Memory & Cognition
1998, 26 (5), 1033-1040

Differential components of the manual and vocal Stroop tasks

DINKAR SHARMA
University of Kent at Canterbury, Canterbury, England

and

FRANK P. McKENNA
University of Reading, Earley Gate, England

In this study, four components of the Stroop effect were examined for manual word and vocal responses. The components were lexical, semantic relatedness, semantic relevance, and response set membership. The results showed that all four components were present in the vocal response task. However, in the manual word response task, the only component that produced significant interference on its own was response set membership. These results do not support predictions made by recent translation models (see W. R. Glaser & M. O. Glaser [1989] and Sugg & McDonald [1994]). A possible solution was suggested that located two sites for Stroop interference. The lexical, semantic relatedness, and semantic relevance effects were located in the lexical system, whereas the response set membership effect was located at a response selection stage.

Over the years, the Stroop task (Stroop, 1935) has played an important role in cognitive psychology. The robust nature of the paradigm has provided a fertile testing ground for investigations of selective attention, automaticity, word-reading, and color-naming processes (Dyer, 1973; Jensen & Rohwrer, 1966; MacLeod, 1991). The task involves the color naming of words presented in incongruent colors. For example, the word *blue* may be presented in red ink. There are a number of variants on how the task is implemented. For example, the response may be executed by manual or vocal output. As MacLeod has noted, this manipulation does make a difference, with manual output producing less interference. To illustrate further the importance of response output, we present in Table 1 the results from six studies. It is clear from these six studies that, when the response required is vocal, the Stroop interference (difference in reaction times between incongruent and control stimuli) is much larger (297 msec) than when the response required is manual (168 msec). Lupker and Katz (1981) have argued that Stroop interference can occur at four possible stages: (1) at input, where the stimulus information is perceptually analyzed; (2) at a decision stage; (3) at a response selection stage; and (4) at a response output stage. The fact that Stroop interference is affected by response modality has led to the proposal that interference should be located at the response output or response selection stage (La Heij, 1988; La Heij, Van der Heijden, & Schreuder, 1985).

Correspondence concerning this article should be addressed to D. Sharma, Department of Psychology, University of Kent, Canterbury, Kent CT2 7NP England (e-mail: d.sharma@ukc.ac.uk).

—**Accepted by previous associate editor Michael E. J. Masson**

One major limitation on the conclusion that the Stroop interference is reduced with a manual or nonverbal response is that it ignores the fact that there are several components to the Stroop effect. Klein (1964) has shown that Stroop interference is not only found when the distractor word and target ink are incompatible but may also be found with other written stimuli as distractors. Klein used four ink colors (red, yellow, green, and blue) to present all the stimuli in a list format. When compared with nonsense syllables, the largest interference was found when the distractor word was one of the colors from the response set (an increase of 31.79 sec for a total of 80 stimuli). If the distractor was a color word not in the response set (such as the word *purple*), interference was only 12.49 sec. When the word (such as *fire*) implicated other colors, interference was 9.75 sec, and when the word was a common English word (such as *friend*), interference was 5.27 sec. This finding shows that there is a semantic gradient in the Stroop effect, such that, as the semantic relationship between the word and ink color increases, so also does the magnitude of the Stroop interference.

A closer look at Klein's (1964) study reveals that the observed semantic gradient can readily be described by a number of different interference effects. First, there is a lexical component—that is, any word that is in the lexicon will show longer color-naming latencies than will items not in the lexicon (such as strings of nonsense syllables). This we refer to as the lexical effect. Although we do not concentrate on this issue in the present article, it is possible that the lexical effect is itself made up of a number of other components, involving both orthographic and phonological components. For example, there may be interference in using repeated letter strings, as compared with a color patch, or in using nonpronounceable words,

Table 1
Comparison of the Results of Several Color-Naming Stroop Studies That Have Manipulated the Type of Word Stimulus, Involving Neutral, Color Incongruent, and Control (Repeated Letters or Symbols) Stimuli

	Vocal Response			Manual Response		
	Control	Neutral	Incongruent	Control	Neutral	Incongruent
McClain (1983)	640	716	884	808	769	912
Redding and Gerjets (1977)	741	824	918	769	754	867
Klein (1964)	621	699	1,018			
Fox, Shor and Steinman (1971)	598	695	966			
Virzi and Egeth (1985)				1,360	1,335	1,565
Hock and Egeth (1970)				1,278	1,264	1,538
Mean	650	734	947	1,053	1,031	1,221

Note—In the case of experiments using a list format for presentation of stimuli, reaction times have been translated to milliseconds per stimuli to allow direct comparison across experiments. Two of these studies (McClain [1983] and Redding & Gerjets [1977]) have also manipulated response output (vocal or manual responding).

as compared with random letter strings, and so forth. Second, there is the semantic relatedness component—that is, any word that is semantically related to the target colors will show an interference effect. For example, the word *fire* is semantically related to the color red and will therefore show greater interference than a semantically unrelated word such as *top*. Third, there is a response set membership component—that is, any word that is also part of the set of possible responses will show greater interference than words not part of the response set. For example, if two responses are to be made, red and green, an irrelevant word that is either green or red will show greater interference than an irrelevant word, such as *gold*, that is not part of the response set. Neumann (1980) and La Heij et al. (1985) have argued for a fourth component in the Stroop effect, which they call semantic relevance. Semantic relevance refers to that part of the Stroop interference that is due to the fact that the color words are also relevant in a task in which subjects have to name ink colors. For example, there will be greater interference with irrelevant words that are color words (e.g., *gold*) than with semantically related noncolor words (e.g., *fire*).

The fact that there are at least four components in the Stroop effect has implications for the general conclusion that response output has an effect on Stroop interference. We can ask, for example, whether all four components of the Stroop effect are affected by response output or whether only some of the components are affected. As yet, we are not aware of any reported study that has looked at the effects of response output on the four components of the Stroop effect. Therefore, to illustrate our point, we would like to consider two aspects of Stroop interference that can be easily extracted from existing data: the lexical component (that is, the difference in color naming between a neutral word and a nonsense syllable) and the other three components (semantic relatedness, semantic relevance, and response set membership), which we collectively refer to as the semantic components (that is, the difference between the incongruent color word and a neutral word). The lexical and semantic components can be calculated from Table 1.

Table 1 clearly shows that the manipulation of response modality has a major influence on the lexical component of Stroop interference, such that there are substantial lexical effects (neutral − control stimuli) with a vocal response (84 msec) but not with a manual response (−22 msec). Although a similar conclusion could be made about the semantic components (incongruent − neutral stimuli), the evidence is less clear. Table 1 shows that vocal responses do produce a larger interference (213 msec) than do manual responses (190 msec). However, this conclusion needs to be considered with caution since, in two studies that have compared vocal and manual responses in the same experiment (McClain, 1983; Redding & Gerjets, 1977), the semantic component was essentially the same for vocal (131 msec) as for manual responses (128 msec).

The question therefore arises as to how to explain the lexical effect found for the vocal response but not for the manual response. To do this, we shall draw on one of the translation models, as proposed by W. R. Glaser and M. O. Glaser (1989). In this model, there are three main assumptions: (1) Semantic memory and the lexicon are separate systems; (2) words have a privileged (i.e., a more direct) access to the lexicon, whereas colors have a privileged access to the semantic system; and (3) interference occurs according to a dominance rule—that is, "Stroop inhibition occurs only if the distractor has privileged perceptual access to the subsystem that is critical for response selcction" (W. R. Glaser & M. O. Glaser, 1989, p. 30).

In describing the workings of the model, W. R. Glaser and M. O. Glaser (1989) trace out a pathway for each of the two dimensions of the stimulus—one for the color and one for the word. The pathway involves the flow of information from stimulus input to response output. When a verbal response is required to the color, the color traces a pathway that initially activates the concept nodes in the semantic system, followed by the word nodes in the lexical system, and then the vocal output system. Similarly, when reading a word, the word stimulus traces out a pathway that mainly involves the lexical system and the vocal output system. However, there may be some activation in the semantic system. Thus, there are three sites at which interference can occur—the lexical system, the vocal output system, and possibly the semantic system. Words cannot interfere with colors in the semantic system due to the dominance rule; thus, interference could occur in the lexical or the vocal output system. In describing their model, W. R. Glaser and M. O. Glaser (1989) assume that interference occurs only in the lexical system and not in the vocal output system. For example, when presented with an incongruent stimulus (the word *blue* written in *red* ink), there is activation of word nodes in the lexical system from two sources—one from the pathway traced by the ink color, which is required to make a response, and the second by the word, which is to be ignored. Naming the ink color (*red*) involves activation in the lexical system from two sources: (1) the spread of activation from the semantic nodes for the color *red*, and (2) response-set membership, which indicates that words that are also part of the set of responses will receive greater activation than those words not part of the response set. In addition, there is activation in the lexical system from the word *blue*. There are three sources for this activation: (1) direct perception of the word (*blue*); (2) response set membership; and (3) spread of activation from the semantic nodes related to the color (*red*). Given that activation for the to-be-ignored word is higher than for the ink color, W. R. Glaser and M. O. Glaser assume that inhibition is the result of the extra time required to resolve this conflict.

When a manual response is required, the pathway for color naming does not enter the lexical system but is directed from the semantic system to a manual output system. The pathway for the words does enter the lexical system and then also the semantic system. However, the words cannot interfere with the color pathway in the semantic system due to the dominance rule. Thus, the lexical effect disappears with a manual response. This model completely describes why there are lexical effects for vocal responses and no lexical effects for manual responses.

What implications are there for the W. R. Glaser and M. O. Glaser (1989) model with respect to the other components of the Stroop effect? The way the model is set up, it would predict that the semantic relatedness, response set membership, and semantic relevance components are a product of interference in the lexical system. An increase in semantic relatedness, response set membership, and semantic relevance would increase the level of activation of the irrelevant word nodes in the lexical system. This additional activation would increase the time needed to make the correct response to the relevant ink color. Thus, the prediction of the model is that, when color naming with a vocal response, there will be interference due to the lexical, semantic relatedness, response set membership, and semantic relevance components. However, when color naming with manual responses, there will not be any interference of any of the four components. This prediction is not in accord with the available data, which shows that there is Stroop interference in color naming with manual responses (McClain, 1983; Pritchatt, 1968; Redding & Gerjets, 1977; Sugg & McDonald, 1994).

A solution to this problem may be possible if one modifies the W. R. Glaser and M. O. Glaser (1989) model, as suggested by Sugg and McDonald (1994). In reviewing the literature, Sugg and McDonald emphasize the importance of two types of manual responses: manual word responses, when the response keys are labeled with words, and manual color responses, when the response keys are labeled with color patches. In the Stroop task, subjects ignore the word and respond to the ink color. Sugg and McDonald refer to this as the translated word response task, if a manual word response task is used to make a response. When a manual color response is used, they refer to this as the untranslated color response task. They show that, with a manual response, it is only in the translated word response task that the Stroop effect is found (about 100 msec in magnitude; see Sugg & McDonald, 1994, Figure 3); the untranslated color response task does not show any Stroop effect (McClain, 1983; Virzi & Egeth, 1985).

Sugg and McDonald (1994) explain this difference between the two manual response tasks by modifying the model of W. R. Glaser and M. O. Glaser (1989). They suggest that, rather than both types of manual responses being output via the semantic system, as in the original model, the translated word response task is output via the lexical system, and the untranslated color response task is output via the semantic system. This modified model can explain why there are Stroop effects with a translated word response task and not with an untranslated color response task. That is, the effects directly parallel the Stroop effects with the vocal response task, in that interference is located in the lexical system. If this is the case, the model also predicts that the four components of the Stroop effect should be present in a manual translated word response task, as they are with the vocal response task. However, this prediction of the Sugg and McDonald model is not in accord with the absence of lexical effects in the manual word response task (see Table 1). Since it is not possible to make a definitive statement from the available literature, because no study has directly looked at all these components in vocal and manual tasks, it is possible that lexical effects are present in the translated word response task but not in the untranslated color response task. However, what is clear is that Sugg and McDonald's model makes very clear predictions in terms of the four components in the translated word response task.

The general aim of the present experiment was to consider the effect of response output on the different components of the Stroop effect and, more specifically, to test the following prediction of the model modified by Sugg and McDonald (1994)—namely, that all four components of the Stroop effect will be present in both the vocal task and the manual translated word response task. If all four components are present in both tasks, this will support Sugg and McDonald's modified model; however, if the four components are not present or only some of the components are present in the manual translated word response task, Sugg and McDonald's modified model will not be supported and may itself need modifying.

METHOD

Subjects

A total of 40 University of Reading students volunteered to take part in the experiment, for which they were paid. All the subjects had English as their mother tongue.

Design

The design of the experiment formed a 2 × 5 × 5 factorial model, with response output (vocal vs. manual), word type (letter strings, neutral words, color related words, color words [nonresponse], and incongruent words) and block as within-subjects factors.

Each of the five word types were presented in blocked format and in counterbalanced order, using a Latin square design (except for the letter strings, which were presented first in each of the four stimulus orders). For half the subjects, the vocal response task was completed before the manual word response task, and for the other half, the order was reversed.

Materials

The words and letter strings used were all presented in capital letters: letter strings: XXXX, HHHH, OOOOO, and PPPPP; neutral words, TOP, CHIEF, CLUB, and STAGE; color-related words, FIRE, GRASS, SKY, and BEAR; color words (nonresponse); PURPLE, GREY, GOLD, and YELLOW; incongruent color words, RED, GREEN, BLUE, and BROWN. The neutral and incongruent words were equated for word length and frequency (Kučera & Francis, 1967). The color-related words and color words (nonresponse) could not be matched for word length and frequency. The mean frequency of the four word conditions are 160.5, 88.75, 50, and 158, for the neutral, color-related, color word (nonresponse) and incongruent conditions, respectively. All stimuli were presented using a Victor V286A PC computer. Subjects sat approximately 60 cm from the computer screen, with the dimensions of each word being 0.6 cm (0.6° of visual angle) high and approximately 2 cm (2° of visual angle) wide.

Procedure

Each of the four color words was combined with the corresponding incongruent ink colors—red, green, blue, and brown—to produce 12 incongruent stimuli. These 12 stimuli were randomized, with two restrictions: that consecutive trials did not repeat the same ink color or word and that the word did not repeat itself on consecutive trials as either the word or the ink color. For example, the word *red* written in green ink would not be followed by the word *red* or by a word written in red or green ink. This formed one block in the stimulus array; five such blocks were formed to produce 60 stimuli that were presented individually on a white background. An identical procedure was used to produce 60 stimuli for the other word and letter conditions.

The subjects were introduced to the task as a color perception task in which they would be presented a word in one of four ink colors at the center of the screen. Their task was to ignore the words and respond to the ink colors as quickly and as accurately as possible. Each stimulus remained on the screen until a response was made, followed by an interstimulus interval of 2 sec.

There were three stages to the experiment. In the first stage, the subjects were familiarized to the four ink colors used throughout the experiment. They were shown 20 nonletter symbols in the four ink colors. The second stage involved two sessions of practice. The stimuli used for practice were nonletter symbols, ####, which varied in length from three to six characters. Sixty such stimuli were presented during each practice phase. In the manual task, during the first practice phase, the subjects were asked to learn the stimulus– response relationship such that they would not have to look at the response buttons every time they made a response. This was done without any time pressure. During the second phase, the subjects were told not to look at the buttons while making a response, unless it was absolutely necessary, and to make the responses as quickly and as accurately as possible. A short break was given between each phase.

In the vocal task, responses were made by speaking into a headset microphone, which triggered a voice key. Because the voice trigger is very sensitive to any sound, all the subjects were asked to be very careful in what they said during the experiment so as not to trigger the voice switch by utterances that were not responses to the stimuli (e.g., coughing, laughing, saying "um", "er", "oops," etc.). All the subjects received practice in using the microphone, any problems being ironed out by the experimenter. This practice phase lasted for about 5 min. During the experiment, a tape recorder was turned on to record subjects' verbal responses. This helped later to locate at which point in the experiment errors were made.

In the third stage, the various word conditions were presented in counterbalanced order across subjects. Before each condition, the subjects were informed that letter strings or real words were going to be presented (the difference between the words was not mentioned to the subjects). All the subjects were instructed to ignore the word/letter stimuli and report only the ink colors as quickly and as accurately as possible.

In the manual word response condition, the subjects pressed one of four black colored buttons. Each button was labeled with one of four words written in black ink—BLUE, BROWN, RED, and GREEN. All the subjects positioned their index and middle fingers from their left and right hands on top of each of the buttons. Half the subjects received the red and green labels on the left hand and the blue and brown labels on the right hand, whereas for the other half the order was reversed.

RESULTS

Analysis of Errors

Errors were classified as any incorrect response made by the subject. However, in the vocal task, some of these errors also included other types of errors: vocal stumbles and equipment problems (e.g., when the microphone did not register subjects' responses). Overall the numbers of errors made in the vocal and manual word response tasks were very small (1.78% and 3.11%, respectively). A 2 × 5 repeated measures analysis of variance (ANOVA) was conducted. The two factors were response output and word type. The analysis showed that there was a main effect of response output [$F(1,39) = 29.5$, $MS_e = 2.17$, $p < .01$]. This indicated that the number of errors made in the vocal condition was significantly lower than in the manual condition. This is as expected, since making vocal responses is a more practiced task than making manual responses. The main effect of word type was also significant [$F(4,156) = 2.84$, $MS_e = 1.66$, $p < .05$]. The mean percentage of errors for the five word type conditions are: letter, 1.98%; neutral, 2.27%; color related, 2.47%; color words (nonresponse), 2.42%; and incongruent, 3.10%. Post hoc analysis using the Tukey (HSD) multiple comparison test showed that only one pair of comparisons involved a significant difference: incongruent and letter conditions. This pattern of results shows that there is not a speed–accuracy tradeoff, since the incongruent condition produced the most errors and the longest reaction times.

The two-way interaction between response output and word type was not significant [$F(4,156) = 1.29$, $MS_e = 1.55$, $p > .2$]. Since the pattern of errors was the same for both response outputs, these results do not affect the interpretation of the differences in the mean reaction times for the vocal and manual word tasks described below.

Analysis of Reaction Times

All latency analyses were conducted on the mean correct reaction times. All observations were included when computing a subject's mean. The main issue addressed in this experiment was to determine whether there were any Stroop effects for the vocal and manual response outputs. To do this, a three-way ($2 \times 5 \times 5$) ANOVA was conducted. The three factors—response output, word type, and block—are all within-subjects factors.

We first concentrate on all main and interaction effects involving only response output and word type. The effects of block will be addressed as a separate issue at the end of this analysis. The analysis showed that there was a main effect of word type [$F(4,156) = 76.30$, $MS_e = 23{,}091.82$, $p < .001$] and no main effect of response output [$F(1,39) = 0.02$, $MS_e = 152{,}179.68$, $p > .8$]. However, of more interest is the significant interaction between word type and response output [$F(4,156) = 5.47$, $MS_e = 16{,}466.45$, $p < .001$; see Figure 1].

To investigate the two-way interaction between response output and word type, simple main effects analyses were conducted on vocal and manual word responses. In both vocal [$F(4,156) = 75.85$, $p < .001$] and manual word [$F(4,156) = 28.69$, $p < .001$] output, the simple main effect of word type was significant. Further post hoc analyses were conducted using the Tukey (HSD) multiple comparison test. For the vocal response, all pairwise comparisons were significantly different (critical value is 33.66 msec). For the manual task, this was not the case, although all pairwise comparisons with the incongruent condition were significant (critical value is 42.6 msec). All other comparisons were not significant, except for the comparison between color words (nonresponse) and letter word type conditions. More importantly, this analysis shows that, for the vocal response, there are four main components, which are all significant: lexical (neutral words − letters), semantic relatedness (color related − neutral words), semantic relevance (color words [nonresponse] − color related), and response set membership (incongruent words − color words [nonresponse]). However, for the manual response, there is only one component that is significant: response set membership (incongruent words − color words [nonresponse]).

Because of the theoretical importance of the null effects in this study, further analyses were conducted to illustrate the power of the statistics for finding significant differences when using a manual word response. These power calculations were based on advice given by Howell (1992), when using matched samples. It was assumed that, for each component, the expected mean of the difference scores would be the difference found for the respective components in the vocal response task. For the manual word response, the power for each of the four components was found to be as follows: lexical (0.77), semantic relatedness (0.95), semantic relevance (0.71),

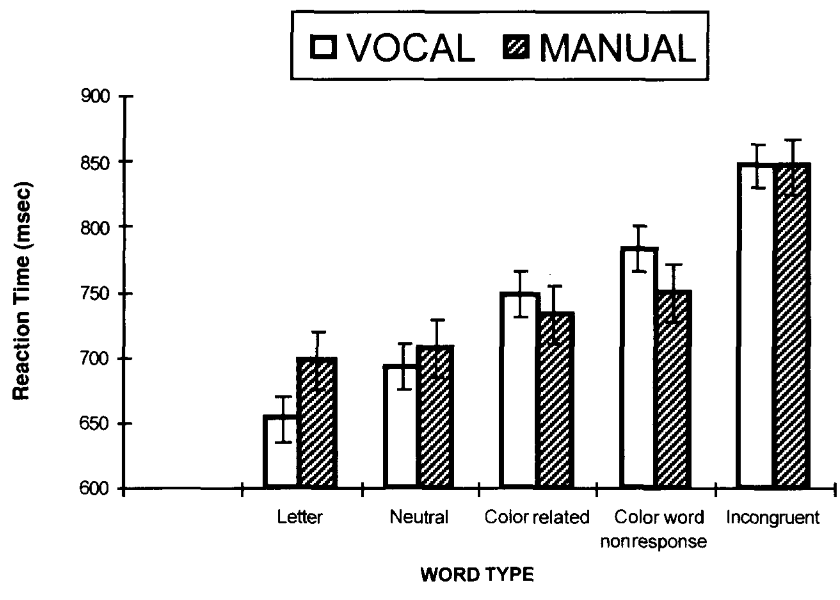

Figure 1: Mean reaction time to color name stimuli in the various word type conditions for vocal and manual word responses. The error bars represent confidence intervals, as described in Loftus and Masson (1994), for within-subjects designs.

and response set membership (0.97). All of these power calculations are moderate to large, as was suggested by Cohen (1988).

It is possible that one explanation for the difference in the Stroop effects found could be due to the different levels of practice subjects have had at identifying ink colors, using a vocal or a manual response. It could, therefore, be argued that the effect of practice would be larger for the manual response than for the vocal response. To address this issue, we looked at whether the factor for block interacted with response output. The analysis showed that there was no interaction between block and response output, either as a two-way interaction [$F(4,156) = 1.29$, $MS_e = 8{,}569.57$, $p > .2$], or as a three-way interaction with word type [$F(16,624) = 1.52$, $MS_e = 6{,}140.71$, $p = .09$]. However, there was an interaction between block and word type [$F(16,624) = 4.40$, $MS_e = 6{,}713.96$, $p < .001$]. Simple main effect analyses showed that there was a simple main effect of block for the color related [$F(4,156) = 2.82$, $p < .05$], color word (nonresponse) [$F(4,156) = 10.71$, $p < .001$], and incongruent [$F(4,156) = 6.84$, $p < .001$] word conditions but not for the letter [$F(4,156) = 2.18$, $p = .07$] or neutral word [$F(4,156) = 0.11$, $p > .9$] conditions. Tukey multiple comparison tests showed that the simple main effects of block were due to a significantly longer reaction time in the first block than in all other blocks. In the incongruent word type condition, the simple main effect of block was also due to faster reaction times in Block 4 than in all other blocks. Overall, these results show that, at least within this study, the effects of practice did not differentially affect performance in the vocal or manual tasks—although, with more extensive practice, it is possible that there may have been a difference.

DISCUSSION

Although Stroop interference has often been considered to be a unitary phenomenon, there is clear evidence that it may be more appropriate to consider the effect as consisting of several components. This point is well illustrated by examining the role of response output. Although there is a large overall Stroop effect (incongruent words − letter strings) for both vocal (194.2 msec) and manual word (148 msec) response outputs, the components that constitute these effects are different. For the vocal response, the Stroop interference is composed of a number of effects—lexical (40.4 msec), semantic relatedness (54.8 msec), semantic relevance (35.4 msec), and response set membership (63.6 msec). However, for manual word responses, the only component that produced a significant interference was response set membership (96.7 msec). There were no significant effects of lexical (9.6 msec), semantic relatedness (25.6 msec) or semantic relevance (16.2 msec). Although research has focused on the overall interference effect, the different components merit attention in their own right. For example, it is of considerable interest that there is a clear lexical effect with vocal responses but that this effect is eliminated with manual word responses. In addition, whereas it is clear that the overall interference is larger for the vocal than for the manual word response, this is not true for one of the components, where the response set membership effect is not larger for the vocal output (63.6 msec) than for the manual word output (96.7 msec).

Before addressing the theoretical implications of our findings, two other issues require some discussion. To what extent can our findings be accounted for by (1) the differential levels of practice in the manual word and vocal response tasks and (2) the fact that the manual word response task used in this experiment is part of a larger class of learned responses. It is evident that most individuals have had more practice at vocally naming a word or color than at manually responding to it. This preference in the initial correspondence between the type of response and the stimulus is thought to be one explanation for the Stroop effect. With extensive practice over a number of days on vocal responding to the ink color of a Stroop stimulus, it is generally found that the Stroop effect can be reduced but not completely eliminated (Dulaney & Rogers, 1994; MacLeod & Dunbar, 1988; Stroop, 1935). A similar finding was reported by Flowers and Stoup (1977) for a manual card-sorting task. One interesting question that relates to this paper is how practice affects each of the four components of the Stroop effect. What role does practice play in our experiment? We can address this question by observing the pattern of results across the five blocks of this study. The analysis showed that the reaction times were significantly longer in the first block than in any other block. This indicates that the effects of practice were limited to the first block. However, of particular importance is the finding that this pattern of results was the same for both manual word and vocal responses. We can therefore conclude that, in this study, practice does not differentially affect the vocal and manual word response tasks—although it is possible that, with more extensive practice, there may have been a different pattern of results for the vocal and manual word responses.

The second issue is concerned with the effect of using static button labels in a manual word response task. In the literature, this issue has been highlighted by MacLeod (1991) and Sugg and McDonald (1994). It has been suggested that the use of static button labels produces a response that relies on a covert verbal response more than it would if the button labels were changed from trial to trial. In their introduction, Sugg and McDonald reviewed some of the manual response studies and showed that this may have been part of the explanation for some Stroop interference in the color response task. The question of using static button labels in the word response task is more problematic, because, as the buttons are already labeled with a word, this should encourage verbal responding. Although it may be possible to reduce this attachment to a verbal response by changing the word labels on each trial, the question arises as to whether this is appropriate in this study. Since our main comparison is with a vocal response, it could be argued that a vocal response also relies on a static label and, therefore, that the more direct comparison is with a static word label rather than with a changing word label. Assuming that a static word label may help the attachment of a covert verbal response, what effect can this have on the various components of the Stroop effect? It could be argued that this can lead to increased interference in the lexical, semantic relatedness, and semantic relevance components of the Stroop effect. To the extent that, in our study, these components did not show any significant effects, one could argue that any attachment of a covert verbal label in a word response task is small.

The present results partially support both W. R. Glaser and M. O. Glaser's (1989) model and Sugg and McDonald's (1994) model of the Stroop effect. First, the fact that there is a lexical effect in the vocal task and no lexical effect in the manual task supports W. R. Glaser and M. O. Glaser's model. However, the model cannot explain why there should be a response set membership effect with manual word responses. Second, the fact that there is a response set membership effect with manual word responses supports Sugg and McDonald's model, but it cannot explain why there are no lexical, semantic relatedness, and semantic relevance effects. We agree with Sugg and McDonald that the model needs to explain why there is a Stroop effect (or more precisely, a response set membership effect) in the manual word response task, but we would argue that the model presented by Sugg and McDonald does not provide a satisfactory solution. Such a model must be able to explain the response set membership effect but, in addition, be able to explain why there are no lexical, semantic relatedness, and semantic relevance effects with manual output.

We now propose a possible explanation for these results, using largely W. R. Glaser and M. O. Glaser's (1989) original model. The results suggest to us some modifications. Since there are differences between the various Stroop components when responses are made either vocally or manually, this suggests different sites for interference. First, why is there no lexical effect in the manual word response task but one in the vocal task? As mentioned earlier, W. R. Glaser and M. O. Glaser's model can explain this difference as being due to the privileged access of words in the lexical system. This is theoretically very important, since it has implications for response conflict models of the classic Stroop effect. As Stirling (1979) has noted, some response conflict models argue that the interference is due to two incompatible responses being activated (one from the color and the other from the word). The fact that a noncolor word can produce interference indicates that response incompatibility is not a necessary condition for interference. In addition, the fact that the lexical effect disappears with manual output suggests that the locus of the lexical effect may be the lexical or vocal output system (W. R. Glaser & M. O. Glaser, 1989). Using the same mechanism, W. R. Glaser and M. O. Glaser's model can explain why there are no semantic relatedness or semantic relevance effects in the manual task.

The only component of the present results that this model cannot explain is the response set membership effect, which is found in both vocal and manual (translated word response) tasks. Can these findings be accommodated within the Sugg and MacDonald (1994) model? The main feature of Sugg and MacDonald's model is that there are different types of processing for the word response task and for the color response task. In particular, a word response has privileged access to the lexical system, whereas a color response has privileged access to the semantic system. One reason for positing this differential processing was to explain the finding that there are Stroop effects in the translated word response task and not in the untranslated color response task. However, our results suggest that differential processing occurs for the translated word response *only* when incongruent color words are part of the set of possible responses. For all the other conditions, W. R. Glaser and M. O. Glaser's (1989) model is sufficient to explain the data. Although this may provide a possible explanation, it undermines the elegance of Sugg and MacDonald's model. In particular, it is not clear why manual word responses should have privileged access to the lexical system only when the stimulus word is part of the set of responses.

An alternative suggestion is that Sugg and McDonald's (1994) model is incorrect in allowing word responses to have direct access to the lexical system. Our findings suggest an amendment that locates the site for response set membership in a different location. Since many of our findings can be adequately explained within W. R. Glaser and M. O. Glaser's (1989) model, we assume, unlike Sugg and MacDonald, that all manual responses (whether they are labeled with a word or a color patch) have privileged access to the semantic system and not to the lexical system. Our results suggest to us that the response set membership effect found in the translated word response task is due to interference at a late response selection stage rather than to interference in the semantic or lexical systems. In particular, we assume that, when translating the stimulus into a response, the translation will be based in some way on the type of response. This can also be thought of in terms of dimensional overlap or set level compatibility between the stimulus and response (Kornblum, Hasbroucq, & Osman, 1990; Kornblum & Lee, 1995; Wang & Proctor, 1996). In the word response task, this translation may be based on some sort of intermediate *word-like* code, since the response keys are labeled with words. This is less likely in the color response task, since response keys are labeled with color patches. In addition, we assume that, when attempting to select a response, there will be competition from other activated responses. This competition will be greater between responses that have translation codes of similar types. When presented with irrelevant word stimuli that are also part of the response set, they will provide activation for the irrelevant response. In the word response task, competition between the irrelevant and relevant responses will be greater when the irrelevant stimulus is a member of the response set. Our findings suggest that the competition between alternative responses is not based on a semantic code, but is more likely to be based on an identity code. Only when the irrelevant stimuli are identical to one of the responses do we find any Stroop interference. However, the precise nature of the translation code and the bases for competition between responses must await further investigations.

In general, these results support the importance of using response output as a method for investigating the nature of the Stroop effect. Our results particularly illustrate the importance of detailed investigations into the various components that make up the Stroop effect. We have shown that vocal and manual (translated word response) outputs have different effects on the various components of the Stroop effect. In this way, detailed predictions from Sugg and McDonald's (1994) model were tested. The model was found not fully to explain the data. Some modification was required to accommodate the results. First, it was assumed that all word response tasks have privileged access to the semantic system and not to the lexical system, and second, that the response set membership effect was located at a response selection stage. As a result, this paper supports two sites for Stroop interference—interference in the lexical system, which accounts for the lexical, semantic relatedness, and semantic relevance effects, and interference in a later response selection stage, which accounts for the response set membership effect. Such a mechanism would bring closer together W. R. Glaser and M. O. Glaser's (1989) model and earlier translation models that located the site of Stroop interference at a central or more specialized response stage (Palef & Olson, 1975; Virzi & Egeth, 1985).

REFERENCES

COHEN, J. (1988) *Statistical power analysis for behavioral sciences* (2nd ed.). New York: Academic Press.

DULANEY, C. L., & ROGERS, W. A. (1994). Mechanisms underlying reduction in Stroop interference with practice for young and old adults. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 470-484.

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

FLOWERS, J. H., & STOUP, C. M. (1977). Selective attention between words, shapes and colors in speeded classification and vocalization tasks. *Memory & Cognition*, **5**, 299-307.

FOX, L., SHOR, R. E., & STEINMAN, R. J. (1971). Semantic gradients and interference in naming color, spatial direction, and numerosity. *Journal of Experimental Psychology: General*, **91**, 59-65.

GLASER, W. R., & GLASER, M. O. (1989). Context effects in Stroop-like word and picture processing. *Journal of Experimental Psychology: General*, **118**, 13-42.

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

HOWELL, D. C. (1992). *Statistical methods for psychology* (3rd ed.). Boston: PWS-Kent.

JENSEN, A. R., & ROHWER, W. D., JR. (1966). The Stroop color–word test: A review. *Acta Psychologica*, **25**, 36-93.

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

KORNBLUM, S., HASBROUCQ, T., & OSMAN, A. (1990). Dimensional overlap: Cognitive basis for stimulus–response compatibility—A model and taxonomy. *Psychological Review*, **97**, 253-270.

KORNBLUM, S., & LEE, J.-W. (1995). Stimulus–response compatibility with relevant and irrelevant stimulus dimensions that do and do not overlap with the response. *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 855-875.

KUČERA, H., & FRANCIS, W. (1967). *Computational analysis of present-day American English*. Providence: Brown University Press.

LA HEIJ, W. (1988). Components of Stroop-like interference in picture naming. *Memory & Cognition*, **16**, 400-410.

LA HEIJ, W., VAN DER HEIJDEN, A. H., & SCHREUDER, R. (1985). Semantic priming and Stroop-like interference in word-naming tasks. *Journal of Experimental Psychology: Human Perception & Performance*, **11**, 62-80.

LOFTUS, G. R., & MASSON M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, **1**, 476-490.

LUPKER, S. J., & KATZ, A. N. (1981). Input, decision, and response factors in picture–word interference. *Journal of Experimental Psychology: Human Learning & Memory*, **7**, 269-282.

MACLEOD, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, **109**, 163-203.

MACLEOD, C. M., & DUNBAR, K. (1988). Training and Stroop-like interference: Evidence for a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 126-135.

MCCLAIN, L. (1983). Effects of response type and set size on Stroop color–word performance. *Perceptual & Motor Skills*, **56**, 735-743.

NEUMANN, O. (1980). *Selection of information and control of action*. Unpublished doctoral dissertation, University of Bochum, Bochum, Germany.

PALEF, S. R., & OLSON, D. R. (1975). Spatial and verbal rivalry in a Stroop-like task. *Canadian Journal of Psychology*, **29**, 201-209.

PRITCHATT, D. (1968). An investigation into some of the underlying associative verbal processes of the Stroop color effect. *Quarterly Journal of Experimental Psychology*, **20**, 351-359.

REDDING, G. M., & GERJETS, D. A. (1977). Stroop effects: Interference and facilitation with verbal and manual responses. *Perceptual & Motor Skills*, **45**, 11-17.

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

STROOP, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, **18**, 643-662.

SUGG, M. J., & MCDONALD, J. E. (1994). Time course of inhibition in color-response and word-response versions of the Stroop task. *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 647-675.

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

WANG, H., & PROCTOR, R. W. (1996). Stimulus–response compatibility as a function of stimulus code and response modality. *Journal of Experimental Psychology: Human Perception & Performance*, **22**, 1201-1217.

(Manuscript received May 20, 1996;
revision accepted for publication May 16, 1997.)