Japanese Psychological Research
2004, Volume 46, No. 1, 56–64

Short Report

Integration versus separation in Stroop-like counting interference tasks[1]

MIYA MUROI[2]
Department of Psychology, College of Humanities and Sciences, Nihon University, Sakurajosui, Setagaya-ku, Tokyo 156-8550, Japan

COLIN M. MACLEOD
Department of Life Sciences, University of Toronto at Scarborough, Scarborough, Ontario M1C 1A4, Canada

Abstract: Two versions of a Stroop-like counting interference task were compared to examine how irrelevant information is ignored. In the integrated task, participants enumerated digits (inconsistent condition) or letters (neutral condition) while attempting to ignore the identity of the characters. In the newly created separated task, participants enumerated asterisks while attempting to ignore a single digit (inconsistent condition) or letter (neutral condition) at fixation. Interference (longer responses in the inconsistent condition than in the neutral condition) was small but significant in both tasks but was not reduced by separation. Contrasted with the pattern in the color–word Stroop task, these results suggest that dimensional separation has different effects on interference depending on the overall amount of interference.

Key words: counting Stroop task, separation, interference.

Since Stroop (1935) published his landmark article in which he showed that naming the ink color of a word was delayed when that ink color was inconsistent with the color denoted by the word (e.g., the word RED printed in green ink, say "green"), there have been a great many studies of the Stroop effect (MacLeod, 1991). Along the way, many variations of the task that now bears Stroop's name have been developed (MacLeod, 1991, pp. 165–170). These variations include the picture-word analog (e.g., say "cat" to a picture of a cat with the irrelevant word TABLE written inside it; Lupker, 1979), and the spatial analog (e.g., say "above" to the word BELOW printed above the fixation point; Palef, 1978).

Among these variations is one in which the participant must enumerate the elements in a display, ignoring their identity. This is considerably harder when the items to be enumerated are incompatible digits (e.g., say "three" to 444) rather than non-numerical characters

[1] This project was supported by research grant A7459 from the Natural Sciences and Engineering Research Council of Canada. The work was carried out while Miya Muroi spent the 1998–1999 academic year at the University of Toronto at Scarborough, supported by an exchange program with Kyoto University. We thank Penny MacDonald for comments on the initial draft of this article.

[2] Correspondence concerning this article should be sent to: Miya Muroi, Department of Psychology, College of Humanities and Sciences, Nihon University, 3-25-40 Sakurajosui, Setagaya-ku, Tokyo 156-8550 Japan (Email: muroi@chs.nihon-u.ac.jp) or Colin MacLeod, Department of Psychology, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada (Email: cmacleod@uwaterloo.ca).

(e.g., say "three" to ***). Introduced by Windes (1968), who reported reliable interference, this counting Stroop task has not been used much over the years. Very recently, however, it has attracted renewed interest in the domains of both cognition and neuroscience because of its utility in examining differences in interference for different numbers (Pavese & Umiltà, 1998) and because responding is done by button pressing, avoiding the need for voice responding that may contaminate data in neuroimaging studies (Bush, Whalen, Rosen, Jenike, McInerney, & Rauch, 1998).

Interference in separated and integrated versions of the Stroop task

Originally, relevant (color) and irrelevant (word) information were integrated in the Stroop task (Stroop, 1935). However, the two kinds of information often are presented separately, as in stimulus onset asynchrony or picture-word variants of the task (MacLeod, 1998). To investigate the mechanisms of information processing in both the separated and the integrated versions, it is necessary to compare responses in the two versions (integrated and separated) of the same task. Using the counting Stroop task, this was our purpose in the present study.

It is by now clear, consistent with an analysis along the lines championed by Garner (1974), that interference in a separated version of the Stroop task is less than that in an integrated version (Underwood, 1976; Flowers & Stoup, 1977; Gatti & Egeth, 1978; Kahneman & Henik, 1981; MacLeod, 1998). In a card-sorting task, Flowers and Stoup (1977) observed less interference when words in a neutral color were printed inside outline color rectangles and participants named the colors of the rectangles than when words themselves were printed in an ink color and participants named their colors. Gatti and Egeth (1978) showed that as the physical distance between the locations of the irrelevant word and the relevant color increased, interference decreased, suggesting a gradient of separation. Most recently, MacLeod (1998) investigated the effect of training on an integrated version (word printed in color) and a separated version (color bar beside word) of the Stroop task and found that interference was smaller for the separated case than for the integrated case throughout training, but particularly early in training.

To investigate the effect on interference of separation of the relevant enumerating process from the irrelevant digit identification process, we carried out the two counting Stroop tasks illustrated in Figure 1. In the integrated task, participants counted digits (inconsistent condition) or letters (neutral condition) while attempting to ignore the actual stimulus identity that they were enumerating. In our newly created separated task, participants enumerated asterisks while attempting to ignore a single digit (inconsistent condition) or letter (neutral condition) at fixation. We explored the pattern of interference in these inconsistent and neutral conditions. In both tasks, we also included a control condition where an asterisk was presented as a character to be enumerated to investigate the pattern of response with exactly the same stimuli in both tasks; consequently, we analyzed this condition separately from the other two conditions. In both tasks, we used the digits 1–5 in the inconsistent condition, five different letters in the neutral condition, and one character in the control condition.

EXPERIMENT 1

Method

Participants

Thirty-five students at the University of Toronto at Scarborough participated for bonus points in their introductory psychology course. They were randomly assigned to the integrated task ($n = 17$) or the separated task ($n = 18$) and were tested individually.

Experimental design

There was one between-subjects variable – task (integrated vs. separated). There were two within-subject variables – consistency (inconsistent, neutral, or control) and number of characters (1–5 characters to be enumerated).

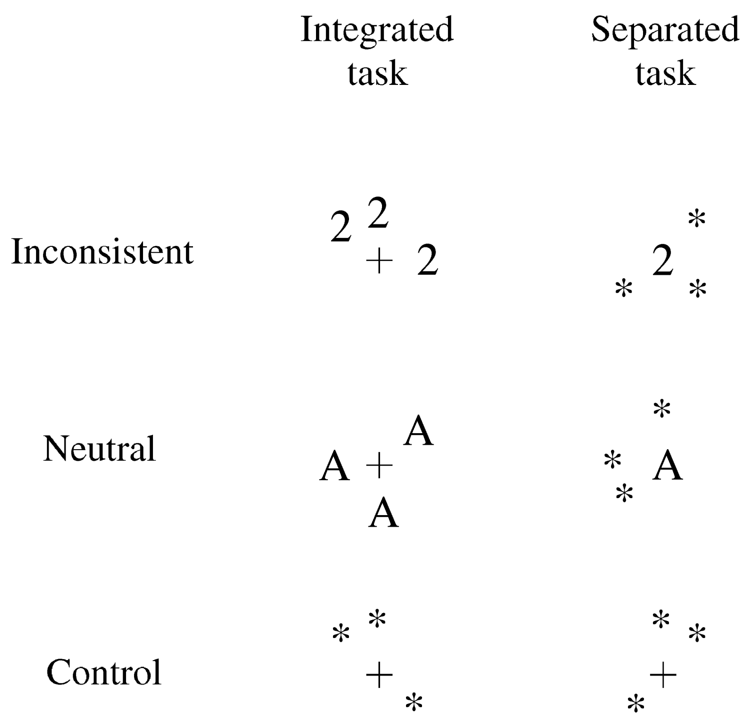

Figure 1. Examples of the inconsistent, neutral, and control stimuli in the integrated task (left column) and the separated task (right column) in Experiment 1. In all cases, the correct response is "three".

Materials

In the integrated task, digits (1, 2, 3, 4, or 5) in the inconsistent condition, letters (A, B, C, D, or G) in the neutral condition, and asterisks in the control condition served as the stimulus items to be enumerated.[3] All characters to be enumerated were identical within a trial. In our analyses, we collapsed over stimulus identity.

In the separated task, instead of presenting digits, letters, or asterisks as the items to be enumerated, asterisks were always presented to be enumerated. A digit (inconsistent condition), a letter (neutral condition), or a plus sign (control condition) was presented at a central fixation point surrounded by the asterisks that were to be enumerated.

In both tasks, we excluded consistent trials, wherein the stimulus digit and response count agreed (e.g., 333), in part because Pavese and Umiltà (1998) had difficulty with them and in part because MacLeod and MacDonald (2000) presented evidence that the apparent facilitation on such trials is illusory. To balance the deletion of consistent items, trials corresponding to one A, two Bs, three Cs, four Ds, and five Gs were excluded from the neutral condition.

The distance between the central character and each character to be enumerated on the imaginary circle was 3 cm.[4] Each letter was 3.5×6.0 mm in size. The distance between the participant and the monitor was 65 cm.

Procedure

In the integrated task, 1 s after the fixation point "+" appeared at the center of the screen, the characters that were to be enumerated appeared in 1–5 of the eight locations forming an imaginary circle around the fixation. The particular

[3] The letters E and F were not used in the neutral condition because each is the initial letter of a number.

[4] We adjusted the monitor to present 80×25 characters.

locations of characters were randomized on each trial. Participants were asked to enumerate the characters around the center and to respond aloud as quickly as possible, ignoring the identity of the characters and avoiding mistakes.

In the separated task, 1 s after the fixation point appeared, asterisks were presented in 1–5 of the eight locations on the imaginary circle. Simultaneously, at the central point, a single character replaced the fixation point: this could be a digit (inconsistent condition), a letter (neutral condition), or a plus sign (control condition). Participants were asked to enumerate the asterisks and to give their response aloud, as quickly as possible, ignoring the character at the central fixation point.

In both tasks, the entire stimulus display remained on the screen until the participant responded. The next trial began after the experimenter had input the participant's answer. In each task, the percentage of trials in each condition (inconsistent, neutral, or control), for each number of characters (1–5), and for each stimulus identity (four types for each number of characters) was equal. After 20 practice trials, two experimental blocks that included 600 trials were carried out. Order of trial presentation was randomized within a session.

Apparatus

Data were collected using IBM-compatible 486 computers equipped with Mitsubishi (Tokyo, Japan) 15-inch color monitors. When a participant spoke into a Realistic Highball-7 microphone (Radio Shack, Fort Worth, TX, USA), the signal was amplified by a Realistic SA-150 stereo amplifier (Radio Shack, Fort Worth, TX, USA) and input through a modified keyboard as if the hyphen key had been pressed. Programming was in QUICKBASIC 4.5 (Redmond, WA, USA) with millisecond accuracy timing routines taken from Graves and Bradley (1991).[5] The screen background color was black (palate #0); instructions and all stimuli were presented in white (palate #7). All materials were presented in regular 80-character-per-line font.

[5] Timing is accomplished in the Graves and Bradley (1991) method by using machine language routines called from the main program in QUICKBASIC.

Data analysis

For each participant, mean correct response times in each task were computed for every stimulus identity and for every number of characters in each consistency condition. Then we averaged the data across stimulus identity. Response times shorter than 200 ms or longer than 2000 ms, together with those from incorrect trials, were excluded from these calculations.

Results

Trials excluded from response time calculations in the inconsistent, neutral, and control conditions accounted, respectively, for 5.9%, 2.8%, and 3.5% of responses in the integrated task, and 5.6%, 5.0%, and 4.7% of responses in the separated task.

Interference

Figure 2 shows the mean response times in the integrated task (top panel) and in the separated task (bottom panel) for each number of characters in each consistency condition. To assess the pattern of interference, we used a mixed $2 \times 2 \times 5$ ANOVA. The between-subjects factor was task (integrated or separated); the within-subject factors were consistency (inconsistent or neutral) and number (1–5 characters).

The main effects of consistency condition and of number, and the consistency × number interaction, all were significant, $F(1,33) = 61.80$, $p < 0.001$, $F(4,132) = 119.70$, $p < 0.001$, $F(4,132) = 2.96$, $p < 0.05$, respectively. The main effect of task, the task × consistency interaction, the task × number interaction, and the three-way interaction were not significant, $F(1,33) = 0.03$, $p > 0.20$, $F(1,33) = 3.91$, $0.05 < p < 0.10$, $F(4,132) = 0.88$, $p > 0.20$, $F(4,132) = 1.67$, $p > 0.20$, respectively. Further analysis and multiple comparisons showed that response times in the inconsistent condition were significantly longer than those in the neutral condition, and that the differences

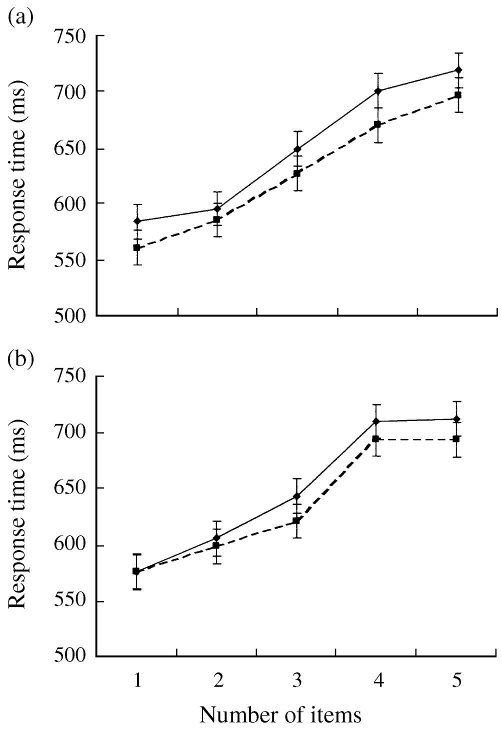

Figure 2. Experiment 1: Mean response time (in ms) in (a) the integrated task and (b) the separated task as a function of the number of characters to be counted (range, 1–5). The error bars in each panel represent the 95% within-subject confidence limit for the data in that panel (see Loftus & Masson, 1994). (—◆—) Inconsistent, (--■--) neutral.

as count increased were significant except between counts of four and five in both tasks ($p < 0.05$).[6]

Control condition

Figure 3 shows the mean response times in the control condition in both tasks. A mixed 2×5 ANOVA compared the pattern of mean response times between two tasks with the same stimuli. The between-subjects factor was task (integrated or separated) and the within-subject factor was number (1–5 characters). The main effect of number was significant, $F(4,132) = 116.60$, $p < 0.001$. Neither the main

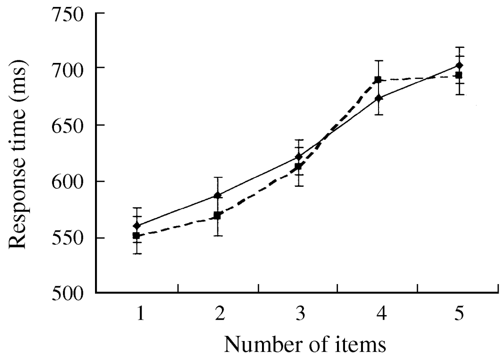

Figure 3. Experiment 1: Mean response time (in ms) in the control condition in the integrated task and in the separated task as a function of the number of characters to be enumerated (range 1–5). The error bars are 95% confidence limits as in Figure 2. (—◆—) Integrated, (--■--) separated.

effect of task nor the interaction was significant, $F(1,33) = 0.06$, $F(4,132) = 1.25$, $p > 0.20$, respectively. Pair-wise tests showed that differences as count increased were significant except between counts of four and five ($p < 0.05$).

Discussion

First, we comment briefly on the number of items to be enumerated and on the control condition. Then we examine our major question: the differences in pattern between the integrated and separated task versions.

Response times increased as the number of to-be-counted characters increased in both the separated and integrated tasks. This result corresponds to that of Pavese and Umiltà (1998). However, because the number of characters in each task ranged from one to five, participants had to discriminate the two endpoints of the count (one and five) from only one other possible count in each case (two and five, respectively) whereas they had to discriminate intermediate counts (two, three, and four) from both a larger and a smaller possible count (Mandler & Shebo, 1982). As a result, when the number of characters to enumerate was five, the response times were relatively faster than would be expected otherwise. These patterns also replicate those of Pavese and Umiltà (1998).

[6] Ryan's method was used for all pair-wise comparisons (Ryan, 1960).

In the control condition, participants enumerated the characters in the same way in the two tasks, despite the differences between the tasks in the inconsistent and neutral conditions. This indicates that any differences between the tasks were not due to the overall strategy used in each task but rather to the stimulus differences themselves.

The integrated-separated distinction

In both tasks, response times in the inconsistent condition were longer than those in the neutral condition. That is, in the counting Stroop task, irrelevant digit information disrupted the relevant enumerating process, analogous to the color-word Stroop task. However, no main effect of task or no task interaction with any factor was significant. In other words, no effect of separation on interference was observed in the counting Stroop task. It is quite clear in a separated version of the color-word Stroop task that the interference is smaller than that in an integrated version of the task (Flowers & Stoup, 1977; MacLeod, 1998).

We will reconsider the difference between the color-word and counting Stroop tasks after the results of the next experiment. Experiment 2 is similar to Experiment 1 except for two changes, and serves to confirm the results just reported. First, we excluded the control condition to ensure that the difference between the results in this study and those in color-word Stroop studies with respect to the effect of separation was not due to inclusion of the control condition. Second, in the integrated task, we changed the character at the center (presented at the same time as the characters to be counted) from + to * to avoid presenting the same character as the fixation point not only in the separated task but also in the integrated task.

EXPERIMENT 2

Method

Participants

Thirty students at Kyoto University participated in either the integrated task ($n = 15$) or the separated task ($n = 15$). Each participant was randomly assigned to a task and was tested individually.

Experimental design

The experimental design was the same as that of the Experiment 1 except for the consistency condition variable, where the control condition was excluded. The locations and the numbers of characters were the same as in Experiment 1.

Procedures

An asterisk was presented at the central location after the fixation point in the integrated task. Because the control condition was excluded, 400 trials consisting of two blocks of 200 trials each were used. Other procedures were the same as those in Experiment 1.

Data were collected using OHTEC HL-5870B computers (Ohtec, Japan) equipped with FUJITSU FMV-DP97Y3 monitors (Fujitsu, Fukushima, Japan). Programming and the color and font used on the screen were the same, and the data analyses were done in the same way as those of Experiment 1.

Results

Trials excluded from response time calculations in the inconsistent and neutral conditions accounted for 2.8% and 2.5%, respectively, of responses in the integrated task, and 2.1% and 2.1%, respectively, of responses in the separated task.

Interference

Figure 4 shows mean response times in the integrated task (top panel) and the separated task (bottom panel) for each number of characters in each consistency condition. To assess the pattern of interference, we again conducted a mixed $2 \times 2 \times 5$ ANOVA. The between-subjects factor was task (integrated or separated); the within-subject factors were condition (inconsistent or neutral) and number (1–5 characters). The main effects of consistency condition and of number were significant, $F(1,33) = 23.31$, $p < 0.001$, $F(4,120) = 43.41$, $p <$

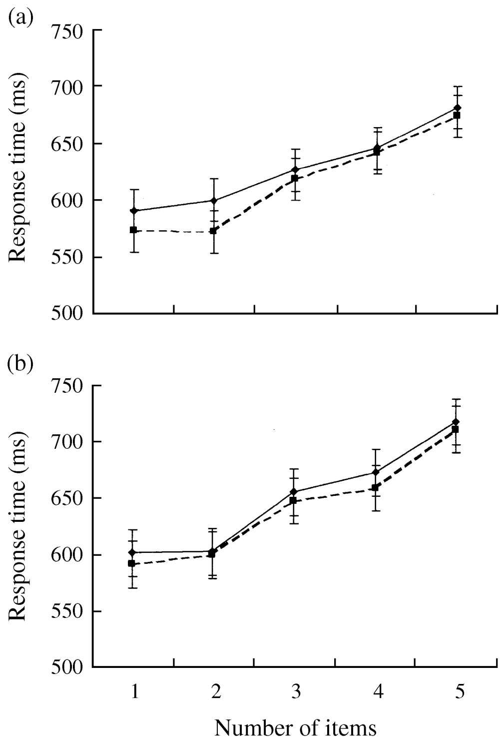

Figure 4. Experiment 2: Mean response time (RT in ms) in (a) the integrated task and (b) the separated task as a function of the number of characters to be counted (range 1–5). The error bars are 95% confidence limits as in Figure 2. (—◆—) Inconsistent, (--■--) neutral.

0.001, respectively. Pair-wise tests showed that the difference between any two counts was significant except between counts of one and two and counts of four and five ($p < 0.05$).

To confirm the effect of separation, we calculated the response time in the same way for the first and second blocks (the trials were separated in two blocks) and used a mixed $2 \times 2 \times 2 \times 5$ ANOVA. The between-subjects factor was task; the within-subject factors were block, condition, and number. The main effects of consistency and of number, and the block × number interaction, were significant, $F(1,30) = 22.62$, $p < 0.001$, $F(4,120) = 43.91$, $p < 0.001$, $F(4,120) = 4.05$, $p < 0.005$, respectively. No other main effects or interactions were significant ($p > 0.10$).

Discussion

In Experiment 2, response times increased as the number of characters increased in both tasks except between one and two and between four and five (although the response times between one and two were different in Experiment 1). The difference between the two experiments may be caused by the different pronunciations of the languages: English in Experiment 1 and Japanese in Experiment 2.

The integrated–separation distinction

The pattern of interference replicated that in Experiment 1. There was a significant main effect of the consistency condition but there was no main effect of the task nor did the task interact with any other factors. These results again suggest that the digit identification process was more automatic than the enumerating process. However, separation did not appear to affect the discrimination of relevant enumerating information from irrelevant digit information in the counting Stroop task. The mixed four-way ANOVA shows that the pattern of interference observed in this study was consistent across the two blocks. This result is quite different from that in the color-word Stroop task in which both separation and training reduce interference (Flowers & Stoup, 1977; MacLeod, 1998).

Another difference between the color-word Stroop task and the counting Stroop task is the amount of interference in the integrated versions. In the color-word task, the difference in response time between the inconsistent condition and the neutral condition is usually large (about 150 ms in MacLeod, 1998; Experiment 2). In our study, the difference between the two consistency conditions was about 17 ms in Experiment 1 and about 7 ms in Experiment 2. Our explanation of this difference relies on the concept of relative automaticity (MacLeod & Dunbar, 1988). Degree of automaticity of processing is a function of learning (indeed, if it can actually be achieved, fully automatic processing may simply be the endpoint of extensive learning), and interference declines as the difference in automaticity between relevant and irrelevant processing decreases. In the standard color-word task, the highly automatic identification of the color word disrupts the less automatic color-naming process and causes large interference. However, in the counting Stroop task, the small interference observed means that the difference in automaticity between the two processes is relatively small, although digit identification is more automatic than enumeration.

These differences between the two tasks imply that separation affects interference differently in two situations. When the difference in automaticity between relevant information processing and irrelevant information processing is small, as in the counting Stroop task, interference is small even in the integrated version and there is no room for separation to reduce interference, or interference reduction is too small to be measured. However, when the difference in automaticity between relevant and irrelevant information processing is large, as in the color-word Stroop task, interference is large in the integrated version and separation can reduce interference. That is, separation affects interference differently depending on the amount of interference.

Finally, small but significant interference in the integrated task in this study suggests that the difference in automaticity between the relevant enumerating process and irrelevant digit identification process was relatively small in the counting Stroop task, certainly when compared to the familiar color-word Stroop task. In addition, the fact that separation did not affect interference in the counting Stroop task suggests that separation, which reduces interference in the usual color-word Stroop task, has no effect on interference when the difference in automaticity of relevant and irrelevant processing is small.

References

Bush, G., Whalen, P. J., Rosen, B. R., Jenike, M. A., McInerney, S. C., & Rauch, S. L. (1998). The counting Stroop: An interference task specialized for functional neuroimaging – Validation study with functional MRI. *Human Brain Mapping*, **6**, 270–282.

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

Garner, W. R. (1974). *The processing of information and structure*. Hillsdale: Erlbaum.

Gatti, S. V., & Egeth, H. E. (1978). Failure of spatial selectivity in vision. *Bulletin of the Psychonomic Society*, **11**, 181–184.

Graves, R., & Bradley, R. (1991). Millisecond timing on the IBM PC/XT/AT and PS/2: A review of the options and corrections for the Graves and Bradley algorithm. *Behavior Research Methods, Instruments & Computers*, **23**, 377–379.

Kahneman, D., & Henik, A. (1981). Perceptual organization and attention. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 181–211). Hillsdale: Erlbaum.

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

Lupker, S. J. (1979). The semantic nature of response competition in the picture-word interference task. *Memory and Cognition*, **7**, 485–495.

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

MacLeod, C. M. (1998). Training on integrated versus separated Stroop tasks: The progression of interference and facilitation. *Memory and Cognition*, **26**, 201–211.

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.

MacLeod, C. M., & MacDonald, P. A. (2000). Interdimensional interference in the Stroop effect: Uncovering the cognitive and neural anatomy of attention. *Trends in Cognitive Sciences*, **4**, 383–391.

Mandler, G., & Shebo, B. J. (1982). Subitizing: An analysis of its component processes. *Journal of Experimental Psychology*, **111**, 1–22.

Palef, S. (1978). Judging pictorial and linguistic aspects of space. *Memory and Cognition*, **6**, 70–75.

Pavese, A., & Umiltà, C. (1998). Symbolic distance between numerosity and identity modulates Stroop interference. *Journal of Experimental Psychology: Human Perception and Performance*, **24**, 1535–1545.

Ryan, T. A. (1960). Significance tests for multiple comparisons of proportions, variances and other statistics. *Psychological Bulletin*, **57**, 318–328.

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

Underwood, G. (1976). Semantic interference from unattended printed words. *British Journal of Psychology*, **67**, 327–338.

Windes, J. D. (1968). Reaction time for numerical coding and naming numerals. *Journal of Experimental Psychology*, **78**, 318–322.

(Received August 6, 2001; accepted May 10, 2003)