Research Article

Cognitive Control in Children
Stroop Interference and Suppression of Word Reading

Daniel N. Bub, Michael E.J. Masson, and Christopher E. Lalonde

University of Victoria, Victoria, British Columbia, Canada

ABSTRACT—The development of cognitive control and its relation to overcoming Stroop interference was assessed in a sample (N = 65) of elementary-school children. Subjects alternately performed Stroop color-naming trials and word-reading trials. In separate blocks, the colored Stroop items were non-color words (incongruent condition) or rows of asterisks (neutral condition). Younger children showed both larger Stroop interference in error rates and a greater slowing of word reading in the incongruent condition compared with older children. We conducted analyses of response time distributions that assessed the degree of word-reading suppression applied by younger and older children. Surprisingly, these analyses indicated that younger children engaged in stronger suppression than older children. We propose that greater Stroop interference among younger children is not due to lack of ability to suppress word reading, but instead is the result of a failure to consistently maintain the task set of color naming.

Resolving competition between two operations invited by a stimulus configuration requires a form of cognitive control. For example, the operations of word reading and color naming come into conflict in the standard Stroop (1935) task, in which naming the color of a written word is slower than naming the color of a neutral stimulus, such as a row of Xs. To resolve the conflict experienced in this paradigm, the powerful tendency to read the word must be overcome in favor of responding to the color dimension. The skill that must be developed to deal with such conflict likely includes a number of different operations.

Previous studies have shown that Stroop interference is larger among children than among adults (e.g., Carter, Mintun, & Cohen, 1995; Comalli, Wapner, & Werner, 1962; Guttentag & Haith, 1978; Vurpillot & Ball, 1979). One possibility that emerges from these age-related differences is that children may be less able to suppress irrelevant stimulus dimensions and therefore may experience more interference than adults, and that this ability may have a developmental trajectory during childhood. Other suppression paradigms, such as the stop-signal and negative priming tasks, have provided evidence that among children in the general age range of 5 to 12 years, younger children (e.g., 5–8 years) are deficient in inhibitory control (Ridderinkhof, Band, & Logan, 1999; Tipper, Bourque, Anderson, & Brehaut, 1989), although Pritchard and Neumann (2004) found no evidence for reduced suppression in younger children in a negative priming task.

An alternative explanation is that inhibitory control may be in place, but younger children have difficulty maintaining the task set of naming the color of a stimulus. This proposal fits with computational models of interference in which the task set modulates pathways relevant to the conflicting tasks (e.g., Cohen, Dunbar, & McClelland, 1990; Yeung, Botvinick, & Cohen, 2004). In the model of Cohen et al., input from the mental set associated with the task of color naming strengthens the mapping from color percepts to responses, allowing the subject to overcome the competing mapping between word percepts and responses. Thus, the primary reason that children, primarily younger ones, show greater interference than adults during a Stroop task may be that they are less able consistently to apply the task set for color naming, rather than that they are less able to inhibit incompatible word responses. In keeping with this suggestion, Ridderinkhof, van der Molen, Band, and Bashore (1997) showed that in a task requiring a response to a central item flanked by distractors, interference effects with children in the age range of 5 to 12 years were larger for the younger children because of lower efficiency in maintaining appropriate stimulus-response mappings, not because of a failure to suppress the irrelevant distractors.

In summary, younger children may be more susceptible to Stroop interference either because they fail to suppress the irrelevant word dimension or because they inconsistently apply the color-naming task set across trials. One avenue to distinguish these two possibilities is to apply some method of

Address correspondence to Michael Masson, Department of Psychology, University of Victoria, PO Box 3050 STN CSC, Victoria, British Columbia V8W 3P5, Canada, e-mail: mmasson@uvic.ca.
determining the degree of suppression used by younger and older children in coping with the competing demands of a task such as Stroop color naming. Ridderinkhof (2002) has suggested that strong versus weak suppression of an incompatible response will have differential influences on trials with short versus long latency responses. To understand the logic of this argument, consider Figure 1a, which shows hypothetical cumulative density functions for color-naming response times in the Stroop task. One function represents responses in a neutral condition (e.g., a row of Xs), and the other three functions represent performance in an incongruent Stroop condition under conditions of strong, weak, or no suppression. Notice that, relative to the weak-suppression case, strong suppression reduces response latencies on incongruent trials, but the size of this reduction increases as response times extend into the part of the distribution corresponding to longer latencies.

This pattern is due to the speed with which processing of relevant and irrelevant dimensions leads to a final response, and to the speed with which suppression counteracts interference from the irrelevant dimension. Assume that producing a correct response based on the relevant dimension requires reaching a threshold of activation and that the time required to reach threshold varies across trials, producing a typical positively skewed response time distribution and the resulting sigmoid cumulative response time distribution (see the small graphs at the top of Fig. 1a and the cumulative distribution function for the neutral condition in Fig 1a). 1 Interference from the irrelevant dimension accrues over time and has a stronger effect on those trials on which relevant activation rises more slowly, producing a longer response latency (see the function for the no-suppression condition in Fig. 1a). On relatively fast trials, relevant activation accumulates quickly enough that response threshold is reached before interference has much impact (note that horizontal distance between cumulative density functions in Fig. 1a represents the magnitude of the interference effect and that the distance between the neutral and no-suppression functions increases as response time increases). Assume further that the effect of suppression accrues over time. The implication of suppression taking time to build is that only relatively slow responses can be made faster by suppression—the faster responses will already have been executed by the time suppression can counteract interference (see the function for weak suppression in Fig. 1a and note how the horizontal distance between that function and the no-suppression function is apparent only for longer response times). With stronger suppression, the influence on interference develops more quickly, and suppression will therefore affect even those trials associated with faster responses (note the horizontal distance between the strong-suppression and no-suppression functions in Fig. 1a).

These comparisons can be summarized by taking the horizontal difference between corresponding points on the neutral-condition function and each of the incongruent-condition functions in turn. Plotting these differences over the course of the response time distribution generates a delta plot representing the size of interference effects at different points in that distribution (Ridderinkhof, 2002). As shown in Figure 1b, the absence of suppression is associated with a monotonically increasing delta function with a steep slope, meaning that interference effects are larger for longer response latencies. The

---

1We use an activation metaphor only for convenience. One could just as easily characterize the generation of a response in an interference task as arising from some other mechanism, such as the accumulation of relevant evidence in a random-walk or diffusion model (e.g., Ratcliff, 1988; Ratcliff, Gomez, & McKoon, 2004). Similarly, suppression of a process could be characterized as a reduction in the rate at which relevant evidence is accumulated in a random walk (the drift rate in a diffusion model).
presence of suppression is indicated by a delta function that has a more shallow slope (the function for weak suppression in Fig. 1b), and in the case of strong suppression, the function may even become downward concave (the strong-suppression function in Fig. 1b). This concave trend results because responses at the longest latencies are quite strongly affected by suppression and therefore exhibit substantial savings, whereas responses at the shortest latencies remain relatively unaffected by suppression (compare the strong-suppression and no-suppression functions in Fig. 1a). Thus, differences in amount of suppression between two conditions or populations can be detected by examining delta functions (Ridderinkhof, 2002). In particular, if different age groups engage different degrees of suppression, then their delta functions should differ in slope. The function for the group with stronger suppression may even be concave downward.

In the study reported in this article, we used delta plots to examine suppression in the Stroop task among younger and older children. In addition, we applied a convergent method to assess suppression of word reading in younger and older children. In a previous study, we showed that word reading is substantially slowed as a result of color naming in a Stroop interference task (Masson, Bub, Woodward, & Chan, 2003). Subjects alternated between naming a color and reading a word. In one block of trials, the incongruent block, color was carried by a word, setting the stage for Stroop interference. In the other block, the neutral block, color was carried by a row of asterisks. Immediately after naming the color on a given trial, subjects switched to reading aloud a word printed in black. Modulation of word reading was demonstrated by the finding that reading latency was about 40 to 50 ms longer in the incongruent than the neutral block. This measure provides a convergent method of assessing the extent to which subjects suppress word reading when overcoming Stroop interference. We therefore used this method to test children of varying ages in a task-switching paradigm and examined the resulting response time distributions.

**METHOD**

**Subjects**

A sample of 65 children (35 boys and 30 girls) ranging in age from 7 years 3 months to 11 years 0 months (mean age = 9 years 1 month, SD = 1 year 0 months) was recruited from two middle-class elementary schools.

**Materials**

Five colors (blue, green, pink, red, and yellow) and five words (back, cold, face, home, and look) were used for the color-naming and word-reading tasks, respectively. The five words were also used in the incongruent condition of the color-naming task. We selected words that have high frequency and are typically acquired at an early age (mean frequency = 491 per million—Kucera & Francis, 1967; mean age of acquisition = 2.1 years—Gilhooly & Logie, 1980). Reading ability was assessed using the spelling and word-decoding subtests of the Wide Range Achievement Test (WRAT3; Wilkinson, 1993).

**Procedure**

Subjects were tested individually in a quiet room in school. An experimenter explained and administered the tasks, consisting of a “computer game” and a “paper and pencil” test. These two tasks were administered in counterbalanced order across the subjects. The paper-and-pencil test was the WRAT3.

Computer-controlled assessment of Stroop interference and modulation of word reading involved a sequence of trials, each of which required subjects to name the color of a stimulus, then read a word aloud. These trials were administered on a Macintosh laptop computer programmed so that vocal responses could be detected and timed using the built-in microphone. Sensitivity of the microphone was adjusted at the start of the task by having the subject read a series of words.

Each trial began with the presentation of a row of five equal signs in the middle of the monitor. A colored stimulus (a word in the incongruent condition or a row of four asterisks in the neutral condition) appeared below the equal signs after a 250-ms delay and disappeared when a color-naming response was detected. The experimenter pressed a key on a keypad attached to the computer to classify the response as correct or as an error (incorrect response, false start, or other speech error), then a word printed in lowercase black letters appeared above the equal signs. The word and the equal signs were erased when a vocal response was detected. The experimenter pressed a key to classify the response as correct or incorrect. After a delay of 250 ms, the next trial began.

The incongruent and neutral versions of the color-naming/word-reading trials were presented in alternating blocks of 10 trials each (e.g., 10 neutral trials, 10 incongruent trials). Each incongruent trial was structured so that the color-naming and word-reading stimuli were two different words. Subjects were given one practice block in each condition (neutral and incongruent), then an alternating sequence of four critical blocks of each type, providing 40 observations in each condition. The order in which the two types of blocks alternated was counterbalanced across subjects.

**RESULTS**

Mean age-normed percentile scores on the WRAT3 spelling and word-decoding subtests were 69.9 (SD = 24.1) and 61.4 (SD = 25.6). Response time data for color naming and word reading were based on correct responses, with outliers excluded. Outliers were defined as response times outside the range of 200 through 4,300 ms for the color-naming task and outside the range of 200 through 2,700 ms for the word-reading task. The upper bounds of these ranges were set such that no more than
0.5% of the correct responses were classified as outliers (Ulrich & Miller, 1994).

To test for Stroop and modulation effects, we computed analyses of variance for the response time and error data. A significance level of .05 was used for all statistical tests reported here. The response time data clearly showed that subjects produced a large Stroop interference effect (1,267 ms vs. 1,042 ms), $F(1, 64) = 166.29$, $\eta_p^2 = .72$, as well as robust modulation of word reading (943 ms vs. 861 ms), $F(1, 64) = 86.66$, $\eta_p^2 = .58$, with longer latencies in the incongruent condition in both cases. The error data also reflected a Stroop effect (18.0% vs. 7.0%), $F(1, 64) = 58.42$, $\eta_p^2 = .48$, but no reliable modulation effect (12.1% vs. 10.6%), $F < 2$. These results indicate that the children in our sample showed a large Stroop interference effect and also showed a marked slowing of word reading in the incongruent condition. Recall that words to be read were always printed in black, but in the incongruent condition were presented immediately following a color-naming response to color that was carried by a word. This modulation of word reading replicates the modulation effect for young adults we reported in our previous study (Masson et al., 2003).

Age-related differences in cognitive control, as assessed by Stroop interference and word-reading modulation (increased reading latency in the incongruent block), were examined by constructing cumulative density functions and corresponding delta plots for the response time and accuracy data, as described earlier. These plots were constructed separately for two age groups based on a median split of the subjects. Children in the younger group were less than 9 years old ($n = 32$), and the older children were 9 years and above ($n = 33$). Cumulative density functions were constructed by forming quintiles for each subject on the basis of a rank ordering of correct response times within a condition. Five successive bins of equal or nearly equal size were created from the rank-ordered response times, and the mean of each bin was computed. The first quintile (the mean of the first bin) represented the first 20% of the distribution, the second quintile represented the second 20%, or 40% cumulatively, and so on. The mean response time for each quintile was computed across subjects, and these means are plotted against the cumulative probabilities in Figures 2a and 2c for color naming and word reading, respectively. The cumulative density functions for each subject were used to compute delta plots for each task by subtracting response time in the neutral condition from the incongruent condition.
from response time in the incongruent condition within each quintile. The mean of the differences for each quintile was computed across subjects, and these mean differences (delta values) are shown in Figures 2b and 2d for color naming and word reading, respectively. Given the error rates in the color-naming and word-reading tasks, there were on average about seven observations per subject in each quintile. Although this is a relatively small number, the analyses we report next show that a sufficiently stable result was obtained to allow clear conclusions.

The delta plot for color naming, representing the Stroop effect, shows a clear divergence for the two age groups. For older children, the Stroop effect consistently increased across the response time distribution. For younger children, however, the Stroop effect diminished at the quintile comprising the longest response latencies, forming a concave-downward function. This pattern is very surprising and indicates, contrary to what might be expected from previous conclusions regarding limitations on suppression among younger children, that these children applied more suppression of word reading than did the older children. The divergence was confirmed by an analysis of variance that revealed a significant interaction between quintile (first through fifth) and age group (younger vs. older), $F(4, 252) = 2.90, \eta_p^2 = .04$.

The word-reading task provides further evidence for the surprising idea that younger children applied greater suppression than older children during the color-naming task. Modulation of word reading, as shown in Figure 2d, increased systematically across quintiles for both groups, and this increase was larger for the younger children, $F(4, 252) = 2.48, \eta_p^2 = .04$. Our claim is that the modulation effect is the outcome of suppression of word-reading processes in the incongruent block and that this suppression plays a causal role in controlling response conflict during color naming.

Word-reading suppression during color naming can also be measured by assessing word reading immediately following different outcomes of color-naming trials. Color-naming attempts vary with respect to both speed and accuracy. According to our characterization of the dynamics of suppression, it is reasonable to assume that in the incongruent block, slower responses involve greater suppression of word reading, which would then be reflected as slowed responding on the subsequent word-reading trial. Similarly, evidence from error-detection studies suggests that suppression following an incongruent trial should be increased in cases of error relative to cases in which a correct response is made (e.g., Yeung et al., 2004). To test these possibilities, we examined word-reading times contingent on either the speed or the correctness of the response on the immediately preceding color-naming trial.

Word-reading performance was conditionalized on a median split of response time on the preceding color-naming trial for both the incongruent and the neutral blocks. In the incongruent blocks, word reading was faster following a fast color-naming response than following a slow color-naming response (1,007 ms vs. 1,057 ms), $F(1, 64) = 4.06, \eta_p^2 = .06$. No significant effect was observed for word reading in the neutral blocks (941 ms vs. 976 ms), $F < 2$, indicating that the contingency seen in the incongruent blocks was not simply a general consequence of transitory fluctuations in response speed. Word-reading latency was also conditionalized on correctness of responding on the preceding color-naming trial. For incongruent trials, latencies were shorter following a correct color-naming response versus an error (982 ms vs. 1,057 ms), $F(1, 61) = 5.14, \eta_p^2 = .03$, but again, no significant effect was found in neutral trials (945 ms vs. 967 ms), $F < 1$. These conditional analyses show that suppression of word reading is tied to the degree of interference experienced during Stroop color naming: More interference leads to more suppression of word reading.

Despite evidence for greater suppression of word reading by younger children, the following analyses of errors (which typically involved execution of the opposing task) suggest that younger children were less successful than older children at escaping task conflict, especially when making fast responses. The occurrence of fast errors implies direct activation of inappropriate responses (Gratton, Coles, & Donchin, 1992). We plotted performance accuracy against response latency to generate conditional accuracy functions (Ridderinkhof, 2002). For the incongruent and neutral blocks in each task, we rank-ordered all trials, regardless of accuracy, by response time to form equal-sized quintiles. The percentage of accurate responses was computed for each quintile, and the resulting conditional accuracy functions are plotted in Figures 3a and 3c for color naming and word reading, respectively. These plots show relatively low accuracy at the quintiles containing the shortest and the longest response latencies for both tasks. The delta plots for each task, created by subtracting accuracy in the neutral condition from accuracy in the incongruent condition at each quintile, are shown in Figures 3b and 3d. For the word-reading task, there were no significant age-related differences in the modulation effect, $F < 1.1$. An analysis of the Stroop effect across quintiles for the two age groups showed that the effect was larger for younger children than for older children, $F(1, 63) = 23.89, \eta_p^2 = .27$, confirming previous reports of age-related differences in Stroop interference. A reliable age-by-quintile interaction showed that this difference was larger at the earliest quintiles, $F(4, 252) = 3.09, \eta_p^2 = .05$, indicating that the tendency for errors to be fast was greater for younger children than for older children. Thus, greater suppression of word reading, evinced by response latency analyses, is not necessarily adequate to prevent task conflict from leading to fast, incorrect responses under conditions of conflict (see Ridderinkhof, 2002, p. 509, for a similar result with contextually induced suppression in adults).
DISCUSSION

Conflict occurring during a Stroop color-naming task was greater for younger than older children, at least with respect to response accuracy. This result is consistent with most previous evidence (e.g., Carter et al., 1995; Comalli et al., 1962). We also examined a common assumption that this age-related difference is due to poorer control of suppression of the irrelevant word-reading task in younger children. Although widely held, this assumption has received relatively little empirical support, and indeed a recent study by Pritchard and Neumann (2004) did not show developmental differences in the magnitude of suppression as measured by negative priming. We sought evidence for developmental differences in suppression by applying a particularly informative analysis of response time distributions (Ridderinkhof, 2002). According to this analysis, longer-latency trials benefit the most from the effect of suppression. As strength of suppression increases, its influence extends to trials with shorter and shorter latencies. This pattern of influence results in a characteristic delta function representing the magnitude of the Stroop effect at different quantiles in a response time distribution. Specifically, under stronger suppression, the delta function has a more shallow slope, and in extreme cases even takes on a concave-downward form.

In this study, the delta function for the Stroop effect not only had a more shallow slope for children younger than 9 years of age than for older children, but also tended to be concave downward for the younger children. This surprising result suggests that younger children not only are quite capable of suppression of word reading, but actually engage in a greater degree of suppression than older children do. We introduced a further measure of suppression of word reading, by examining the response to a word printed in black immediately after each color-naming response. In previous experiments using this measure with adults, we showed modulation of word reading associated with Stroop color naming (Masson et al., 2003). In the present study, this measure provided results consistent with the evidence from distributional analyses of the Stroop effect: Younger children showed greater modulation of word reading than older children did. These two results provide converging evidence for relatively strong suppression of word reading by children younger than age 9.

What, then, might be the reason for larger Stroop interference among younger children, if not a failure of word-reading suppression? Some computational accounts of Stroop interference,
such as the model proposed by Cohen et al. (1990), emphasize the shift in attention weights to different task sets. For example, successful Stroop performance involves an increase in the attention weight for the color-naming task set, allowing the observer to overcome the habitually stronger tendency to read a word rather than name a color. The result of this shift is that the color-naming task set prevails over word reading. Younger children may be less consistent than older children at maintaining the color-naming task set, leaving them to face stronger competition from a word, as suggested by the delta plots for accuracy. In an effort to overcome this competition, younger children resort to stronger suppression (see Müller, Dick, Gela, Overton, & Zelazo, 2005, for another example of greater suppression in younger children).

Our interpretation that younger children are deficient in maintaining an appropriate task set receives support from a number of studies (e.g., Diamond & Taylor, 1996; Zelazo, Müller, Frye, & Marcovitch, 2003) showing that very young children tend to perseverate on a particular stimulus dimension (e.g., color) despite being able to articulate instructions that require them to shift to a competing stimulus dimension (e.g., shape). Similarly, we observed large Stroop effects on error rates among young children. Errors typically involved execution of the wrong task (e.g., reading the word instead of naming the color), as would be expected if an important source of developmental differences in cognitive control, at least in the context of Stroop interference, is the ability to maintain attention on the appropriate task set. This interpretation, coupled with the evidence for greater suppression of word reading in younger than older children, raises the possibility of a distinction between mechanisms responsible for maintaining relevant task set and those responsible for suppressing an irrelevant task.

Acknowledgments—This research was supported by the Canadian Language and Literacy Research Network. We thank Marnie Jedynak, Mary Jane Richards, Jennifer Thomas, Megan Vanderslys, and Jin Wang for their assistance. We are grateful to Fairburn and Margaret Jenkins elementary schools for their cooperation.

REFERENCES


(Received 3/9/05; Revision accepted 4/21/05; Final materials received 5/4/05)