

# Strategies in the color-word Stroop task

GORDON D. LOGAN and N. JANE ZBRODOFF  
*University of British Columbia, Vancouver, British Columbia, Canada*

and

JAMES WILLIAMSON  
*Erindale College, University of Toronto, Toronto, Ontario, Canada*

**Three experiments examined subjects' ability to strategically influence interference from conflicting color words while naming colors. Contingencies introduced between words and colors allowed subjects to predict the value of the reported (color) dimension when given the value of the unreported (word) dimension and thereby to speed processing. Strategies to exploit these contingencies were adopted when only two colors were used and buttonpress responses (Experiment 1) and vocal responses (Experiment 2) were required. When four colors were used (Experiment 3), there was no evidence of the strategy. Implications for the interpretation of existing Stroop data are discussed.**

When people report one dimension of a multidimensional stimulus, they often suffer interference if unreported dimensions specify related but conflicting meanings. For example, reaction time to name colors is increased substantially when the colors are presented as words that specify conflicting color names ("red" written in blue ink; Stroop, 1935). It is widely believed that this Stroop interference is strategy invariant, resulting from automatic processing of the unreported dimension (e.g., Posner & Snyder, 1975) or from the failure of selective attention to exclude the unreported dimension (e.g., Treisman, 1969). This view is challenged by recent evidence (Logan, 1980; Logan & Zbrodoff, 1979) that Stroop interference may be strategy dependent. It was found to vary as the relative frequency of stimuli with conflicting and compatible dimensions was manipulated. When conflicting stimuli were more frequent than compatible ones so that the reported and unreported dimensions were likely to have opposite meanings, conflicting stimuli were processed faster than compatible ones, reversing the usual Stroop effect. The reversal was stronger the greater the relative frequency of conflicting trials (i.e., the greater the cue validity; Logan & Zbrodoff, 1979), and was stronger for dimensions that were processed faster than for dimensions that were processed more slowly (Logan, 1980). Apparently, subjects adopted a strategy that took advantage of the predictive information in the frequency manipulation;

This research was supported by Grant UO 053 from the Natural Science and Engineering Research Council of Canada to Gordon D. Logan. N. Jane Zbrodoff was supported by a doctoral fellowship from the Humanities and Social Sciences Research Council of Canada. Requests for reprints should be addressed to Gordon D. Logan, Department of Psychology, University of British Columbia, Vancouver, British Columbia V6T 1W5, Canada.

they attended to the unreported dimension as well as to the reported one in deciding upon a response.

Such strategic flexibility has been found in Stroop-like tasks in which words that specify spatial positions are reported and the words are presented in positions in the viewing field that conflict with their meaning (Logan & Zbrodoff, 1979), in tasks in which the relative position of two objects is reported and the objects appear in a position in the viewing field that conflicts with the relative judgment (Logan, 1980), and in tasks in which simple arithmetic equations are verified (e.g.,  $3 + 4 = 7$ ;  $3 \times 4 = 12$ ) and false equations are presented that would be true for the opposite operation (e.g.,  $3 \times 4 = 7$ ;  $3 + 4 = 12$ ; Zbrodoff, Logan, & Barber, 1984). The experiments reported here were designed to extend these findings to the usual color-word Stroop task.

Three experiments were conducted in which the color of color words was reported. Each experiment involved two groups of subjects, one for which colors and words were compatible on 80% of the trials and conflicted on 20% (the 20% conflicting group) and one for which colors and words were compatible on 20% of the trials and conflicted on 80% (the 80% conflicting group). In each group, each word appeared with only two colors, the compatible one on compatible trials and a single conflicting one on conflicting trials (i.e., when the word "blue" conflicted, it was always written in green). Thus, the words could predict the colors. If subjects were sensitive to these predictive relations, reaction time should be facilitated for the frequent or expected stimulus type and inhibited for the infrequent or unexpected stimulus type. The main effect of compatibility should be enhanced in the 20% conflicting group (which could expect compatible stimuli) and diminished or reversed in the 80% conflicting group (which could expect conflicting stimuli), producing an interaction between compatibility and frequency or groups. This interaction was the

index of strategic flexibility, and was expected to vary in magnitude between experiments if some conditions were more (or less) susceptible to strategic intervention than others.

Experiment 1 used two colors and required a button-press response to indicate which color had been presented. The button-press response was used because previous experiments demonstrating a reversed Stroop effect also used a button-press response (e.g., Logan & Zbrodoff, 1979).

Experiment 2 used two colors, but required a vocal response to the colors (i.e., the colors were named aloud). Since vocal responses are usually used in Stroop studies, this procedure provides a more appropriate test of the generality of strategy-dependent effects. Moreover, the contrast between Experiments 1 and 2 bears on the importance of perceptual confusion and response competition as determinants of Stroop interference (see Dyer, 1973). If response competition is the only factor, no interference would be found in Experiment 1, in which the button-press responses were not habitually associated with color names. If both response competition and perceptual confusion are important, less interference would have been found in Experiment 1 than in Experiment 2, and if perceptual confusion is the only factor, interference would have been of the same magnitude in Experiments 1 and 2.

Experiment 3 used four colors and required vocal responses. The contrast between performance with two colors in Experiment 2 and four colors in Experiment 3 bears on the importance of capacity limitations in strategy-dependent and strategy-invariant effects. To fully exploit the predictive information available in the frequency manipulation, subjects must keep in mind one expectancy for each color-word contingency as well as expectancies for the stimulus-response contingencies described in the instructions. Since four extra expectancies should be harder to keep in mind than two, expectancy effects (i.e., the compatibility  $\times$  frequency interaction) might have been smaller in Experiment 3 than in Experiment 2. By contrast, strategy-invariant effects are generally not affected by capacity limitations (Logan, 1979; Posner & Snyder, 1975), and should have been the same in Experiments 2 and 3. That is, the main effect of compatibility should have been the same in the two experiments. Indeed, Golden (1974) found that Stroop interference was not affected by the number of colors employed.

Since the three experiments were variations of a common theme, they will be described in one Method section. Departures from the general method will be noted where appropriate.

## METHOD

### Subjects

Thirty-six undergraduate students from Erindale College participated voluntarily or to fulfill course requirements. Six subjects were assigned randomly to each group in each experiment.

### Apparatus and Stimuli

The stimuli were the words "blue," "green," "red," and "gold," made from blue, green, red, and gold Letraset (PT103)

and mounted on white cards to appear centered in the viewing field of a Scientific Prototype three-channel tachistoscope (Model GB). At a viewing distance of 122 cm, each letter subtended  $.24 \times .28$  deg of visual angle. The length of each word in degrees of visual angle was .82 for "red," 1.12 for "blue," 1.19 for "gold," and 1.49 for "green."

Each trial began with a black fixation point centered in a white field, which was replaced by a blank white field for a 500-msec foreperiod, which itself was replaced by the color word for that trial, exposed for 500 msec. After the color word had been exposed, the fixation field returned and remained on until the next trial. Luminance of all three fields was matched at 10 fL.

In all experiments, reaction time was measured in milliseconds from the onset of the color word. In Experiment 1, the reaction time clock stopped when the subject pressed one of two buttons attached to microswitches and mounted on a panel in front of him or her. Pressing each button also illuminated a separate light visible to the experimenter so that response accuracy could be monitored. In Experiments 2 and 3, the reaction time clock stopped when the subject spoke into a microphone attached to a voice operated relay.

### Procedure

In each experiment, each subject reported the colors of color words on 240 trials. For the 20% conflicting groups, 192 of the trials involved compatible stimuli and 48 involved conflicting stimuli. For the 80% conflicting group, 48 of the trials involved compatible stimuli and 192 involved conflicting stimuli. Within compatible and conflicting stimulus types, each selected color-word token was used equally often. The color-word tokens assigned to each subject were chosen from the set of all possible color-word combinations within the following constraints: (1) Each color was assigned to two words, one with which it was compatible and one with which it conflicted. This allowed the possibility of a predictive relation between dimensions in the 80% conflicting conditions: For each color, the compatible combination was used on 20% of the trials that used that color, and the conflicting combination was used on 80%. Thus, given the word, the probability of the conflicting color was 0.8. Note that the same considerations apply to compatible stimuli in the 20% conflicting conditions. (2) Across subjects, all possible color-word combinations were used equally often. (3) When only two of the four colors were used (Experiments 1 and 2), each possible pair of colors was used equally often. Within these constraints, each subject received the 240 trials in a separate random order.

The instructions described the sequence of events on a trial and the response required of the subject. The subjects were shown examples of stimuli and stimulus sequences. The subjects were told the relative frequency of conflicting trials that they would experience (expressed to them as a percentage: 20% or 80%), and were asked to respond as quickly and as accurately as possible. Each experiment involved two blocks of 120 trials, with breaks between blocks if the subjects wished. No practice was given.

## RESULTS AND DISCUSSION

### Experiment 1

Mean reaction times for correct responses in each combination of compatibility and frequency conditions were computed for each subject. The means across subjects and the mean error rates (expressed as proportions) appear in Table 1.

It is clear from Table 1 that the compatibility effect was modulated by the frequency of conflicting trials. When 20% of the trials were conflicting, reaction times to conflicting stimuli were 96 msec slower than reaction times to compatible stimuli. But when 80% of the trials were conflicting, reaction times to conflicting stimuli

were 7 msec faster than reaction times to compatible stimuli. This reversal was apparent in the individual data of four of the six subjects in the 80% conflicting group (note that the traditional view of Stroop performance holds that none of the subjects should have shown a reversal).

These observations were supported by analysis of variance. The main effect of compatibility [ $F(1,10) = 5.44$ ,  $p < .05$ ] and the interaction between compatibility and frequency (i.e., groups) [ $F(1,10) = 7.42$ ,  $p < .05$ ] were both significant. For reasons not readily apparent to us, the effect of groups was marginally significant [ $F(1,10) = 3.90$ ,  $p < .08$ ], reflecting faster performance in the 80% conflicting group.

The error rates, also presented in Table 1, reflect the trends in the reaction time data.

### Experiment 2

Mean reaction times for correct responses in each combination of compatibility and frequency conditions were computed for each subject. Means across subjects appear in Table 1.

In this experiment, compatibility effects were modulated by frequency once again, but there was no reversal of the Stroop effect. When 20% of the trials were conflicting, reaction times to conflicting stimuli were 176 msec slower than reaction times to compatible stimuli. When 80% of the trials were conflicting, the difference dropped to 56 msec. This reduction was apparent in the individual subject data. Only one subject in the 80% conflicting group showed a compatibility effect larger than the smallest compatibility effect observed in the 20% conflicting group.

These observations were supported by analysis of variance. The main effect of compatibility was significant [ $F(1,10) = 37.74$ ,  $p < .01$ ], as was the interaction between compatibility and frequency (i.e., groups) [ $F(1,10) = 10.27$ ,  $p < .01$ ]. In this experiment, the ten-

dency for the 80% group to be faster did not approach significance [ $F(1,10) = 2.12$ ,  $p > .10$ ].

The error rates, also presented in Table 1, reflect the trends in the reaction time data.

The contrast between Experiments 1 and 2 bears on the importance of response competition in determining the Stroop effect. In particular, the buttonpress responses of Experiment 1, assigned arbitrarily to colors, should not have been subject to response competition as much as should the vocal responses of Experiment 2, habitually associated with specific colors. There was less Stroop interference in Experiment 1 than in Experiment 2 (56 msec vs. 116 msec, on average), but the result is difficult to evaluate because reaction times were considerably faster in Experiment 1 than in Experiment 2. Thus, Experiment 2 may have produced more interference because there was more time for the decision to be influenced by the unreported dimension, quite apart from possible competition (see Logan, 1980). An analysis of variance comparing the two experiments showed a significant interaction between compatibility and experiments [ $F(1,20) = 7.05$ ,  $p < .05$ ], suggesting differential response competition, but also showed a significant main effect of experiments [ $F(1,20) = 25.96$ ,  $p < .01$ ], suggesting different opportunities for influence. Interestingly, there was no interaction among experiments, compatibility, and frequency conditions [ $F(1,20) < 1$ ], suggesting that strategic effects were equally strong in the two experiments.

### Experiment 3

Mean reaction times for correct responses in each combination of compatibility and frequency conditions were computed for each subject, and the means across subjects appear in Table 1.

In contrast with the previous experiments with two colors, the compatibility effects with four colors were relatively unaffected by the frequency of conflicting trials. When 20% of the trials were conflicting, reaction times to conflicting stimuli were 102 msec slower than reaction times to compatible stimuli; when 80% of the trials were conflicting, reaction times to conflicting stimuli were still 89 msec slower than reaction times to compatible stimuli. Analysis of variance of these data showed a significant main effect of compatibility [ $F(1,10) = 39.69$ ,  $p < .01$ ] and a marginal (unexplained) main effect of frequency [ $F(1,10) = 4.59$ ,  $p < .06$ ], but no significant compatibility x frequency interaction [ $F(1,10) < 1$ ]. Subjects did not develop systematic expectations for particular stimuli as they had in previous experiments.

The error rates, presented in Table 1, reflect similar trends.

The contrast between Experiments 2 and 3 bears on the importance of capacity limitations in strategy-dependent and strategy-invariant effects. Experiment 3 required four expectations to be kept in mind simultaneously (i.e., one for each color-word contingency),

Table 1  
Reaction Times and Error Rates for Each Combination  
of Compatibility and Frequency Conditions  
in Experiments 1, 2, and 3

Description	Measure	20%		80%	
		Comp	Conf	Comp	Conf
Experiment 1					
2 Colors-Buttons	RT	474	570	393	386
	Error	.021	.058	.033	.018
Experiment 2					
2 Colors-Voice	RT	672	848	627	683
	Error	.027	.075	.023	.019
Experiment 3					
4 Colors-Voice	RT	729	831	631	720
	Error	.010	.078	.043	.010

Note—Comp = compatible trials; Conf = conflicting trials; 20% and 80% refer to the frequency of conflicting trials.

whereas Experiment 2 required only two, so strategy-dependent effects should have been more constrained by capacity limitations in Experiment 3 than in Experiment 2. But since strategy-invariant effects are not subject to capacity limitations, they should have been roughly equal in the two experiments. The data supported both predictions: The strategy-dependent effects, indicated by the compatibility  $\times$  frequency interaction, were clearly weaker in Experiment 3 than in Experiment 2. However, the strategy-invariant effects, indicated by the compatibility main effect, were about the same in the two experiments (116 msec in Experiment 2 vs. 96 msec in Experiment 3). In an analysis of variance comparing the two experiments, the interaction among experiments, compatibility, and frequency was significant [ $F(1,20) = 5.01, p < .05$ ], indicating stronger strategy-dependent effects in Experiment 2 than in Experiment 3, but the interaction between compatibility and experiments was not [ $F(1,20) < 1$ ], indicating equivalent strategy-invariant effects.

### GENERAL DISCUSSION

The present experiments showed that interference in the color-word Stroop task may be strategy-dependent. In the experiments that used two colors (Experiments 1 and 2), interference decreased as the frequency of conflicting trials increased; with buttonpress responses (Experiment 1), several subjects in the 80% condition actually reversed the Stroop effect. This corroborates previous findings with other Stroop-like tasks (Logan, 1980; Logan & Zbrodoff, 1979; Zbrodoff et al., 1984) and suggests that the strategies seen in those contexts may be quite general.

The failure to find strategy-dependent effects with four colors (Experiment 3) is an important result because it indicates a limitation on the use of strategies. To perform the task successfully, four stimulus-response contingencies had to be kept in mind (one for each color). To exploit the predictive information in the frequency manipulation, four color-word contingencies had to be kept in mind as well, and, apparently, this was too much; the subjects abandoned the strategy. Indeed, all of the previous experiments showing reversals with Stroop-like tasks used two responses and required only two additional contingencies to be kept in mind to exploit the predictive information. Thus, strategy-dependent Stroop interference may not be observed in studies that use four or more stimulus-response contingencies (i.e., studies in which four or more response classes exist). Since many Stroop tasks are designed this way, strategy-dependent interference may not pose a serious problem in interpreting existing data. However, two-response studies should be interpreted with caution.

The task environment imposes a second important limitation on strategy-dependent effects: It must contain the predictive information necessary to support the strategy. The previous Stroop-like tasks in which reversals were observed used binary dimensions so that increasing the frequency of compatible or conflicting stimuli made one value of the reported dimension more likely given the value of the unreported dimension (i.e., the same value was more likely when compatible stimuli were

frequent, and the opposite value was more likely when conflicting stimuli were frequent). Many-valued dimensions, such as color, do not necessarily provide predictive information when stimulus frequencies are varied. Clearly, when compatible stimuli are more frequent, the value of the unreported dimension makes one value of the reported dimension most likely. However, this is not the case when conflicting stimuli are more frequent; many-valued dimensions can conflict in many ways. Some constraints, such as in the present experiments, must be introduced before the frequency of conflicting stimuli can convey predictive information.

Finally, a general implication of the present experiments and the previous ones is that performance may be jointly determined by strategy-dependent and strategy-invariant processes. In every experiment demonstrating strategic variability in Stroop interference, there was an overall effect of compatibility, indicating a strategy-invariant component. Possibly, no task is completely immune to strategic effects and no task is completely free of strategy-invariant effects. An effective approach to a new task is to try new combinations of old abilities until one is found that works appropriately (Logan, 1979). Performance based on this approach would reflect a blend of influences; the strategy-invariant properties of the individual abilities and the strategy-dependent properties of their combination. Much of the interest in information processing has focused on the separate (strategy-invariant) abilities; little is known about the processes that organize them. It may be important to understand the strategies available to subjects in our tasks, and the effects the strategies may have on performance. This kind of approach has proved useful in understanding problem-solving behavior (e.g., Simon, 1975). Perhaps it is time we recognized that information-processing tasks require problem solving as well.

### REFERENCES

- DYER, F. N. (1973). The Stroop phenomenon and its use in the study of perceptual, cognitive, and response processes. *Memory & Cognition*, 1, 106-120.
- GOLDEN, C. J. (1974). Effect of differing number of colors on the Stroop color and word test. *Perceptual and Motor Skills*, 39, 550.
- LOGAN, G. D. (1979). On the use of a concurrent memory load to measure attention and automaticity. *Journal of Experimental Psychology: Human Perception and Performance*, 5, 189-207.
- LOGAN, G. D. (1980). Attention and automaticity in Stroop and priming tasks: Theory and data. *Cognitive Psychology*, 12, 523-553.
- LOGAN, G. D., & ZBRODOFF, N. J. (1979). When it helps to be misled: Facilitative effects of increasing the frequency of conflicting stimuli in a Stroop-like task. *Memory & Cognition*, 7, 166-174.
- POSNER, M. I., & SNYDER, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola Symposium*. Hillsdale, NJ: Erlbaum.
- SIMON, H. A. (1975). The functional equivalence of problem solving skills. *Cognitive Psychology*, 7, 268-288.
- STROOP, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643-662.
- TREISMAN, A. M. (1969). Strategies and models of selective attention. *Psychological Review*, 76, 282-299.
- ZBRODOFF, N. J., LOGAN, G. D., & BARBER, C. Y. (1984). *On the autonomy of mental processes: A case study of arithmetic*. Manuscript submitted for publication.

(Manuscript received for publication November 30, 1983.)