Journal of Experimental Psychology:
Human Perception and Performance
1998, Vol. 24, No. 3, 978–992

0096-1523/98/$3.00

Stroop-Type Interference: Congruity Effects in Color Naming With Typewritten Responses

Gordon D. Logan and N. Jane Zbrodoff
University of Illinois at Urbana-Champaign

Four experiments examined the Stroop effect with typewritten responses. Experiment 1 compared vocal, arbitrary-keypress, and typewritten responses and found the largest Stroop effect for typewritten responses. The effect appeared in the latency to type the first keystroke and not in the duration of the typing response. Experiment 2 compared normal (name the color) and reverse (name the word) Stroop effects with typewritten responses and found that the normal Stroop effect was much larger than the reverse Stroop effect. Experiment 3 compared typing the entire color name with typing its first letter and found equivalent Stroop effects in the 2 conditions. Experiment 4 varied the relative frequency of congruent and incongruent trials and found that the typewritten Stroop effect was larger when congruent trials outnumbered incongruent trials. The results are related to theories of the Stroop effect and theories of language production.

Stroop (1935) found that subjects took longer to name the color of the ink in which color words were written than to name the color of the ink in which control stimuli were written. This so-called *Stroop effect* has been replicated hundreds of times, and current theories are still striving to explain it (for a review, see MacLeod, 1991). An important issue, theoretically and empirically, concerns the effect of response modality in the Stroop task. Stroop's investigation and the majority of the replications required vocal responses; subjects said the color names out loud. Most of the other replications used arbitrary-keypress responses; subjects pressed keys that were associated with color names (e.g., Keele, 1972; Pritchatt, 1968; Sugg & McDonald, 1994; Virzi & Egeth, 1985). The Stroop effect is found with both response modalities. Studies that compared response modalities often found a stronger Stroop effect with vocal responses than with arbitrary keypresses (Logan, Zbrodoff, & Williamson, 1984; Majeres, 1974; McClain, 1983; Melara & Mounts, 1993; Neill, 1977; Redding & Gerjets, 1977; Simon & Sudalaimuthu, 1979; White, 1969), though some found no difference between response modalities in the magnitude of the effect (Roe, Wilsoncroft, & Griffiths, 1980).

Response modality effects bear on important issues in the Stroop literature, including the factors responsible for the Stroop effect, the role of automaticity and modularity in the Stroop effect, and the locus of the Stroop effect. The present experiments were intended to address these issues by investigating the Stroop effect with typewritten responses; subjects saw colored words, as in the usual Stroop task, and responded by typing the names of the colors on a keyboard. Typewritten responses are similar to vocal ones, in that they are automatic in skilled typists, and they are similar to arbitrary keypress responses, in that they involve the manual response modality. These contrasting similarities allow us to unravel differences between vocal and arbitrary-keypress responses that were confounded in previous investigations and thereby to shed some light on the controversies surrounding the modality effect.

Translation Versus Automaticity

The response modality effect may reflect something special about the relations between printed words and vocal responses, or it may reflect a greater degree of automaticity for vocal responses to words than for arbitrary keypresses to words. *Translation theories* of the Stroop effect appear to endorse the first position (e.g., Glaser & Glaser, 1989; Sugg & McDonald, 1994; Virzi & Egeth, 1985), whereas *automaticity theories* would endorse the second (Cohen, Dunbar, & McClelland, 1990; Logan, 1980; MacLeod & Dunbar, 1988).

Translation Theories

Translation theories argue that the relation between stimulus and response (S-R) modalities is an important determinant of the Stroop effect. Printed words have special access to spoken words because they are in the same (linguistic) modality; words are mapped directly onto vocal responses. Perceived colors and spoken words are in different modalities, and a translation process is required to map colors onto vocal responses (Glaser & Glaser, 1989; Sugg & McDonald, 1994; Virzi & Egeth, 1985). These assumptions lead to several predictions that have been confirmed in the literature.

This research was supported by National Science Foundation Grant SBR 94-10406. We are grateful to Julie Delheimer for recruiting and testing the subjects and to Gracetrue Kim for a discussion that suggested the utility of examining the Stroop effect with typewritten responses. A policy enacted by the Council of Editors of the American Psychological Association prevents us from thanking Tom Carr for his help with this article.

Correspondence concerning this article should be addressed to Gordon D. Logan or N. Jane Zbrodoff, Department of Psychology, University of Illinois, 603 East Daniel Street, Champaign, Illinois 61820. Electronic mail may be sent to glogan@s.psych.uiuc or to jzbrodo@s.psych.uiuc.edu.

First, they account for the normal Stroop effect with vocal responses, in which subjects report colors and ignore words: Irrelevant words have direct access to vocal responses, and if they are color names, they create conflict with the vocal responses that subjects are required to make to the colors. Some sort of response selection or inhibitory process must be invoked to resolve the conflict. Second, translation theories account for the lack of interference in the *reverse Stroop task* with vocal responses, in which subjects report words and ignore colors: Colors do not have direct access to vocal responses, so they do not generate response tendencies that conflict with vocal responses to the words. Third, and most relevant to our purposes, they account for the response modality effect: Words do not have privileged access to arbitrary-keypress responses, so they are unlikely to generate responses that compete with the required responses to the colors. Words and colors both require translation before they access arbitrary-keypress responses.[1]

Automaticity Theories

Automaticity theories argue that the Stroop effect depends on the relative automaticity with which the to-be-reported and to-be-ignored dimensions are processed (Cohen et al., 1990; Logan, 1980; MacLeod & Dunbar, 1988). Automaticity determines the extent to which stimuli activate responses unintentionally; the greater the degree of automaticity, the stronger the activation of the response. These assumptions account for many Stroop phenomena as well.

First, they account for the normal Stroop effect with vocal responses by arguing that vocal responses to words are more automatic than vocal responses to colors. Thus, incongruent color words activate responses that conflict with the deliberately selected responses to the colors. Second, the same reasoning accounts for the lack of interference in the reverse Stroop task: Vocal responses to colors are not automatic enough to be generated unintentionally, so they are not activated enough to interfere with deliberate responses to words. Third, and most relevant, automaticity theories account for the response modality effect by arguing that arbitrary-keypress responses to words are less automatic than vocal responses to words, so words are less likely to activate keypress responses that compete with the deliberately selected responses to the colors.[2]

Typewritten Responses

Translation theories and automaticity theories appear to differ in their predictions about the Stroop effect with typewritten responses. Automaticity theories predict a Stroop effect in skilled typists because typewritten responses should be strongly associated with words. The Stroop effect with typewritten responses should be as strong as the Stroop effect with vocal responses as long as typewritten responses are sufficiently automatic; the Stroop effect with typewritten responses should be much stronger than the Stroop effect with arbitrary keypresses because the former should be much more automatic than the latter. By contrast, translation theories predict a weaker Stroop effect with typewritten responses than with vocal responses because typewritten responses, like arbitrary keypresses, are in a different modality than printed words.

Locus of the Stroop Effect

The locus at which the Stroop effect occurs has been an important issue in the Stroop literature. Some researchers have argued that it occurs early in processing, during perceptual encoding (Hock & Egeth, 1970). Others have argued for a more central locus, involving translation between codes (Glaser & Glaser, 1989; Kornblum, Hasbroucq, & Osman, 1990; Sugg & McDonald, 1994; Treisman & Fearnley, 1969; Virzi & Egeth, 1985). Still others have argued for a later locus, in the response selection stage (Cohen et al., 1990; Duncan-Johnson & Kopell, 1980, 1981; Logan, 1980; Morton, 1969; Morton & Chambers, 1973; Posner & Snyder, 1975; Warren & Marsh, 1979). There is some support for each position, but the bulk of the evidence points to a locus in response selection. Most likely, the effect has several loci.

The contrasts between typewritten responses, vocal responses, and arbitrary-keypress responses allow us to ask a new question about the locus of the effect, distinguishing between response selection and execution. Response selection processes may be similar with different response types, but response execution processes are likely to be distinct. If the Stroop effect occurred entirely in processes at or before response selection, we would expect a Stroop effect of the same magnitude regardless of response type. However, if execution processes were involved, we might expect Stroop effects of different magnitudes with different response types, depending on the extent to which the conflict was resolved in the execution processes. The current evidence for a response modality effect is consistent with the latter hypothesis, although factors like compatibility, automaticity, and translation may also be responsible for the differences.

Typewritten responses provide separate measures of response selection and execution processes. The latency to the first keystroke reflects response selection, whereas the dura-

[1] Translation theories make several more subtle predictions that have been confirmed at least partially (see, e.g., Glaser & Glaser, 1989; Sugg & McDonald, 1994; Virzi & Egeth, 1985). These predictions do not address the modality effects, with which we are primarily concerned, so we chose not to discuss them.

[2] Automaticity theories make more subtle predictions that have been confirmed in previous research. MacLeod and Dunbar (1988) offered the most compelling evidence for automaticity theories, training subjects to respond with color names when given arbitrary shapes. Early in practice, when shape naming was not automatic, incongruent colors interfered with shape naming but incongruent shapes did not interfere with color naming. After 5 sessions of practice, when shape naming achieved some degree of automaticity, mutual interference was observed; incongruent colors interfered with shape naming and incongruent shapes interfered with color naming. After 20 sessions of practice, when shape naming was more automatic than color naming, incongruent shapes interfered with color names interfered with color naming but incongruent colors did not interfere with shape naming.

tion of the typing responses—the time between the first and the last keystroke—reflects response execution primarily (Salthouse, 1986; Sternberg, Monsell, Knoll, & Wright, 1978). If the Stroop effect results only from processing at or before response selection, then it should appear only in initial latency. If the Stroop effect results only from execution processes, subjects should take longer to type color names in the incongruent condition than in the neutral or the congruent conditions, but initial latency should not differ. If the Stroop effect results from response selection and execution, both initial latency and duration should be longer with incongruent stimuli.

Modularity

The contrast between response types and the contrast between initial latency and duration with typewritten responses allow us to address a broader issue in the Stroop literature and the literature on language production that concerns the modularity of response processes. Theorists often distinguish between *modular* processes and *interactive* processes. Modular processes are encapsulated, operating independently of other concurrent or prior processes, whereas interactive processes are influenced by concurrent and prior processes as they unfold over time (see, e.g., Fodor, 1983).

Theories of the Stroop effect appear to assume, implicitly at least, that response execution processes are modular with respect to response selection and prior processes (e.g., Cohen et al., 1990; Logan, 1980; Phaf, van der Heijden, & Hudson, 1990). They assume that the perception of words and colors results in activating nodes that represent words and colors and that selection of responses results from activating nodes that represent responses. Perceptual processes prior to color and word nodes are not included in the models; neither are motor processes subsequent to response nodes. The Stroop effect is predicted entirely in terms of the flow of activation between color and word nodes on the one hand and response nodes on the other. Prior perceptual processes and subsequent motor processes are assumed, implicitly, to play no role in the Stroop effect.

These assumptions about modularity are reasonable insofar as the models are able to account for Stroop phenomena, and MacLeod (1991) argued that they have been quite successful. However, past accomplishment does not guarantee future success, and it is possible that new data will challenge the modularity assumption. Indeed, an important direction for future research is to describe in more detail the processes involved in activating color and word nodes and the processes involved in translating activated response nodes into overt responses. One could imagine, for example, coupling a model of word perception, like that of McClelland and Rumelhart (1981), Seidenberg and McClelland (1989), or Coltheart, Curtis, Atkins, and Haller (1993), with a model of word production, like that of Dell (1986) or Levelt et al. (1991), to provide a more complete account of the Stroop effect. It is not clear whether the modularity assumption would remain valid in such a model. On the one hand, Levelt et al. argued that the processes that select words are separate from, and modular with respect to, the processes that utter them, but on the other hand, Dell argued that selection and execution processes are interactive (see also Dell & O'Seaghdha, 1991). Current models of typewriting assume a high degree of interactivity between selection and execution processes (Rumelhart & Norman, 1982).

The contrast between response types provides a test of modularity: If response execution processes are modular with respect to response selection and prior processes, then the Stroop effect should have the same magnitude regardless of response type. If response execution is not modular, then the magnitude of the Stroop effect should vary with response type. The observed differences between vocal and arbitrary-keypress responses are consistent with the latter hypothesis, but they may also be due to automaticity, compatibility, and translation processes.

The contrast between initial latency and duration in typewritten responses provides a clearer test of the modularity hypothesis: If response execution is modular with respect to response selection and prior processes, then the Stroop effect should appear entirely in initial latency and not at all in duration. If response execution is not modular, then the Stroop effect should appear in the duration of the typing response as well as in initial latency.

Note that the test of the modularity hypothesis addresses the same data as the test of the locus of the Stroop effect, that is, the contrast between the initial latency and the duration of the typing response. Thus, the tests are not independent. However, we believe that the hypotheses are not independent either. Rather, they are different perspectives on the same issue. We felt that each perspective was important enough to warrant separate discussion, even though they address the same data.

The Experiments

We conducted four experiments in which we examined the Stroop effect with typewritten responses. Experiment 1 investigated response type, comparing the magnitude of the Stroop effect with vocal responses, arbitrary keypresses, and typewritten responses. Experiment 2 compared normal and reverse typewritten Stroop effects, replicating the contrast between vocal and typewritten responses. Experiment 3 compared the magnitude of the typewritten Stroop effect when subjects typed the whole word and when they typed only the first letter. Experiment 4 examined the role of strategies in the typewritten Stroop effect, comparing conditions in which congruent trials predominated with conditions in which incongruent trials predominated. Each experiment examined initial latency and the duration of the typewritten response to test hypotheses about the locus of the Stroop effect and related hypotheses about modularity.

Experiment 1: Vocal Responses, Arbitrary Keypresses, and Typewritten Responses

The first experiment addressed response type. Subjects performed the Stroop task three times, once with vocal responses, once with arbitrary-keypress responses, and once with typewritten responses. The main purpose was to compare translation and automaticity as explanations of the difference between vocal and arbitrary-keypress Stroop effects. If the difference was due to response modality, then the typewritten Stroop effect should be no larger than the arbitrary-keypress Stroop effect, and both should be smaller than the vocal Stroop effect. If the difference was due to automaticity, then the typewritten Stroop effect should be as large as the vocal Stroop effect, and both should be larger than the arbitrary-keypress Stroop effect. A secondary purpose was to test hypotheses about modularity and the locus of the Stroop effect by comparing the effects of congruity on the initial latency and the duration of typewritten responses.

Method

Subjects. The subjects were 18 volunteers from the general university community who were paid $5 for their services. Their speed on a typing test (described below) ranged from 29.4 to 79.9 words per min (wpm), with a mean of 49.8 wpm. Their accuracy on the typing test, measured in terms of the percentage of words typed correctly, ranged from 81.8% to 97.3%, with a mean of 92.0%.

Apparatus and stimuli. The stimuli were presented on an Amdek Model 722 color monitor controlled by an IBM AT computer. Vocal responses were registered by a Scientific Prototype Voice Activated Relay. Keypress and typewritten responses were registered by the computer's keyboard. The stimuli were the words RED, GREEN, BLUE, YELLOW, and WHITE and strings of percent symbols (%) matched in length to the lengths of the color words. The stimuli were presented in red, green, blue, yellow, and white. The IBM colors, which were used to construct the stimuli, were 12, 10, 9, 14, and 15, respectively. In the vocal and typewritten response conditions, subjects responded by speaking or typing the names of the colors in which the stimuli appeared. In the arbitrary-keypress condition, subjects pressed Z for red, X for green, C for blue, V for yellow, and B for white. In the typewritten response condition, subjects were told to rest their fingers on the "home row." In the arbitrary-keypress condition, subjects were told to rest the middle and index fingers of their left hands and the index, middle, and ring fingers of their right hands lightly on the appropriate keys.

Each task (vocal responses, arbitrary keypresses, and typewritten responses) involved 240 trials divided into four sets of 60. Each set of 60 trials included 20 congruent trials, in which words matched the colors they appeared in (e.g., RED in red), 20 incongruent trials, in which words did not match the colors (e.g., RED in green), and 20 neutral trials, in which three to six colored percent symbols were presented. The incongruent trials used every combination of words and colors except the congruent combination. Each of the four blocks was randomized separately.

Each trial began with a fixation point (a period) presented for 500 ms in row 13, column 40 of the standard IBM 24 × 80 text screen. When the fixation point extinguished, it was replaced immediately with the Stroop stimulus for that trial (a colored word or a colored string of percent symbols), which began in row 13, column 38 of the IBM text screen. The Stroop stimulus was exposed until a response from the subject was registered, at which point the screen went blank. If the response was vocal, the computer waited for the subject to score the accuracy of his or her response (1 if correct, 2 if error). After the subject typed the number, the screen remained blank for a 1,500-ms intertrial interval. If the response was an arbitrary keypress, the 1,500-ms blank interval began as soon as the response was registered. If the response was typewritten, the 1,500-ms blank interval began as soon as the subject typed the "return" key after typing the word for that trial.

A typing test was given on the computer before and after the experimental trials were conducted. The test involved four texts adapted from Collier's (1995) book, *Border Collies;* the texts ranged from 111 to 117 words in length (see Appendix). Subjects received a different text on each test. Because there were four texts and 18 subjects, the assignment of texts to orders (first vs. second) was not completely balanced. Texts 1 and 3 occurred four times as the first test and five times as the second; Texts 2 and 4 occurred five times as the first test and four times as the second.

On a given test (i.e., the one before or the one after the experimental trials), subjects saw the text they would type twice, once so they could preread it, and once so they could type it. During the typing part of the test, the text remained on the screen, but the characters that the subject typed were not echoed to the screen. The computer measured the time from the first to the last keystroke, and the number of words in the text was divided by that time to estimate typing speed. Accuracy was scored by displaying a record of the keys that subjects typed on the screen and counting the number of words that contained errors. Speed and accuracy scores reported in the *Subjects* section were averaged over the two tests.

Procedure. Subjects were given one test of typing speed, then three experimental conditions of 240 trials each, and then another test of typing speed. They were tested individually in a small room. Subjects were allowed to rest briefly after every 80 trials in the Stroop tasks.

Subjects were given instructions relevant to each segment of the experiment just before the segment began, without being told about subsequent segments. Thus, subjects were first instructed about the typing-speed test, and then the test was administered. After that, they were instructed about the first condition of the Stroop task, and then they performed 240 trials in that condition. This procedure was repeated for the two other conditions, and then instructions for the final typing-speed test were given, followed by the speed test itself. The order in which subjects performed the three conditions was counterbalanced, with 3 subjects receiving each of the six possible orders of conditions.

Data analysis. Several aspects of the data were analyzed, including the mean and standard deviation of the latency to the first response (the only response in the vocal and arbitrary-keypress conditions), the accuracy of responding, and the mean and standard deviation of the time between the first and the last keystroke (duration) in the typewritten response condition. The data from the first three measures were subjected to 3 (task: vocal vs. arbitrary keypress vs. typewritten) × 3 (congruity: congruent, incongruent, neutral) analyses of variance (ANOVAs). The data from the duration measures in the typewritten response condition were subjected to one-way ANOVAs with congruity (congruent, incongruent, neutral) as the factor.

Results

The means and standard deviations of the latency to initiate the response and the accuracy of responses for each condition are presented in Table 1 as a function of congruity. Table 1 also contains the means and standard deviations of the durations of the typewritten responses as a function of congruity. The results of ANOVAs performed on these data are presented in Table 2.

Initial latency. Vocal responses (M = 800 ms) were faster than arbitrary keypresses (M = 884 ms) and typewritten responses (M = 893 ms), though the differences between

Table 1
Experiment 1: Means and Standard Deviations of Time to Initiate Response and Response Accuracy in Experimental Conditions, and Duration Data for Typewritten Responses

Measure	Congruent	Incongruent	Neutral
Vocal responses			
Time to initiate response			
M latency	757	912	732
SD	233	312	223
Response accuracy (%)	100	96	100
Arbitrary-keypress responses			
Time to initiate response			
M latency	823	961	868
SD	354	372	357
Response accuracy (%)	97	95	96
Typewritten responses			
Time to initiate response			
M latency	809	1,023	847
SD	201	222	206
Response accuracy (%)	95	93	95
Time between first & last keystrokes			
M duration	557	566	561
SD	208	204	203

them were not significant. The Stroop effect occurred in each condition. The difference between congruent and incongruent trials was largest with typewritten responses (214 ms), intermediate with vocal responses (155 ms), and smallest with arbitrary-keypress responses (138 ms). We tested the significance of these differences with nonorthogonal planned comparisons. The Stroop effect with typewritten responses was significantly larger than the one with vocal responses, $F(1, 68) = 7.06, p < .05, MSE = 2{,}219.35$, and significantly larger than the one with arbitrary-keypress responses, $F(1, 68) = 11.71, p < .01, MSE = 2{,}219.35$, but the difference in the Stroop effect with vocal versus arbitrary-keypress responses was not significant ($F < 1$).

Table 2
Analysis of Variance for Conditions in Experiment 1

Measure	Task (T)	Congruity (C)	T × C
Latency			
F	2.12	129.86**	7.41**
df	2, 34	2, 34	4, 68
MSE	66,564.91	3,539.75	2,219.35
SD of latency			
F	2.92	3.26	1.18
df	2, 34	2, 34	4, 68
MSE	111,026.71	8,784.84	7,946.08
Accuracy			
F	10.29**	21.21**	1.88
df	2, 34	2, 34	4, 68
MSE	20.21	4.29	7.04
Duration (typewritten responses)			
F		2.20	
df		2, 34	
MSE		146.02	
SD of duration			
F		0.28	
df		2, 34	
MSE		449.17	

$^*p < .05$. $^{**}p < .01$.

Researchers often divide the Stroop effect into facilitation and interference scores, computing the former by subtracting reaction times (RTs) in the congruent condition from RTs in the neutral condition and computing the latter by subtracting RTs in the neutral condition from RTs in the incongruent condition. When we performed these calculations, we found −25, 45, and 38 ms of facilitation for vocal responses, arbitrary keypresses, and typewritten responses, respectively. The negative facilitation with vocal responses is not typical but it is not unheard of in the Stroop literature (e.g., Vanayan, 1993). Moreover, it was not significant by Fisher's least significant difference (LSD) test. Differences had to exceed 31 ms to be significant at $p < .05$. We found larger facilitation with arbitrary keypresses than with vocal responses, which has precedents in the literature (Redding & Gerjets, 1977). It is interesting that the facilitation for arbitrary keypresses and typewritten responses was about the same magnitude.

We calculated interference scores, finding 180, 93, and 176 ms of interference for vocal responses, arbitrary keypresses, and typewritten responses, respectively. There are precedents in the literature for smaller interference with arbitrary-keypress responses than with vocal responses (Redding & Gerjets, 1977). It is interesting that the interference observed with typewritten responses was about the same magnitude as the interference with vocal responses.

Accuracy. The percentage of correct responses was highest for vocal responses ($M = 98.4$), intermediate for arbitrary keypresses ($M = 96.0$), and lowest for typewritten responses ($M = 94.6$). However, the accuracy with typewritten responses is not directly comparable with the other conditions. Subjects had to type an average of 4.6 characters correctly to register a correct typewritten response, whereas they only had to type 1 character to register a correct arbitrary-keypress response. If the keypresses that compose the typewritten response were independent of each other, we could calculate the accuracy of each individual response by taking the 4.6th root of the overall accuracy. This calculation yields a value of 98.8%, which is close to the value for vocal responses. This estimate must be viewed with caution because the individual keystrokes are unlikely to be independent, but it does suggest that accuracy per keystroke was higher than the overall accuracy measure implies it was.

Congruity affected accuracy as well as latency, but the effect appeared to be the same for all three tasks (i.e., the interaction between task and congruity was not significant).

Duration. In the typewritten-response condition, subjects took 9 ms longer to type incongruent words than to type congruent ones. This difference was not significant. Apparently, congruity has strong effects on the processes that lead up to a response (i.e., on initial latency) but not on the processes that actually execute the response.

Discussion

This experiment demonstrated a Stroop effect with typewritten responses, which is important because it has not been demonstrated before. We found a large Stroop effect in the initial latency of the typewritten response but virtually no Stroop effect in duration. This is important because it suggests that the locus of the Stroop effect is prior to response execution, and it supports the modularity assumption implicit in theories of the Stroop task. Apparently, one need not account for the details of response execution to explain the Stroop effect. More broadly, the dissociation between initial latency and duration is consistent with modular theories of language production (e.g., Levelt et al., 1991) and inconsistent with interactive theories (e.g., Dell, 1986; Dell & O'Seaghdha, 1991; Rumelhart & Norman, 1982).

The main purpose of the experiment was to compare the magnitude of the Stroop effect with typewritten, vocal, and arbitrary-keypress responses, to assess translation and automaticity theories of response modality effects. The data were inconsistent with both theories. The Stroop effect was larger with typewritten responses than with arbitrary keypresses, which is inconsistent with translation theories. Translation theories argue that the magnitude of the Stroop effect depends on response modality (vocal vs. manual) and predict Stroop effects of equal magnitude for typewritten and keypress responses because both involve manual responses. The Stroop effect was larger with typewritten responses than with vocal responses, which is inconsistent with automaticity theories. Automaticity theories argue that the magnitude of the Stroop effect depends on the strength of association or amount of practice with S-R mappings, and vocal responses should be practiced more and therefore more strongly associated than typewritten responses.

It may be possible to account for the magnitude of the typewritten Stroop effect in terms of the directness or consistency of mapping between visual words and typewritten words. Each letter corresponds directly and uniquely to a particular key on the keyboard. By contrast, the mapping between visual words and vocal responses is less direct and much less consistent. For example, the letter *e* occurred in each of the five color words used in Experiment 1, but it was pronounced three different ways (i.e., /ɛ/ in RED and YELLOW, /i/ in GREEN, and silent in BLUE and WHITE). Automaticity is acquired more rapidly and more strongly when mapping is more consistent (Schneider & Fisk, 1982), so typewritten responses may be more automatic than vocal responses even though they are less well practiced.

Two factors may conspire to make the typewritten Stroop effect larger than the vocal one: Visual words may activate typewritten responses sooner than they activate vocal responses, and typewritten responses to colors may take longer to generate than vocal responses. Subjects may generate typewritten responses to colors by first naming them (subvocally) and then translating the verbally coded names into typewritten responses. These two factors would create a "window of opportunity" during which the word can interfere with the response to the color, and the window may be longer with typewritten responses than with vocal responses. The longer the window, the greater the interference (Logan, 1980). The present data provide some support for this hypothesis: Initial latency tended to be longer for typewritten responses than for vocal responses, $F(1, 34) = 3.51$, $p < .10$, $MSE = 66{,}564.91$. Experiment 2 was designed, in part, to test other aspects of this hypothesis.

Two other aspects of the present results were unanticipated. First, we observed negative facilitation in the vocal Stroop task. Subjects were faster to respond to neutral stimuli than to congruent stimuli, although the difference was not significant. Negative facilitation is not unheard of in the Stroop literature (Vanayan, 1993), and facilitation is a fragile effect that is not always significant (MacLeod, 1991). Our results may have been due in part to our choice of percent symbols as neutral stimuli (see Jonides & Mack, 1984). Subjects respond more slowly to neutral stimuli that share orthographic, phonemic, and semantic properties with color words (MacLeod, 1991), and our neutral stimuli were different in all of these respects.

Second, we found no significant difference between the vocal Stroop effect and the arbitrary-keypress Stroop effect, although the difference was in the right direction. Although most studies find a larger effect with vocal responses, a null difference is not unheard of (e.g., Roe et al., 1980). Moreover, the arbitrary-keypress Stroop effect is often larger when the keys are labeled with color names (vs. colors; Sugg & McDonald, 1994) and when subjects use verbal mediation to remember the arbitrary mapping. Our subjects had only 240 trials in the arbitrary-keypress condition, which may not have been enough to allow them to abandon verbal mediation, and those 240 trials occurred in the context of the vocal response and typewritten response conditions, which may have encouraged subjects to think of responses in terms of color names. The verbal mediation required for the vocal and typewritten response conditions may have carried over to the arbitrary-keypress condition.

Experiment 2: Normal Versus Reverse Stroop Task

The second experiment compared the normal Stroop task, in which subjects reported the color of the word, with the reverse Stroop task, in which subjects reported the word itself. Half of the subjects responded vocally and half responded by typewriting. There were three purposes behind these comparisons. The first was to test the generality of the typewritten Stroop effect by replicating a standard effect in the Stroop literature. If the typewritten Stroop effect is like the vocal one, the effect in the normal Stroop task should be much larger than the effect in the reverse Stroop task.

The second purpose was to replicate and extend the dissociation between initial latency and duration observed in the previous experiments under conditions in which we expected large variation in the magnitude of the Stroop effect. If latency and duration are dissociable, then the contrast between normal and reverse Stroop tasks should affect initial latency but not duration.

The third purpose was to test the window-of-opportunity hypothesis advanced above to account for the larger Stroop effect with typewritten responses. The difference between color-naming and word-naming latencies provides an index of the length of the window of opportunity during which interference can operate. If the Stroop effect is larger because typewritten responses to color allow a longer window of opportunity, the difference between color-naming and word-naming latencies should be larger for typewritten responses than for vocal responses.

Method

Subjects. The subjects were 36 volunteers, who were recruited from the general university community and paid $5 for their services. Half of the subjects responded by typing the word or the color name and half responded by speaking it. For those who typed their responses, typing speed ranged from 30.3 to 59.1 wpm, with a mean of 46.8 wpm. Accuracy ranged from 79.5% to 98.2%, with a mean of 93.1%.

Apparatus and stimuli. Subjects who responded vocally were tested individually on the IBM AT computer used in Experiment 1. Subjects who responded by typing were tested on Gateway 2000 486 computers that controlled Gateway 2000 Crystalscan 1024 NI color monitors. There were three computers facing orthogonal walls of a larger testing room so that several subjects could be tested at once without distracting each other. The stimuli were the same as those used in Experiment 1, except that there was no neutral condition.

Procedure. The procedure was the same as in Experiment 1. Subjects who typed their responses began and ended the experiment with a test of typing speed. In the experiment proper, subjects performed two 240-trial blocks of Stroop trials. In the normal Stroop block, they reported the name of the color in which the word was written; in the reverse Stroop block, they reported the word itself.

Data analysis. The main data were the mean and standard deviation of the latencies for the initial keystroke and the onset of the vocal response, accuracy, and the mean and standard deviation of the durations of the typewritten responses. The first three variables were analyzed in 2 (response: typewritten vs. vocal) × 2 (task: normal vs. reverse Stroop) × 2 (congruity: congruent, incongruent) ANOVAs. Means and standard deviations of the durations of the typing responses were analyzed in 2 (task) × 2 (congruity) ANOVAs.

Results

The means and standard deviations of the latency to initiate typewritten and vocal responses, the accuracy of responses, and the means and standard deviations of the durations of typewritten responses in the normal and reverse Stroop conditions are presented in Table 3 as a function of congruity. The results of ANOVAs performed on these data are presented in Table 4.

Initial latency. For both typewritten and vocal responses, initial latencies were much longer in the normal Stroop condition (M = 783 ms) than in the reverse Stroop condition (M = 572 ms), replicating standard effects in the Stroop literature. For both response types, the congruity effect was much larger in the normal Stroop condition (M = 184 ms) than in the reverse Stroop condition (M = 9 ms), also replicating standard effects in the Stroop literature. The normal Stroop effect was much larger with typewritten responses (244 ms) than with vocal responses (125 ms), replicating the findings of Experiment 1. The reverse Stroop effect was negligible with both response types (7 ms for typewritten responses, 10 ms for vocal responses). The difference between normal and reverse Stroop conditions was larger with typewritten responses (M = 224 ms) than with vocal responses (M = 198 ms), but the interaction between response and task was not significant (see Table 4).

Accuracy. The percentage of correct responses was higher for vocal responses (M = 98.8%) than for typewritten responses (94%), replicating the findings of Experiment 1. We calculated the accuracy of individual keystrokes by taking the 4.6th root of the probability of typing the whole word correctly and found a value of 98.7%, which is quite

Table 3
Experiment 2: Means and Standard Deviations of Time to Initiate Response and of Response Duration, and Response Accuracy for Normal and Reverse Stroop Conditions

Measure	Stroop condition			
	Normal: Report color		Reverse: Report word	
	Congruent	Incongruent	Congruent	Incongruent
Typewritten responses				
Time to initiate response				
M latency	759	1,014	654	661
SD	168	310	116	119
Time between first & last keystrokes				
M duration	554	559	551	551
SD	204	202	198	204
Response accuracy (%)	96	92	95	93
Vocal responses				
Time to initiate response				
M latency	623	747	481	492
SD	155	174	112	136
Response accuracy (%)	100	96	100	99

Table 4
Analysis of Variance for Conditions in Experiment 2

Measure	Response (R)	Task (T)	R × T	Congruity (C)	R × C	T × C	R × T × C
Latency							
F	41.66**	395.21**	1.38	196.28**	17.78**	149.33**	18.08**
df	1, 34	1, 34	1, 34	1, 34	1, 34	1, 34	1, 34
MSE	29,145.94	4,041.09	4,041.09	1,707.45	1,707.45	1,860.45	1,860.45
SD of latency							
F	0.10	24.04**	1.27	5.33*	0.10	0.31	0.80
df	1, 34	1, 34	1, 34	1, 34	1, 34	1, 34	1, 34
MSE	9,799.12	4,191.35	4,191.35	2,358.91	2,358.91	2,972.28	2,972.28
Accuracy							
F	58.70**	4.62*	5.14*	46.80**	0.08	21.77**	0.40
df	1, 34	1, 34	1, 34	1, 34	1, 34	1, 34	1, 34
MSE	12.81	41.17	41.17	4.04	4.04	3.93	3.93
Duration (typewritten responses)							
F		0.18		1.88		1.67	
df		1, 17		1, 17		1, 17	
MSE		3,438.77		75.50		54.54	
SD of duration							
F		0.08		0.23		1.99	
df		1, 17		1, 17		1, 17	
MSE		1,066.83		321.46		146.81	

$^*p < .05$. $^{**}p < .01$.

close to the value observed with vocal responses. Accuracy was about the same in normal ($M = 96\%$) and reverse Stroop tasks ($M = 96.8\%$). There was a small congruity effect that was larger in the normal Stroop condition (4%) than in the reverse Stroop condition (1.5%), corroborating the RT results.

Duration. The time between the first and the last keystrokes was about the same in the two typewritten response conditions, averaging 557 ms in the normal Stroop condition and 551 ms in the reverse Stroop condition. The congruity effects were negligible: 5 ms in the normal Stroop condition and 0 ms in the reverse Stroop condition. Again, these results suggest that the factors that produce the Stroop effect operate before response execution.

Discussion

The experiment replicated the standard contrast between normal and reverse Stroop tasks with typewritten and vocal responses, showing that the typewritten Stroop effect resembles the vocal Stroop effect in another important respect. The experiment also dissociated response execution from response selection and prior processes with typewritten responses by showing that the contrast between normal and reverse Stroop tasks was apparent in initial latency only, not in duration. The normal Stroop effect was larger with typewritten responses than with vocal responses, replicating Experiment 1. The window-of-opportunity account of this difference in magnitude did not fare very well. The difference between response latencies to words and colors was larger for typewritten responses than for vocal responses, as the window-of-opportunity hypothesis predicts, but the difference was not large and it was not significant. It remains possible, however, that the larger Stroop effect with typewritten responses results from greater automaticity of typewritten responses. We return to this point in the General Discussion section.

Experiment 3: Whole Word Versus First Letter

Experiments 1 and 2 showed that congruity affected the initial latency of the typewritten response but not its duration. Experiment 3 was designed to replicate this effect and provide converging evidence. If the typewritten Stroop effect is due entirely to the processes that lead up to typing the first letter and not to the processes that type subsequent letters, then the Stroop effect should be just as large if subjects simply type the first letter of the color name as if they type the whole word. If letters beyond the first play a role in the typewritten Stroop effect, then the effect should be smaller when subjects type only the first letter. Experiment 3 was designed to test these hypotheses. Subjects participated in two conditions, typing the first letter of the name of the color in one, and typing the entire color name in the other. The whole-word condition provided a replication of the typewritten response conditions of Experiments 1 and 2, allowing us to confirm that the Stroop effect occurs in the initial latency rather than in the duration of the typewritten response.

Method

Subjects. The subjects were 18 volunteers recruited from the general university community. They were paid $5 for participating. Speed on the typing test ranged from 27.9 to 69.6 wpm, with a mean of 49.5 wpm. Accuracy on the typing test ranged from 84.6% to 99.1%, with a mean of 93.1%.

Apparatus and stimuli. Stimulus presentation and response registration were controlled by the Gateway computers used in Experiment 2. In all other respects, the apparatus and stimuli were the same as in Experiments 1 and 2.

Procedure. The procedure was the same as in Experiments 1 and 2, except that the first-letter condition replaced the arbitrary-keypress condition, and the vocal response condition was dropped. Thus, there were two 240-trial conditions: the whole word condition, in which subjects responded by typing all of the letters in the word that named the color of the stimulus, and the first-letter condition, in which subjects typed only the first letter of the word that named the color of the stimulus.

Subjects received a test of typing speed before and after the experimental trials, as in Experiments 1 and 2. The order of the two experimental conditions—first letter versus whole word—was counterbalanced, with 9 subjects receiving each order. In all other respects, the procedure was the same as in Experiments 1 and 2.

Data analysis. The main data were the mean and standard deviation of the latencies for the initial keystroke, accuracy, and the mean and standard deviation of the durations of the typewritten responses in the whole-word condition. The first three variables were analyzed in 2 (task: whole word vs. first letter) × 3 (congruity: congruent, incongruent, neutral) ANOVAs; the last two were analyzed in one-way ANOVAs with congruity (congruent, incongruent, neutral) as the factor.

Results

The means and standard deviations of the latency to initiate the response and the accuracy of responses for the whole-word and first-letter conditions are presented in Table 5 as a function of congruity. Table 5 also contains the means and standard deviations of the durations of the typewritten responses from the whole-word condition as a function of congruity. The results of ANOVAs performed on these data are presented in Table 6.

Initial latency. The latency of the first keystroke was slightly longer in the whole-word condition (M = 879 ms) than in the first-letter condition (M = 856 ms), but the difference was not significant. A Stroop effect was observed in both conditions. The difference between incongruent and congruent RTs was 250 ms in the whole-word condition and 263 ms in the first-letter condition. We calculated facilitation and interference scores, subtracting congruent RTs from neutral RTs and subtracting neutral RTs from incongruent RTs. There was 41 ms of facilitation in the whole-word condition and 46 ms in the first-letter condition. There was 209 ms of interference in the whole-word condition and 217 ms in the first-letter condition. All of these effects were significantly greater than zero (Fisher's LSD = 22 ms, $p < .05$), but there were no significant differences between the effects in the whole-word condition and the effects in the first-letter condition.

Accuracy. The percentage of correct responses was lower in the whole-word condition (M = 92.7) than in the first-letter condition (M = 96.4). Again, part of this difference may be due to the difference in the number of keystrokes that had to be executed correctly for the response to be scored "correct": 4.6 in the whole-word condition versus 1 in the first-letter condition. Assuming that the keystrokes in the whole-word condition were independent of each other, then the accuracy per keystroke was 98.4%, which compares more favorably with the value for the first-letter condition.

There was a congruity effect in the accuracy data that mirrored the effect in the latency data. Subjects were less accurate in the incongruent condition than in the congruent or neutral conditions. The congruity effect on accuracy was the same for the whole-word and first-letter conditions; the interaction between task and congruity was not significant.

Duration. In the whole-word condition, subjects took 10 ms longer to type color names that were incongruent with the word than color names that were congruent with the word. This difference was not significant, suggesting once again that the Stroop effect occurred primarily before the response execution stage.

Table 5
Experiment 3: Means and Standard Deviations of Time to Initiate Response and Response Accuracy for First-Letter and Whole-Word Conditions and Response Duration in Whole-Word Condition

Measure	Congruent	Incongruent	Neutral
Type first letter			
Time to initiate response			
M latency	753	1,016	799
SD	219	305	202
Response accuracy (%)	99	93	97
Type whole word			
M latency	782	1,032	823
SD	224	240	169
Response accuracy (%)	95	89	94
Time between first & last keystrokes			
M duration	538	548	540
SD	204	202	196

Table 6
Analysis of Variance for Conditions in Experiment 3

Measure	Task (T)	Congruity (C)	T × C
Latency			
F	0.71	246.88**	0.40
df	1, 17	2, 34	2, 34
MSE	20,399.30	2,755.73	1,016.82
SD of latency			
F	2.14	8.78**	2.31
df	1, 17	2, 34	2, 34
MSE	12,033.22	7,837.14	4,748.71
Accuracy			
F	7.19*	25.29**	0.29
df	1, 17	2, 34	4, 68
MSE	43.58	13.38	8.14
Duration (typewritten responses)			
F		1.93	
df		2, 34	
MSE		258.05	
SD of duration			
F		0.47	
df		2, 34	
MSE		621.92	

*$p < .05$. **$p < .01$.

Discussion

The most important result in this experiment was that the Stroop effect was just as large in the first-letter condition as it was in the whole-word condition. Apparently, letters beyond the first one have little impact on the magnitude of the Stroop effect. The contrast between initial latency and duration in the whole-word condition provides converging evidence. The Stroop effect was apparent only in initial latency. There was no effect in the duration of the typewritten response. These results are consistent with the hypothesis that the Stroop effect is located in processes prior to response execution and consistent with modularity assumptions in theories of the Stroop effect and theories of language production.

Experiment 4: Strategic Modulation of Color–Word Impact

Experiment 4 examined the effects of varying the relative frequency of congruent versus incongruent trials on the magnitude of the typewritten Stroop effect. In one condition (80–20), 80% of the trials involved congruent stimuli and 20% involved incongruent stimuli. In the other condition (20–80), only 20% of the trials involved congruent stimuli, whereas 80% involved incongruent stimuli. Past research suggested that this manipulation would affect the strategies with which subjects approached the Stroop task, resulting in an enhanced Stroop effect when congruent trials predominated and a reduced or reversed Stroop effect when incongruent trials predominated (Logan, 1980; Logan & Zbrodoff, 1979; Logan et al., 1984; Zbrodoff & Logan, 1986). We wanted to see whether similar strategies would operate in the typewritten Stroop task. Another purpose of the manipulation was to dissociate initial latency and duration once again. The previous experiments, and Experiment 3 in particular, suggest that the relative frequency of congruent trials should affect the magnitude of the Stroop effect in initial latency but not in duration.

Method

Subjects. The subjects were 18 volunteers from the general university community who were paid $5 for participating. Typing speed ranged from 31.6 to 59.4 wpm, with a mean of 42.1 wpm. Accuracy ranged from 82.0% to 97.3%, with a mean of 90.4%.

Apparatus and stimuli. The apparatus and stimuli were the same as those used in Experiment 3.

Procedure. The procedure was largely the same as in previous experiments. Subjects began with a test of typing speed, performed in two experimental conditions, and ended with a second test of typing speed. The two experimental conditions each involved 200 trials but differed in the composition of those trials. In the 80–20 condition, 80% of the trials involved congruent stimuli and 20% involved incongruent stimuli (i.e., 160 congruent trials randomly mixed with 40 incongruent trials), whereas in the 20–80 condition, 20% of the trials (40 trials) involved congruent stimuli and 80% (160 trials) involved incongruent stimuli. Half of the subjects performed the 20–80 condition before the 80–20 condition, and half did the opposite. None of the subjects was told about the relative frequency of congruent and incongruent stimuli, and the change in relative frequency halfway through the experiment was not announced to them. It did fall after a break, however (subjects were allowed brief rests every 100 trials). In all other respects, the procedure was the same as the one used in the previous experiments.

Data analysis. The main data were the mean and standard deviation of the latencies for the initial keystroke, accuracy, and the mean and standard deviation of the durations of the typewritten responses. These variables were analyzed in 2 (frequency: 80–20 vs. 20–80) × 2 (congruity: congruent, incongruent) ANOVAs.

Results

The means and standard deviations of the latency to initiate the response, the accuracy of responses, and the means and standard deviations of the durations of the responses in each relative frequency condition are presented in Table 7 as a function of congruity. The results of ANOVAs performed on these data are presented in Table 8.

Initial latency. The latency to type the first keystroke was about the same in the two relative frequency conditions (Ms = 915 and 909 ms for the 80–20 and the 20–80 conditions, respectively). Once again, there was a strong congruity effect, averaging 214 ms across the two relative frequency conditions. The congruity effect was strongly affected by the relative frequency of congruent versus incongruent trials, however, averaging 305 ms with 80% congruent and 20% incongruent versus 123 ms with 20% congruent and 80% incongruent. For the first time with typewritten responses, these results replicate standard effects in the Stroop literature.

Accuracy. The percentage of correct responses was about the same in the two relative frequency conditions, averaging 91% in the 80–20 condition and 92% in the 20–80 condition. Congruity affected accuracy, averaging 93% for congruent trials and 90% for incongruent trials. This congruity effect was modulated by the relative frequency of

Table 7
Experiment 4: Means and Standard Deviations of Time to Initiate Response and of Response Duration, and Response Accuracy in Experimental Conditions

Measure	Congruent	Incongruent
80% congruent, 20% incongruent		
Time to initiate response		
M latency	762	1067
SD	160	196
Time between first & last keystrokes		
M duration	600	592
SD	206	199
Response accuracy (%)	93	88
20% congruent, 80% incongruent		
Time to initiate response		
M latency	847	970
SD	202	224
Time between first & last keystrokes		
M duration	592	592
SD	213	201
Response accuracy (%)	93	91

Table 8
Analysis of Variance for Conditions in Experiment 4

Measure	Task (T)	Congruity (C)	T × C
Latency			
F	0.05	136.32**	147.43**
df	1, 17	1, 17	1, 17
MSE	11,787.48	6,042.48	1,017.86
SD of latency			
F	4.53*	3.93	0.41
df	1, 17	1, 17	1, 17
MSE	4,895.72	3,958.87	1,985.36
Accuracy			
F	2.74	11.26**	9.91**
df	1, 17	1, 17	1, 17
MSE	14.76	19.59	8.53
Duration (typewritten responses)			
F	0.16	0.67	1.79
df	1, 17	1, 17	1, 17
MSE	2,012.51	323.69	141.21
SD of duration			
F	0.48	3.28	0.24
df	1, 17	1, 17	1, 17
MSE	653.39	469.65	345.35

*$p < .05$. **$p < .01$.

congruent versus incongruent trials. There was a 5% congruity effect when congruent trials predominated (in the 80–20 condition) and a 2% effect when incongruent trials predominated (in the 20–80 condition).

Duration. The time between the first and last letter did not vary much in this experiment. It was about the same when congruent trials predominated ($M = 596$ ms in the 80–20 condition) as when incongruent trials predominated ($M = 592$ ms in the 20–80 condition). It was not affected much by congruity (*M*s = 596 ms for congruent and 592 ms for incongruent), and the congruity effect was not modulated by the relative frequency of congruent versus incongruent trials.

Discussion

The magnitude of the Stroop effect was affected strongly by the relative frequency of congruent versus incongruent stimuli, replicating previous results in the literature (e.g., Logan, 1980; Logan & Zbrodoff, 1979; Logan et al., 1984; Zbrodoff & Logan, 1986) and suggesting that the typewritten Stroop task is susceptible to the same sort of strategic modulation as vocal and arbitrary-keypress Stroop tasks. The strategic modulation and the Stroop effect itself was restricted to initial latency. There was no effect of congruity or relative frequency of congruent versus incongruent trials on the duration of the typing response. This result provides further support for the modularity of response execution and response selection (and prior) processes.

General Discussion

The experiments demonstrated a Stroop effect with typewritten responses that was much stronger than the Stroop effect with vocal responses and the Stroop effect with arbitrary keypresses. Moreover, the typewritten Stroop effect occurred entirely in the latency to type the first keystroke; the duration of the typewritten response was not affected at all by the congruity of the Stroop stimuli. These results have implications for theories of the Stroop effect and for issues of modularity in the literature on the Stroop effect and the literature on language production.

Automaticity Versus Translation

Typically, the Stroop effect is stronger with vocal responses than with arbitrary keypresses (Logan et al., 1984; Majeres, 1974; Melara & Mounts, 1993; Neill, 1977; Redding & Gerjets, 1977; Simon & Sudalaimuthu, 1979; White, 1969). Translation theories explain this difference by arguing that words are directly connected to vocal responses but require some translation to be connected to keypresses. Thus, words activate vocal responses more readily than arbitrary keypresses and therefore produce more interference with color naming. Automaticity theories explain the larger Stroop effect with vocal responses by arguing that words are more strongly associated with vocal responses than with arbitrary keypresses and therefore are more likely to activate vocal responses and interfere with color naming.

The typewritten Stroop task provides an interesting way to contrast these positions. From the perspective of translation theories, it should produce the same sort of Stroop effect as arbitrary keypresses, because typewriting and arbitrary keypresses both involve manual responses and so both should require translation. From the perspective of automaticity theories, the typewritten Stroop task should produce the same sort of Stroop effect as vocal responses, because both typewritten and vocal responses are strongly associated with printed words. The data were inconsistent with both of these hypotheses. The Stroop effect was larger with typewritten responses than with arbitrary keypresses, contrary to translation theories. It was larger with typewritten responses than with vocal responses, which is contrary to automaticity theories.

It may be possible to salvage automaticity theories by arguing that typewritten responses may be more automatic than vocal responses. Vocal responses may be more practiced than typewritten responses, but the mapping between visually presented words and vocal responses is less direct and less consistent than the mapping between words and typewritten responses (see, e.g., Seidenberg & McClelland, 1989). Consistency of practice is an important determinant of automaticity (Shiffrin & Schneider, 1977), and even minor degrees of inconsistency seem to impair automatization (Schneider & Fisk, 1982).

It may be possible to salvage translation hypotheses as well by defining modalities in terms of central representations and processes rather than input and output modalities. From this perspective, typewriting and speaking are both linguistic activities, whereas making arbitrary keypresses is not. Printed words, also linguistic, may activate typewritten and spoken response directly without translation between modalities, whereas some kind of mediation may be required for printed words to activate arbitrary keypresses.

Nevertheless, it may be hard to dismiss automaticity entirely. Speaking may be universal in humans, but typewriting is certainly not. Speaking is acquired in the second year of life by virtually all humans, whereas typewriting is acquired typically in the second or third decade of life by people who have access to the technology and are willing to invest hundreds of hours of practice in acquiring the skill. Moreover, people speak to others every day in an on-line, real-time fashion, whereas typewritten communication is less common even in those who type, and it typically occurs off-line, outside of real time. Thus, typing and speaking do not appear to be alternative expressions of a common linguistic ability. Typewritten responses may show a Stroop effect because they are well practiced, not because they are inherently linguistic.[3]

Automaticity may be important but it may not be the whole story. The magnitude of the typewritten Stroop effect is not correlated with typing skill. Over the 72 subjects in all four experiments, the correlation between speed on the typing test and the magnitude of the Stroop effect (difference in initial latency to incongruent and congruent stimuli) was −.072, and the correlation between typing-test accuracy and the magnitude of the Stroop effect was −.039. It may be that some minimal degree of automaticity is necessary to produce the basic Stroop effect, and after that some other factor, such as compatibility (e.g., Kornblum et al., 1990), may interact with automaticity to determine the strength of the effect.

The best solution may be to combine translation and automaticity theories. Researchers have pitted the theories against each other, but they have much in common. Automaticity may determine whether translation is necessary; translation may be the process that allows S-R mapping when automaticity is insufficient. On the one hand, theories of automaticity contrast direct S-R association, which is sufficient for well-practiced activities, with controlled, strategic, or algorithmic processing, which is required for novel activities (e.g., Logan, 1988; Posner & Snyder, 1975; Shiffrin & Schneider, 1977). The translation processes in current translation theories may well be characterized as controlled, strategic, or algorithmic. On the other hand, translation theories assume, implicitly at least, that S-R pathways are learned. This assumption is more apparent if modalities are defined in terms of central representations and processes. Reading, writing, and speaking are learned activities that interface the linguistic modality to the outside world. Translation appears to be necessary when input and output cannot be mediated by prior learning.

The connection between translation and automaticity theories may be made already in compatibility theories (Kornblum et al., 1990; Phaf et al., 1990; Treisman & Fearnley, 1969) that argue that vocal and typewritten responses are more compatible with printed words than with colors. They explain the Stroop effect in terms of activation of conflicting responses by compatible stimuli; they explain the reverse Stroop effect by arguing that colors are less compatible with vocal responses than words, so they are less likely to activate conflicting responses; and they explain the modality effect by arguing that vocal responses are more compatible than arbitrary keypresses with printed words. Compatibility theories blur the distinction between translation and automaticity theories: Kornblum et al. (1990) defined compatibility in terms of *dimensional overlap*, arguing that stimuli and responses that have more features in common are more compatible (e.g., responding to a spatially distinct light by touching it is more compatible than responding with a spoken word because the touch response shares spatial location with the light, whereas the spoken response does not). This definition suggests a similarity between compatibility theories and translation theories, because both appear to define compatibility in terms of the features of the motor response. However, Kornblum et al. and others have argued that practice can increase the degree of compatibility, which suggests a similarity between compatibility theories and automaticity theories.

Locus of the Stroop Effect

The locus at which the Stroop effect occurs has been an important source of controversy in the Stroop literature (Cohen et al., 1990; Duncan-Johnson & Kopell, 1980, 1981; Glaser & Glaser, 1989; Hock & Egeth, 1970; Logan, 1980; Morton, 1969; Morton & Chambers, 1973; Posner & Snyder, 1975; Sugg & McDonald, 1994; Treisman & Fearnley, 1969; Virzi & Egeth, 1985; Warren & Marsh, 1979). Researchers have suggested perceptual, semantic, or response selection stages, either exclusively or in combination, as the locus of the effect. The present experiments add to this literature by distinguishing response execution stages from response selection and prior stages. If the Stroop effect occurs in response execution, then its magnitude should vary with the nature of the response (e.g., vocal, keypress, or typewritten), and it should affect the duration of the response. If the Stroop effect occurs prior to response execution, then its magnitude should not vary with response type, and it should not appear in measures of response duration. Predictions about response modality are not unique, because factors such as automaticity, compatibility, or the requirement for translation that are correlated with response modality can also affect the magnitude of the effect. However, predictions about response duration are clear and uncompromised by other factors.

[3] The typing literature suggests that typewriting may be less linguistic than speaking (for reviews, see Cooper, 1983; Rumelhart & Norman, 1982; Salthouse, 1986). Linguistic structures larger than the word seem to have little impact on typing. Skilled typists type scrambled words more slowly than intact words, but they type scrambled text as quickly as coherent text (Fendrick, 1937; Hershman & Hillix, 1965; Larochelle, 1983, 1984; Salthouse, 1984, 1985; Shaffer, 1973; Shaffer & Hardwick, 1968; Thomas & Jones, 1970; West & Sabban, 1982). Skilled typists type more slowly when preview is restricted, but previews of five to seven characters allow them to type at normal rates (Coover, 1923; Hershman & Hillix, 1965; Salthouse, 1984, 1985; Shaffer, 1973; Shaffer & French, 1971; Shaffer & Hardwick, 1970).

The data from all four experiments consistently found null effects of congruity on duration measures. The mean difference in the duration of incongruent and congruent responses was 9, 5, 10, and −4 ms in Experiments 1–4, respectively. By contrast, the mean difference in the initial latency of incongruent and congruent responses was 214, 244, 250, and 214 ms in Experiments 1–4, respectively. These results suggest that the locus of the Stroop effect is prior to response execution, in response selection, or in some process that precedes it.

Modularity

Theories of the Stroop effect. Theories of the Stroop effect appear to assume a kind of modularity between response selection and prior processes on the one hand and response execution on the other. This appearance stems from the fact that none of the theories says anything about response execution in their accounts of the Stroop effect, basing their arguments instead on response selection and prior processes (e.g., Cohen et al., 1990; Logan, 1980; Phaf et al., 1990). The theories appear to predict that the Stroop effect should occur prior to response execution. The present experiments allowed us to test this modularity assumption in two ways. First, modularity predicts no effect of response modality. This is a weak prediction because response modality may be confounded with automaticity, compatibility, and translation processes. Second, modularity predicts no Stroop effect on response duration. This is a stronger prediction because no other factors compromise it. It was confirmed readily in each experiment. Moreover, D. Spieler (personal communication, May 16, 1996) found no Stroop effect in the duration of vocal responses. Thus, the modularity assumption, implicit or not, is supported by the data.

It would be tempting to interpret the apparent modularity of response selection and execution as evidence that execution is ballistic. If execution was ballistic, the typing response, once begun, would run on to completion with no further intervention. Other data suggest that it would be better not to yield to this temptation. Logan (1982) and Salthouse and Saults (1987) found that skilled typists could inhibit typing responses in midword, one or two keystrokes after a signal to stop typing was presented. Long (1976) and Rabbitt (1978) found that skilled typists often stopped typing in midword when they made an error. The median number of keystrokes typed after the error was zero. Those data suggest that typists have close control over execution, so execution is not ballistic even if it is automatic and relatively independent of earlier processes. Ballisticity is only part of automaticity (Zbrodoff & Logan, 1986).

Theories of language production. Theories of language production are divided on the issue of modularity. Some argue that selection and execution processes are modular (i.e., separate and additive; Levelt et al., 1991), whereas others argue that they interact with each other (Dell, 1986; Dell & O'Seaghdha, 1991). Rumelhart and Norman's (1982) theory of typewriting sides with the nonmodular theories, assuming strong interactions between selection and execution. The present data, showing the Stroop effect entirely in initial latency and not at all in duration, are more consistent with theories that argue for modularity.

The story is not that simple, however. Other factors, such as frequency, affect both latency and duration in spoken (Balota, Boland, & Shields, 1989; Balota & Chumbley, 1985; see also Balota & Abrams, 1995) and typewritten responses (e.g., Gentner, Larochelle, & Grudin, 1988; Inhoff, 1991). These factors may have separate, independent effects on both measures, as modular theories would predict, or they may have interactive effects, as nonmodular theories would predict. Further research and further theoretical development will be necessary to settle the modularity issue.

Conclusions

The present experiments showed a strong Stroop effect with typewritten responses, suggesting that automaticity is an important factor in producing the Stroop effect. The experiments also showed that the typewritten Stroop effect occurred exclusively in the latency to type the first keystroke and not in the duration of the typewritten response, suggesting that the locus of the Stroop effect is prior to response execution and supporting the modularity assumption implicit in theories of the Stroop effect and explicit in certain theories of language production. The experiments encourage further exploration of the typewritten Stroop effect, showing that it provides new insights into important issues in the Stroop literature.

References

Balota, D. A., & Abrams, R. A. (1995). Mental chronometry: Beyond onset latencies in the lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1289–1302.

Balota, D. A., Boland, J. E., & Shields, L. W. (1989). Priming in pronunciation: Beyond pattern recognition and onset latency. *Journal of Memory and Language, 28,* 14–36.

Balota, D. A., & Chumbley, J. I. (1985). The locus of word-frequency effects in the pronunciation task: Lexical access and/or production? *Journal of Memory and Language, 24,* 89–106.

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.

Collier, M. (1995). *Border collies.* Neptune City, NJ: T.F.H. Publications.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual route and parallel-distributed-processing approaches. *Psychological Review, 100,* 589–608.

Cooper, W. E. (Ed.). (1983). *Cognitive aspects of skilled typewriting.* New York: Springer-Verlag.

Coover, J. E. (1923). A method of teaching typewriting based upon a psychological analysis of expert typewriting. *National Education Association: Addresses and Proceedings, 61,* 561–567.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93,* 283–321.

Dell, G. S., & O'Seaghdha, P. G. (1991). Mediated and convergent lexical priming in language production: A comment on Levelt et al. (1991). *Psychological Review, 98,* 604–613.

Duncan-Johnson, C. C., & Kopell, B. S. (1980). The locus of interference in a Stroop task: When you read "blue," do you see "red"? *Psychophysiology, 17,* 308–309.

Duncan-Johnson, C. C., & Kopell, B. S. (1981). The Stroop effect: Brain potentials localize the source of interference. *Science, 214,* 938–940.

Fendrick, P. (1937). Hierarchical skills in typewriting. *Journal of Educational Psychology, 28,* 609–620.

Fodor, J. A. (1983). *The modularity of mind.* Cambridge, MA: MIT Press.

Gentner, D. R., Larochelle, S., & Grudin, J. (1988). Lexical, sublexical, and peripheral effects in skilled typewriting. *Cognitive Psychology, 20,* 524–548.

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

Hershman, R. L., & Hillix, W. A. (1965). Data processing in typing: Typing rate as a function of kind of material and amount exposed. *Human Factors, 7,* 483–492.

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

Inhoff, A. W. (1991). Word frequency during copytyping. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 478–487.

Jonides, J., & Mack, R. (1984). The cost and benefit of cost and benefit. *Psychological Bulletin, 96,* 29–44.

Keele, S. W. (1972). Attention demands of memory retrieval. *Journal of Experimental Psychology, 93,* 245–248.

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

Larochelle, S. (1983). A comparison of skilled and novice performance in discontinuous typing. In W. E. Cooper (Ed.), *Cognitive aspects of skilled typewriting* (pp. 67–94). New York: Springer-Verlag.

Larochelle, S. (1984). Some aspects of movements in skilled typewriting. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X* (pp. 43–54). Hillsdale, NJ: Erlbaum.

Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review, 98,* 122–142.

Logan, G. D. (1980). Attention and automaticity in Stroop and priming tasks: Theory and data. *Cognitive Psychology, 12,* 523–553.

Logan, G. D. (1982). On the ability to inhibit complex movements: A stop-signal study of typewriting. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 778–792.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95,* 492–527.

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.

Logan, G. D., Zbrodoff, N. J., & Williamson, J. (1984). Strategies in the color–word Stroop task. *Bulletin of the Psychonomic Society, 22,* 135–138.

Long, J. (1976). Visual feedback and skilled keying: Differential effects of masking the printed copy and the keyboard. *Ergonomics, 19,* 93–110.

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 and Cognition, 14,* 126–135.

Majeres, E. (1974). The combined effects of stimulus and response conditions on the delay in identifying the print color of words. *Journal of Experimental Psychology, 102,* 868–874.

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

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 88,* 375–407.

Melara, R. D., & Mounts, J. R. W. (1993). Selective attention to Stroop dimensions: Effects of baseline discriminability, response mode, and practice. *Memory & Cognition, 21,* 627–645.

Morton, J. (1969). Categories of interference: Verbal mediation and conflict in card sorting. *British Journal of Psychology, 60,* 329–346.

Morton, J., & Chambers, S. M. (1973). Selective attention to words and colours. *Quarterly Journal of Experimental 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.

Phaf, R. H., van der Heijden, A. H. C., & Hudson, P. T. W. (1990). SLAM: A connectionist model for attention in visual selection tasks. *Cognitive Psychology, 22,* 273–341.

Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 55–85). Hillsdale, NJ: Erlbaum.

Pritchatt, D. (1968). An investigation into some of the verbal processes that underlie associative verbal processes of the Stroop color effect. *Quarterly Journal of Experimental Psychology, 20,* 351–359.

Rabbitt, P. M. A. (1978). Detection of errors by skilled typists. *Ergonomics, 21,* 945–958.

Redding, G. M., & Gerjets, D. A. (1977). Stroop effect: Interference and facilitation with verbal and manual responses. *Perceptual and Motor Skills, 45,* 11–17.

Roe, W. T., Wilsoncroft, W. E., & Griffiths, R. S. (1980). Effects of motor and verbal practice on the Stroop task. *Perceptual and Motor Skills, 50,* 647–650.

Rumelhart, D. E., & Norman, D. A. (1982). Simulating a skilled typist: A study of skilled cognitive–motor performance. *Cognitive Science, 6,* 1–36.

Salthouse, T. A. (1984). Effects of age and skill in typing. *Journal of Experimental Psychology: General, 113,* 345–371.

Salthouse, T. A. (1985). Anticipatory processing in transcription typing. *Journal of Applied Psychology, 70,* 264–271.

Salthouse, T. A. (1986). Perceptual, cognitive, and motoric aspects of transcription typing. *Psychological Bulletin, 99,* 303–319.

Salthouse, T. A., & Saults, J. S. (1987). Multiple spans in transcription typing. *Journal of Applied Psychology, 72,* 187–196.

Schneider, W., & Fisk, A. D. (1982). Degree of consistent training: Improvements in search performance and automatic process development. *Perception & Psychophysics, 31,* 160–168.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96,* 523–568.

Shaffer, L. H. (1973). Latency mechanisms in transcription. In S. Kornblum (Ed.), *Attention and performance IV* (pp. 435–446). New York: Academic Press.

Shaffer, L. H., & French, A. (1971). Coding factors in transcription. *Quarterly Journal of Experimental Psychology, 23,* 268–274.

Shaffer, L. H., & Hardwick, J. (1968). Typing performance as a function of text. *Quarterly Journal of Experimental Psychology, 20,* 360–369.

Shaffer, L. H., & Hardwick, J. (1970). The basis of transcription skill. *Journal of Experimental Psychology, 84,* 424–440.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84,* 127–190.

Simon, J. R., & Sudalaimuthu, P. (1979). Effects of S–R mapping and response modality on performance in a Stroop task. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 176–187.

Sternberg, S., Monsell, S., Knoll, R. L., & Wright, C. E. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In G. E. Stelmach (Ed.), *Information processing in motor control and learning* (pp. 117–152). New York: Academic Press.

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

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 and Performance, 20,* 647–675.

Thomas, E. A. C., & Jones, R. G. (1970). A model for subjective grouping in typewriting. *Quarterly Journal of Experimental Psychology, 22,* 353–367.

Treisman, A. M., & Fearnley, S. (1969). The Stroop test: Selective attention to colours and words. *Nature, 222,* 437–439.

Vanayan, M. (1993). *Relating interference and facilitation in the Stroop task: An individual-differences comparison.* Unpublished doctoral dissertation, University of Toronto, Toronto, Ontario, Canada.

Virzi, R. A., & Egeth, H. E. (1985). Toward a translational model of Stroop interference. *Memory & Cognition, 13,* 304–319.

Warren, L. R., & Marsh, G. R. (1979). Changes in event related potentials during processing of Stroop stimuli. *International Journal of Neuroscience, 9,* 217–223.

West, L. J., & Sabban, Y. (1982). Hierarchy of stroking habits at the typewriter. *Journal of Applied Psychology, 67,* 370–376.

White, B. W. (1969). Interference in identifying attributes and attribute names. *Perception & Psychophysics, 6,* 166–168.

Zbrodoff, N. J., & Logan, G. D. (1986). On the autonomy of mental processes: A case study of arithmetic. *Journal of Experimental Psychology: General, 115,* 118–130.

Appendix

Texts Used for the Test of Typing Speed

These texts are adapted from *Border Collies* (pp. 6, 13–14, 20–21) by M. Collier, 1995, Neptune City, NJ: T.F.H. Publications.

Text 1

Dogs have been part of human life since before history was recorded. People started to make use of dogs very early in their relationship. First, they were used as guards for the home, and later as guards for other properties, such as livestock. In early times, they were guards only, and there is no written record of dogs working with livestock in any way until the Book of Job in the Old Testament. Even here they are referred to only briefly, and it would seem that they were guard dogs only, even then. The first modern reference to a dog working sheep in a similar manner to the modern Border Collie was made in the seventeenth century.

Text 2

A Border Collie is never happier than when doing something. Such is their desire to please that they are willing to be trained to do most things, and they usually excel in the field chosen for them by their owner. Although primarily a sheep dog, the Border Collie is not averse to turning his paw to anything that involves work or pleasure in any sphere. Border Collies are the dog to beat at obedience trials. They relish working hard with their handlers, and seem to find little difficulty in adapting their sheep-working habits to obedience, which, on the face of it, appears to be a totally different set of skills.

Text 3

Border Collies have become the chief participant in the sport of agility. Their natural athleticism and keenness to please have made them a very suitable subject for this sport, where the main requirements are speed and the ability to clear a variety of jumps, weave through poles, and go through tunnels. If the Border Collie has a fault in this sport, it is that it can be faster than the handler, and this can lead to errors. It is no coincidence, though, that the top agility dog has been a Border Collie since they started recording winners. Working trials are yet another canine sphere where Border Collies have become the dog to beat.

Text 4

The natural working ability in Border Collies makes them most easy to train, yet there are some dogs that have no working instinct whatever, and will never become effective sheep workers however long they are trained. It is not unusual for farmers to buy a fully or partly trained dog, as some have neither the time nor the ability to train the dogs themselves. It also means that they do not waste valuable time on a dog that may not turn out to be a successful worker. Many farmers find that the cost of such a dog is well worth it. Even good trial dogs sometimes change ownership, but the prices are very high.

Received June 11, 1996
Revision received March 10, 1997
Accepted April 25, 1997 ■