuli were used in the reading task. A new random order of trials for naming and reading tasks was constructed for each subject.

The reason neutral, congruent, and conflict stimuli were mixed instead of presented in blocks is that we wanted to avoid reducing the congruency condition into a reading condition. Subjects might develop the strategy of simply reading the words if all the congruent trials were grouped in a single block. As a matter of fact, the instability of the congruency effect across studies occurred primarily when different types of stimuli were mixed rather than blocked and so subjects could not simply read the words.

Design and Procedure

The experiment employed a $2 \times (2 \times 3)$ mixed design, with strategy as the between-subjects factor and task and relation as the within-subjects factors. The two response strategies were speed and accuracy. The two tasks were reading and naming. The three target-distractor relations were neutral, congruent, and conflict, corresponding to the three types of trials described previously.

The order of the tasks was counterbalanced across subjects, with half of the subjects in each group receiving the reading task first and the other half receiving the naming task first. In the reading task, all the subjects were asked to read aloud the color words regardless of the colors that the words were printed in. In the naming task, the subjects had to name the ink colors of the words instead.

The subjects were randomly assigned to the speed group or the accuracy group. The assignment put 9 men and 14 women in the speed group and 6 men and 18 women in the accuracy group. Subjects in the speed group were told that the purpose of the experiment was to assess how quickly people could make up their minds and say a response. They were told that accuracy was of little concern to the experimenter and that they should not be worried about making mistakes. Subjects in the accuracy group, on the other hand, were instructed that the experimenter was interested in finding out how accurately people could make a response under uncertainty. They were encouraged to think over a potential response before actually saying it. Both groups, however, were told not to correct a mistake should one be made.

Subjects were seated at a viewing distance of 275 cm from the screen. A trial started with a cursor flashing at the center of the screen for about one second. Subjects were told to look at the cursor when preparing to receive a stimulus, which followed the cursor after about 1 second. The stimulus remained on the screen until subjects responded within the limit of 2 seconds. (None of the subjects responded later than the limit.) Subjects' vocal response removed the stimulus and brought the cursor back on the screen. Each stimulus measured 3.4 to 6.8 cm in width and 0.6 cm in height. It subtended 0.13° of visual angle vertically and 0.71° to 1.42° horizontally. Subjects were given 20 practice trials for each task.

RESULTS

Mean response time for correct responses and number of errors out of 40 trials were computed for each subject in each relation condition within each task. These measures, further averaged across subjects, are presented in Figs. 1 and 2 as well as Tables 1 and 2.

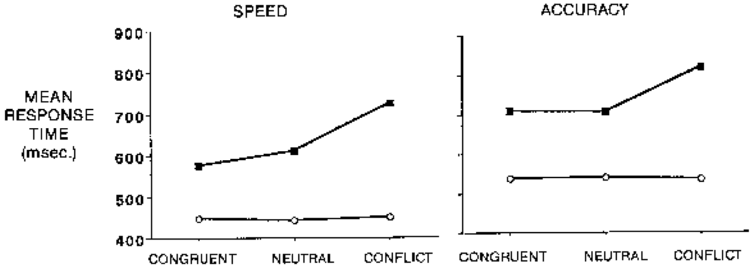

FIG. 1. Mean response times as a function of strategy (speed or accuracy), task [reading (○) or naming (■)] and target-distractor relation (congruent, neutral, or conflict)

Multivariate analysis of variance was employed. An initial check of possible differences in outcomes for subjects from the two universities (sample) gave only a significant main effect for sample ($F_{1,43} = 4.85$, $MSe = 26426.62$, $p < .04$); none of the interactions involving the sample factor were significant (p ranges from .08 to .98). The two samples were then combined in all subsequent analyses.

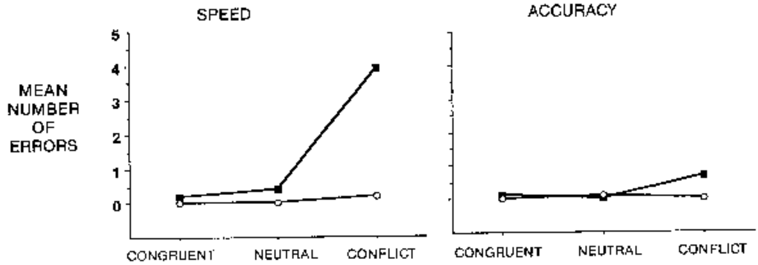

FIG. 2. Mean number of errors as a function of strategy (speed or accuracy), task [reading (○) or naming (■)] and target-distractor relation (congruent, neutral, or conflict)

The first set of analyses aimed to verify the effectiveness of the instructions in manipulating subjects' response strategies. Analyses of the response times for correct responses as well as the number of errors both showed a significant main effect for strategy ($F_{1,45} = 23.38$, $MSe = 30040.10$, $p < .001$ and $F_{1,45} = 26.57$, $MSe = 1.05$, $p < .001$, respectively). Subjects in the speed group responded faster and also made more errors over-all than subjects in the accuracy group. The mean response times were 541.81 msec. and 641.63 msec., and the mean numbers of errors out of 40 trials were 0.78 and 0.15 for the speed and accuracy groups, respectively. This indicated that the instructions successfully induced different response strategies in the subjects.

TABLE 1
MEAN RESPONSE TIMES AND STANDARD DEVIATIONS BY STRATEGY, TASK, AND TARGET-DISTRACTOR RELATION

Task		Accuracy			Speed		
		Congruent	Neutral	Conflict	Congruent	Neutral	Conflict
Naming	*M*	709.21	706.95	819.41	575.79	609.45	725.37
	SD	90.61	83.23	97.65	61.00	65.34	84.87
Reading	*M*	537.37	541.11	535.76	447.51	442.69	450.03
	SD	85.15	90.85	80.85	56.86	47.57	58.35

The next set of analyses[3] focused on response times to check for a number of findings that are typical of a Stroop color-naming task. This was done to validate the general experimental procedure. If the typical findings were also found in the present experiment, the experiment could be said to be procedurally comparable to previous Stroop experiments. This multivariate analysis yielded a significant interaction for task × relation ($F_{2,90} = 170.15$, $MSe = 686.36$, $p < .001$), a significant main effect for relation ($F_{2,90} = 215.28$, $MSe = 554.39$, $p < .001$), and a significant main effect for task ($F_{1,45} = 887.19$,

[3]Results based on the response times are reported hereafter. Further analyses of errors were omitted because error rates were very low and variance was lacking in some conditions.

$MSe = 3133.35$, $p < .001$). Four simple main effects for relations were also examined for each task in each strategy group. There were no significant relation effects for the reading tasks in either strategy group ($F_{2,44} = 1.65$ and $F_{2,46} = .80$ for the speed and the accuracy groups, respectively). In contrast, the relation effects were significant for the naming tasks in both groups ($F_{2,44} = 146.18$, $MSe = 968.83$, and $F_{2,46} = 90.97$, $MSe = 1090.32$, $p < .001$, for the speed and the accuracy groups, respectively). These results are consistent with the typical Stroop findings.

TABLE 2
MEAN NUMBER OF ERRORS AND STANDARD DEVIATIONS BY STRATEGY, TASK, AND TARGET-DISTRACTOR RELATION

Task		Accuracy			Speed		
		Congruent	Neutral	Conflict	Congruent	Neutral	Conflict
Naming	*M*	0.13	0.00	0.67	0.17	0.39	3.91
	SD	0.34	0.00	1.17	0.57	0.78	2.99
Reading	*M*	0.00	0.08	0.00	0.00	0.00	0.17
	SD	0.00	0.41	0.00	0.00	0.00	0.64

Finally, the analyses focused on the naming tasks to assess whether there was a significant congruency effect or interference effect in the two strategy groups and whether the two groups differed in the two kinds of Stroop effects. Multivariate analysis of variance involving strategies and relations showed a significant two-way interaction ($F_{2,90} = 5.42$, $MSe = 1030.92$, $p < .01$), indicating that the Stroop effects did vary with subjects' response strategy. When the interference effect and the congruency effect were examined separately, the interaction of strategy × relation was not significant in the analysis that included the neutral and the conflict conditions ($F_{1,45} = .05$), but it was significant in the analysis that included the neutral and the congruent conditions ($F_{1,45} = 16.90$, $MSe = 448.66$, $p < .001$). The magnitude of the interference effect was 115.92 msec. (or 19.02% increase over the neutral condition) and 112.46 msec. (or 15.91% increase over the neutral condition) for the speed and the accuracy groups, respectively. The magnitude of the congruency effect was −33.66 msec. (or 5.52% decrease over neutral) and 2.26 msec. (or 0.32% increase over neutral) for the speed and accuracy groups, respectively.

DISCUSSION

Present results showed that the Stroop congruency effect was observed when speed was emphasized but not when accuracy was emphasized in subjects' response strategy. This result is consistent with the idea that the congruency effect may not always have been found in past studies because in some cases subjects tended to be overcautious. The presence of conflict trials among neutral and congruent trials would make subjects aware that the first response that comes to mind is not necessarily the correct response. A natural extra step would be to double-check a potential response before it is articulated. This step would add time to the response process and cancel out any savings in processing time on the congruent trials.

Consider whether an explanation of the present results can be framed in terms of the evaluation mechanism that Warren and Lasher (1974) proposed to explain how subjects arrive at a correct response. The mechanism consists of two types of response check, one against a list of permissible responses (a set check), the other a source check. The set check rejects a response that does not belong to the target set, whereas the source check rejects a response that belongs to the target set but is not the target response on a given trial (in effect, the source check resolves conflict). The fact that the manipulation of response strategy in the present experiment affected the congruency effect but not the interference effect suggests that different evaluation processes may underlie responses to congruent stimuli and responses to conflict stimuli. According to this idea, when subjects confront a conflict stimulus, they typically apply both a set check and a source check whether they have adopted a speed strategy or an accuracy strategy. On the other hand, when subjects confront a congruent stimulus, those adopting the accuracy strategy would still apply both types of check, whereas those adopting the speed strategy might bypass the source check and apply only the set check, at least on some of the trials. The consequence of applying a source check, in addition to a set check, on a congruent trial is that it consumes more processing time and hence obscures a possible congruency effect.

Such an explanation, however, begs a critical question. The explanation seems to presume that subjects know ahead of time what type of stimulus will appear on a given trial; yet, given the random mixture of different types of stimuli in the present experiment, how could they know this without having gone through the response-evaluation process? One possibility is that the response-evaluation process can give partial information about whether there is a competing response. When there is such a conflict, the source of each response could be checked. When there is no competing response, depending on the response strategy they adopt, the subjects would either output the only available response immediately or conduct a source check to be sure.

An alternative explanation might be based on a spreading activation network, in particular, the Parallel Distributed Processing (PDP) model (Rumelhart, McClelland, & the PDP Research Group, 1986). Cohen, *et al.* (1990) recently applied a PDP model to account for the Stroop effect. In their account, word reading and color naming constitute two processing pathways that differ in strength (with word reading being the stronger one). The two pathways share the same output module (verbal response), where they intersect in a Stroop color-naming task. When both pathways activate the same output unit, as in the Stroop congruent condition, activation accumulates faster to exceed the response threshold than when only one pathway activates the output unit or when the two pathways activate two different output units. The researchers were able to simulate the congruency effect as well as other basic Stroop phenomena such as the faster reading than naming time, the interference effect, and the greater magnitude of the interference effect than the congruency effect. However, the simulations assumed a blocked presentation of different types of Stroop stimuli, as opposed to the mixed presentation employed in the present experiment. Processes are likely to be different in a blocked presentation than in a mixed presentation. Moreover, although the PDP model allowed task demands (i.e., whether a color-naming or a reading response was called for) to modulate the initial activation level of the network, response strategy was not considered as a possible factor that could change the pattern of activation. Presumably, response strategy would be handled by adjusting the response threshold of an output unit and the parameters of the activation function.

There is an increasing tendency in the literature to use the congruent condition as the control to assess interference when the first of a series of experiments does not detect the congruency effect (e.g., McClain, 1983; Virzi & Egeth, 1985). However, lack of the congruency effect in one experiment does not justify abandoning the neutral condition and using the congruent condition as a control in subsequent experiments. The null result is likely to be unstable across experiments and subjects unless researchers specifically manipulate response strategy. As pointed out earlier, using the congruent condition as a control could lead to a spurious interference effect or inflate its magnitude at the least. It seems more appropriate to check on the congruency effect when the interference effect is being assessed. Furthermore, given the interference effect and the congruency effect were both observable under the speed strategy in the present experiment, explicit emphasis should be placed on response speed if researchers wish to study both effects.

REFERENCES

Anderson, J. R. (1983) *The architecture of cognition.* Cambridge, MA: Harvard Univer. Press.

Broadbent, D. E. (1958) *Perception and communication.* New York: Pergamon.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990) On the control of automatic processes: a parallel distributed processing account of the Stroop effect. *Psychological Review*, 97, 332-361.

Collins, A. M., & Loftus, E. F. (1975) A spreading-activation theory of semantic processing. *Psychological Review*, 82, 407-428.

Dalrymple-Alford, E. C. (1972) Associative facilitation and interference in the Stroop color-word task. *Perception & Psychophysics*, 11, 274-276.

Deutsch, J. A., & Deutsch, D. (1963) Attention: some theoretical considerations. *Psychological Review*, 70, 80-90.

Dyer, F. N. (1973) The Stroop phenomenon and its use in the study of perceptual, cognitive, and response processes. *Memory & Cognition*, 1, 106-120.

HINTZMAN, D. L., CARRE, F. A., ESKRIDGE, V. L., OWENS, A. M., SHAFF, S. S., & SPARKS, M. E. (1972) "Stroop" effect: input or output phenomenon. *Journal of Experimental Psychology*, 95, 458-459.

HOCK, H. S., & EGETH, H. (1970) Verbal interference with encoding in a perceptual classification task. *Journal of Experimental Psychology*, 83, 299-303.

KLEIN, G. (1964) Semantic power measured through the interference of words with color naming. *American Journal of Psychology*, 77, 576-588.

MCCLAIN, L. (1983) Stimulus-response compatibility affects auditory Stroop performance. *Perception & Psychophysics*, 33, 266-270.

MORTON, J., & CHAMBERS, S. M. (1973) Selective attention to words and colors. *Quarterly Journal of Psychology*, 25, 387-397.

NEILL, W. T. (1977) Inhibitory and facilitatory processes in selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, 3, 444-450.

NORMAN, D. A. (1968) Toward a theory of memory and attention. *Psychological Review*, 75, 522-536.

RUMELHART, D. E., MCCLELLAND, J. L., & THE PDP RESEARCH GROUP. (1986) *Parallel distributed processing: explorations in the microstructure of cognition*. Vol. 1. *Foundations*. Cambridge, MA: MIT Press.

SEYMOUR, P. H. K. (1977) Conceptual encoding and locus of the Stroop effect. *Quarterly Journal of Experimental Psychology*, 29, 245-265.

STIRLING, N. (1979) Stroop interference: an input and output phenomenon. *Quarterly Journal of Experimental Psychology*, 31, 121-132.

STIRLING, N., & COLTHEART, M. (1977) Stroop interference in a letter naming task. *Bulletin of the Psychonomic Society*, 10, 31-34.

STROOP, J. R. (1935) Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643-661.

VIRZI, R. A., & EGETH, H. E. (1985) Toward a translational model of Stroop interference. *Memory & Cognition*, 13, 304-319.

WARREN, R. E., & LASHER, M. D. (1974) Interference in a typeface variant of the Stroop test. *Perception & Psychophysics*, 15, 128-130.

ZAJANO, M. J., HOYCEANYLS, E. M., & OUELLETTE, J. F. (1981) A confound in the standard control condition of the Stroop experiment. *Perceptual and Motor Skills*, 53, 835-841.

Accepted July 1, 1991.