Experimental note: Cross-modal, visual–auditory picture–word interference

H. SCHRIEFERS
Freie Universität Berlin, Berlin, West Germany

and

A. S. MEYER
Max-Planck-Institut für Psycholinguistik, Nijmegen, The Netherlands

(Charles W. Eriksen, *Sponsor*)

Conflicting results have been reported concerning the question of whether or not cross-modal, visual–auditory interference effects can be obtained in the Stroop task. In the present paper, we document such effects for a similar task, namely the picture–word interference task. Furthermore, we show that these effects are highly sensitive to the precise temporal relationship between the presentation of the target and the presentation of the distractor. The discussion of whether or not there are cross-modal Stroop effects would greatly benefit from experiments in which the timing of the stimulus presentation is carefully controlled.

Recently, there has been some discussion about whether or not cross-modal, visual–auditory interference effects can be obtained in the Stroop paradigm. Cowan and Barron (1987; see also Cowan, 1989a, 1989b) showed that naming of the colors in which rows of Xs were printed was slowed down when subjects simultaneously heard spoken color names, as compared with control conditions in which they repeatedly heard the word *the*, repetitions of the alphabet, or music, or in which no auditory distractors were presented.

This result, however, was not replicated by Miles and Jones (1989) or Miles, Madden, and Jones (1989). As the vivid discussion of these conflicting results shows, the question of whether or not there are cross-modal Stroop effects has important theoretical consequences, in particular for theories concerning selective attention and short-term storage of information. Hence, it seems worthwhile to investigate the conditions under which such effects might or might not arise.

In the classic version of the Stroop task (Stroop, 1935), the target (the color of the print) and the distractor (the meaning of the printed word) are necessarily presented simultaneously, as two aspects of one and the same physical stimulus. This is not always the case in modifications of the visual–visual Stroop task, in which the subject must react to one stimulus while ignoring another. Whether or not in these instances interference effects are obtained depends crucially on the temporal relationship between the

presentation of the target and the presentation of the distractor (see, e.g., M. O. Glaser & W. R. Glaser, 1982, and W. R. Glaser & Düngelhoff, 1984). This suggests that the outcomes of cross-modal, visual–auditory Stroop experiments might also, at least to some extent, depend on the timing of the targets and the distractors. In the studies mentioned above, the subjects were presented with cards on which large sets of stimuli appeared—such as, for instance, 100 strings of Xs printed in different colors. While naming the colors, the subjects heard the distractors. The dependent variable was the total time the subjects needed to name the colors of all stimuli on a card. Since the naming of the colors was self-paced and the distractors were presented at a fixed rate, the experimenter had no control over the temporal relationship between the processing of the visual stimuli and the preparation of the responses on the one hand, and the processing of the auditory stimuli on the other hand.

Below, the results of an experiment are presented, which demonstrate the existence of cross-modal, visual–auditory interference effects in a picture–word interference task; this task is closely related to the cross-modal Stroop task. Furthermore, the results show that the cross-modal interference effects are highly sensitive to the relative timing of the visual targets and the acoustic distractors.

METHOD

Subjects
Sixteen paid subjects, 8 men and 8 women, participated in the experiment. They were undergraduate students at Nijmegen University and native speakers of Dutch.

Materials
The visual materials consisted of 32 pictures of common objects (such as a house, a nail, or a hat). In a pilot study with an independent group

of subjects, 16 of these pictures had been named spontaneously in the same way by at least 95% of the subjects. These items were used in the critical trials; the remaining items served as fillers. The pictures were presented as white line-drawings on a black background, using a Vector General display.

Each critical picture was combined with four types of auditory distractor. The first distractor (associate) was a close associate of the picture name. The second distractor (phonological) was a noun that was phonologically related to the picture name, sharing at least the first two or three phonological segments with it. The third distractor (unrelated) was a noun unrelated to the picture name in sound and meaning. The fourth distractor (blanco), which was the same for all pictures, was the word *blanco* (English *blank*). In addition, the pictures were also presented alone, without any interfering stimuli; this was the *silence* condition. The filler pictures were not combined with associates or phonologically related words, but with three different unrelated words each and with the word *blanco*. In addition, they were also shown without interfering stimuli (silence).

The words, spoken by a female speaker of Dutch, were tape-recorded and digitized, using a sampling rate of 20 kHz. They were stored on the disk of a PDP-11/55 computer, which also controlled the experiment. The words were presented to the subjects through Sennheiser headphones.

Design

Each subject saw each picture 10 times—that is, twice under each of the distractor conditions described above. Each target-distractor combination was presented once with a stimulus onset asynchrony of 0 msec (SOA = 0) and once with a stimulus onset asynchrony of 300 msec (SOA = 300). In the former case, the onset of the target and of the distractor coincided; in the latter case, the onset of the distractor was delayed by 300 msec relative to the onset of the target. SOAs were blocked. For 8 subjects, the SOA = 0 condition was administered first, followed by the SOA = 300 condition; for the remaining subjects, the order of the two SOA conditions was reversed. Within each SOA block, the items were presented in a pseudorandom order. Repetitions of identical pictures in direct succession were avoided, and the order of the distractor conditions was counterbalanced across the pictures.

Procedure

The subjects were tested individually in a dimly lit soundproof room. A trial had the following, very simple structure: First, the picture was presented for 500 msec. Depending on the SOA block, the distractor was presented at the same time or with a delay of 300 msec. The presentation of the picture was followed by a pause of 2,600 msec. Then the next trial began. A voice key was used to measure the subject's reaction time, which was defined as the time interval between the onset of the picture and the onset of the naming response. There was a brief pause between the two SOA blocks.

RESULTS

Eight percent of the reaction times on critical trials were replaced by estimates, according to the procedure recommended by Winer (1971). In these cases, the subjects had used incorrect naming responses (i.e., names other than those expected on the basis of the results of the pilot study), or they had started their naming response with a filled pause (e.g., "uhm"), or the reaction times deviated by more than two standard deviations from the respective subject's and picture's mean naming latency. The data were submitted to an analysis of variance with SOA order (SOA = 0 followed by SOA = 300 vs. SOA = 300 followed by SOA = 0) as a between-subjects factor, and SOA, distractor condition, and picture as within-subjects factors. The mean naming latencies for the five distractor conditions with the two SOAs are given in Table 1.

Table 1
Mean Naming Latencies by Distractor Condition and SOA

Condition	SOA = 0	SOA = 300
Associate	664	617
Phonological	605	604
Unrelated	662	591
Blanco	648	610
Silence	617	600

Note—All values are in milliseconds.

The analysis of variance yielded significant main effects for SOA [$F(1,14) = 27.2, p < .001, MS_e = 28,696$] and for distractor condition [$F(4,56) = 17.7, p < .001, MS_e = 6,416$], as well as a significant interaction of these variables [$F(4,56) = 11.6, p < .001, MS_e = 8,154$]. Furthermore, there were significant interactions between SOA order on the one hand, and SOA [$F(1,14) = 16.2, p < .005, MS_e = 28,696$] and distractor condition [$F(4,56) = 2.5, p < .05, MS_e = 6,414$] on the other hand. The interaction of SOA order, SOA, and distractor condition was not significant [$F(4,56) = 1.4, p = .24, MS_e = 8,154$]. Hence, the pattern of the interaction between SOA and distractor condition was the same for both SOA orders.

An analysis of simple effects revealed a significant effect of the five distractor conditions at SOA = 0 [$F(4,56) = 22.6, p < .001$] and at SOA = 300 [$F(4,56) = 3.05, p < .05$]. Newman-Keuls tests showed that under SOA = 0, the associative, unrelated, and blanco conditions did not differ significantly from each other, but in all three of these conditions, the mean reaction times were longer than in the silence and phonological conditions ($p < .01$), which in their turn did not differ significantly from each other. Under SOA = 300, the only significant difference obtained between the associative and the unrelated conditions ($p < .05$). In particular, none of the four distractor conditions was significantly different from the silence condition.

DISCUSSION

The present results show that cross-modal, visual-auditory interference effects can be obtained in a picture-word interference task. This is reflected in the fact that at SOA = 0, longer naming latencies were observed in the associate, unrelated, and blanco conditions than in the silence condition. Furthermore, the interference effect was sensitive to the type of distractors used. Although interference effects were obtained in three of the distractor conditions, no such effect was obtained in the fourth condition, which is the phonological condition. A possible interpretation of this pattern of results is that the presentation of an irrelevant word interferes with the naming response relative to the silence condition, but that this interference effect is canceled in the phonological condition by a facilitatory effect due to the phonological similarity between the target name and the auditorily presented distractor. This interpretation is supported by the results of other experiments (see Schriefers, Meyer, & Levelt, 1990). Finally, and most importantly for the present discussion of the effects of the temporal relation of target and distractor, interference effects were only obtained at SOA = 0, but not at SOA = 300. When the auditory distractors are presented 300 msec after picture onset, none of the distractor conditions leads to a significant prolongation of the naming latencies relative to the silence condi-

tion. The distractors must have appeared too late to affect the naming response. Again, this pattern of results was replicated in other experiments (see Schriefers et al., 1990).

As has been mentioned above, W. R. Glaser and Düngelhoff (1984) showed that the outcome of intramodal picture–word interference experiments depends on the relative timing of the targets and the distractors. We have demonstrated that this is also true for cross-modal picture–word interference experiments. Given these findings, it seems highly likely that the results of cross-modal Stroop experiments also depend, to a certain extent, on the temporal relation of target and distractor presentation.

Obviously, this conclusion does not imply an explanation for why cross-modal interference effects were obtained by Cowan (1989a) and Cowan and Barron (1987), but not by Miles and Jones (1989) and Miles et al. (1989). But in none of these studies was the temporal relation between target and distractor systematically controlled for. Therefore, the paradigm used in these experiments might not have been the most suitable one for investigating the question of interest, which is whether or not the processing of a visual target and/or the preparation of a response can be affected by the presentation of an auditory distractor. More compelling evidence concerning this issue can be expected from experiments in which the relative timing of target and distractor presentation are systematically varied.

REFERENCES

COWAN, N. (1989a). The reality of cross-modal Stroop effects. *Perception & Psychophysics*, **45**, 87-88.

COWAN, N. (1989b). A reply to Miles, Madden, and Jones (1989): Mistakes and other flaws in the challenge to the cross-modal Stroop effect. *Perception & Psychophysics*, **45**, 82-84.

COWAN, N., & BARRON, A. (1987). Cross-modal, auditory-visual Stroop interference and possible implications for speech memory. *Perception & Psychophysics*, **41**, 393-401.

GLASER, M. O., & GLASER, W. R. (1982). Time course analysis of the Stroop phenomenon. *Journal of Experimental Psychology: Human Perception & Performance*, **8**, 875-894.

GLASER, W. R., & DÜNGELHOFF, F. J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception & Performance*, **10**, 640-654.

MILES, C., & JONES, D. M. (1989). The fallacy of the cross-modal Stroop effect: A rejoinder to Cowan (1989). *Perception & Psychophysics*, **45**, 85-86.

MILES, C., MADDEN, C., & JONES, D. M. (1989). Cross-modal, auditory-visual Stroop interference: A reply to Cowan and Barron (1987). *Perception & Psychophysics*, **45**, 77-81.

SCHRIEFERS, H., MEYER, A. S., & LEVELT, W. J. M. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory & Language*, **29**, 86-102.

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

WINER, J. B. (1971). *Statistical principles in experimental design*. New York: McGraw-Hill.