Journal of the International Neuropsychological Society (1996), 2, 467–470.

LETTER TO THE EDITOR

Paying attention to the Stroop effect?

AVISHAI HENIK
Department of Behavioral Sciences, Ben-Gurion University of the Negev, Beer-Sheva, Israel
(RECEIVED November 15, 1995; REVISED January 8, 1996; ACCEPTED February 21, 1996)

Studies of attention commonly use conflict situations, such as the Stroop task (Stroop, 1935), in which subjects are asked to attend to an object or a dimension and simultaneously ignore other objects or dimensions. Because of its reliability and robustness, the Stroop task is included in many neuropsychology batteries for investigating attentional processing in a variety of populations. Unfortunately, despite its wide application, the task is sometimes misapplied because some studies are not motivated by a conceptual framework, but only by empirical diagnostic criteria. The data resulting from some studies are either uninterpretable or, worse, misleading. To clarify matters I would like first to direct attention to Stroop's original study, thereby explaining the reasoning behind the various Stroop conditions, and then point out the various deviations from the original which reduce the potential value of these studies.

A Brief Description of Stroop's (1935) Original Study

Stroop (1935) was interested in creating an experimental method that would enable one to measure the interference of one stimulus dimension on another. The seminal paper that was published in 1935 described three experiments. In all experiments subjects were presented with cards that had words in black, patches of color, or words in color. The task was to read the words or name the colors as fast as possible. The time to perform the task was measured in seconds.

The first experiment measured interference of colors on reading words. Accordingly, there were two conditions: (1) names of colors in black ink (e.g., the word red in black), and (2) color words in an incongruent ink (e.g., the word red in purple). Subjects were asked to read 100 color names, and in the incongruent condition, also to ignore the colors. Stroop referred to the first condition as reading color names printed in black (RCNb) and to the second condition as reading color names where the color and the word are different (RCNd). On average, subjects were 2.3 s slower to read color names in the incongruent condition than when the same names were printed in black. Stroop concluded that this effect was not reliable, and further experiments have replicated this finding (MacLeod, 1991). That is, colors do not reliably interfere with reading of words.

The second experiment measured interference of words on naming colors. This experiment was also composed of two conditions: (1) squares of colors (e.g., a square of red color), and (2) an incongruent condition similar to that mentioned for the first experiment. Subjects were asked to name 100 colors, and in the incongruent condition, also to ignore the words. The first condition was termed naming color test (NC) and the second condition was termed naming color of word when the color of the print and the content of the word are different (NCWd). On average, subjects named the colors in the incongruent condition 47 s slower than they named the patches of colors. Stroop concluded that this was a reliable difference.

Many authors refer to the effect generated in the second experiment (and to similar effects), that is, the interference of an irrelevant word on naming color, as the "Stroop effect." The effect generated in the first experiment, that is, the interference of the irrelevant color on reading the words, is termed "reverse Stroop."

The third experiment was designed to look at "The effects of practice upon interference." In this experiment subjects practiced naming colors mainly by working with the incongruent condition (NCWd) for 8 days. Although the Stroop interference effect was still sizable even after 8 practice days, it was clear that practice reduced the interference of the word on color naming.

Stroop suggested that this pattern of results was due to training and to the nature of the associations between word stimuli and reading on the one hand and between color stimuli and naming responses on the other hand:

> ". . . the associations that have been formed between the word stimuli and the reading response are evidently more effective than those that have been formed between the color stimuli and the naming response . . . The word stimulus has been associated with the specific response 'to read,' while the color stimulus has been associ-

Reprint requests to: Avishai Henik, Department of Behavioral Sciences, Ben-Gurion University of the Negev, Beer-Sheva, Israel.

Table 1. Color naming and word reading in milliseconds as a function of group (PD or control) and condition (neutral or incongruent)

	Color naming			Word reading[a]		
	Neutral	Incongruent	Interference[b]	Neutral	Incongruent	Interference[b]
PD	828	974	146	650	730	80
Control	670	782	112	500	510	10

[a]Data for word reading are hypothetical.
[b]Interference is inconguent minus neutral.

ated with various responses: 'to admire,' 'to name,' 'to reach for,' 'to avoid,' etc." (pp. 659–660).

The Stroop Task in Neuropsychological Batteries

It is important to note that Stroop employed neutral and incongruent conditions in order to measure interference. The appropriate neutral trials were changed to suit the task; patches of color for color naming (Experiment 2) and words in black for reading (Experiment 1). Many neuropsychological batteries employ only three conditions, words in black for reading (RCNb), patches of color (NC), and words in color for color naming (NCWd).[1] Experiments that use the neutral (NC) and incongruent (NCWd) conditions enable researchers to look at the interference component that is reflected in the difference between the neutral and the incongruent conditions. This is similar to the effect reported by Stroop (1935) in Experiment 2. The size of the interference is thought to measure subjects' ability to selectively attend to the relevant color and to suppress processing of the irrelevant word.

There are certain common ways in which the data have been handled. I would like to present them and the problems they create. Since most researchers are interested in the Stroop effect and not in the reverse Stroop effect, most of the examples will be related to the Stroop effect. To demonstrate various points, I employ the data presented in Table 1. This table includes mean reaction time (RT) in milliseconds (ms) for the various experimental conditions. The color naming data were collected in a study of Parkinson Disease (PD) patients (Henik et al, 1993). We studied the Stroop effect in patients with early and late onset PD and age-matched control subjects. We used the trial by trial version of the task in which subjects are presented with one word on every trial and asked to name the color (under the color naming condition) or read the word (under the word reading condition). The color naming data were taken from the early onset PD patients and their age-matched control subjects. The word reading data are hypothetical.

Group comparisons for each condition separately. Many authors use the three conditions mentioned above and compare experimental (e.g., patient) and control groups on all three conditions and nothing more. Suppose one wants to look at the differences between patients (e.g., Huntington's disease, schizophrenia, Parkinson's disease) and normal controls. Often only the results of pairwise comparisons between experimental and control groups are reported for each of the three conditions (see Bamford et al., 1989, as an example for research in Huntington's disease; Dikmen et al., 1995, as an example for research in closed head injury; Hietanen & Teravainen, 1986, as an example for research in Parkinson's disease; Everett et al., 1989; Golden, 1976; Wapner & Krus, 1960; Wysocki & Sweet, 1985, for examples of research in schizophrenia).[2] Let us assume that there are significant differences between the groups on all of the conditions. What does it tell us? It tells us that the patients are typically slower to respond than normal controls. But this is well known. It is impossible to tell whether the two groups differ in interference. The data are not tested directly in order to look at the important aspect for which the test was designed—interference. Thus, it does not enlighten us about attentional processing in patients compared with controls. In order to test for the difference in interference one has to run a two-way ANOVA (Abramczyk et al., 1983) in which one factor is group (patient *vs.* control) and the other factor is Stroop condition (incongruent *vs.* neutral, or in Stroop's terms, NCWd *vs.* NC, correspondingly). There are other ways to carry out the appropriate test for the difference in interference (Buchanan et al., 1994), but it is clear that in order to look at differences in interference, separate comparisons of the two groups on each of the Stroop conditions are not useful.

This point can be demonstrated by reference to the color naming data presented in Table 1. PD patients showed an interference of 146 ms and the control subjects showed an

[1] Sometimes authors mention that they employ the Stroop task and do not describe the conditions. They may mention part 1 and part 2 and cite Stroop (1935) as a reference. It is not clear what they have in mind and what conditions they have actually employed.

[2] Wapner and his colleagues (1960) reported a significant interaction between group and condition, but the condition variable included not only neutral and incongruent but also the word reading, so that it is not clear if the difference in interference was the source of this interaction.

https://doi.org/10.1017/S1355617700001557 Published online by Cambridge University Press

interference of 112 ms. The two-way ANOVA with group (PD *vs.* control) and Stroop condition (incongruent *vs.* neutral) as the two factors revealed no significant interaction. This means that the difference in interference between the two groups was not reliable. There was a main effect of group which reflected the fact that the PD group was slower than the control group.

Suppose that one compared the PD and the control groups on the neutral RTs and the incongruent RTs. This is usually achieved by two *t*-tests. The PD and control groups were significantly different on these two comparisons. These results would suggest that the PD group was slower than the control group, but nothing more. Unless one compared the interference effects directly it is impossible to take these separate pairwise comparisons as indicative of a difference in interference.

Measuring word reading. The neutral word reading condition (RCNb) is employed without the incongruent word reading condition (RCNd). Differences between experimental and control groups in this condition (i.e., RCNb) may measure both general speed of responding and speed of color word reading. Only infrequently this condition is used to look at the reverse Stroop effect (Abramczyk et al., 1983). To study the reverse Stroop (i.e., the interference of irrelevant colors on reading words) it is necessary to add an incongruent condition (RCNd) and ask subjects to read the words and ignore the colors.

Hypothetical data that can help clarify this point are presented in Table 1 under word reading. Suppose that our study included the neutral condition (reading words in black) but not the incongruent condition. Moreover, suppose that a *t*-test between the two means (650 ms and 500 ms for the PD and the control groups, respectively) showed that the PD patients were significantly slower than the controls in reading the black words. This could mean that PD reading was deficient or that their general performance, and not only reading, was slower than the controls' general performance. This adds nothing to what we already know from the previous analysis of color naming (which showed that PD patients were generally slower than the control subjects). If one predicts a specific deficiency in word reading, one has to compare performance in this word reading condition to some other conditions that measures general performance or performance on some other variable that is not supposed to be deficient in the studied group of patients.

One way to look at deficient word processing is to study the reverse Stroop effect. To this end, one has to employ the two word reading conditions depicted in Table 1. For example, suppose that we suggest that PD patients have word reading deficiency so that their word reading is not as automatic as the control subjects' word reading. That is, their word reading is more vulnerable to interference by the irrelevant color than the control subjects' word reading. Accordingly, we would expect the interference (the difference between the neutral and the incongruent conditions) to be larger for the PD than the control groups. Similar to the color naming study we should apply a two-way ANOVA and the difference in interference (color interfering with word reading) will be reflected in a significant two-way interaction (group by condition). If the interaction was significant it is advisable to run additional analyses that compare neutral and incongruent conditions in the two groups separately. It is expected that the effect will be either nonsignificant or very small for the control group and large and significant for the PD group. This pattern of results is depicted in Table 1: the color interference for the control group is 10 ms and for the PD group 80 ms.

The incongruent conditions as a measure for interference. There are researchers who use only one condition of the Stroop task (Kartsounis et al., 1991; Rush et al., 1983; Van der Linden et al., 1993; Wolfe et al., 1990), which in most cases is the incongruent condition (NCWd). If no baseline (neutral) condition is employed, discussion of the data as measuring interference is unwarranted. The incongruent condition measures both general performance (e.g., general slowness of the patient group) and also interference (Liddle & Morris, 1991). With only one condition it is impossible to tease them apart.

Suppose that in our study (Henik et al., 1993) we used only the incongruent condition so that we asked PD and control subjects to name the color and ignore the word. Next, we compared the two mean RTs (974 ms and 782 ms) and found a significant difference. This difference reflects the fact that the PD group was slower than the control group. It should not be taken as an indication for a difference in interference because RTs in the incongruent condition do not measure interference uniquely. They reflect both interference and general speed of responding. To tease these two effects apart one has to control general speed by a comparison to a neutral condition, as suggested above.

Confounding sources of interference and task. Researchers also compare word reading with color naming. Sometimes subjects are given the stimuli which are commonly used for the incongruent condition (Corcoran & Upton, 1993; McLean et al., 1983) and asked once to read the word and ignore the color (Stroop Experiment 1, RCNd) and as a second condition they are asked to name the color and ignore the word (Stroop Experiment 2, NCWd). This comparison confounds two sources of interference, as well as task. There is an interference of color on word reading, the interference of word on color naming, and there is a difference between the two tasks, naming and reading. The difference between color naming and word reading is not necessarily a measure of attention or an ability to select the relevant dimension and ignore the irrelevant dimension. This measure might reflect a difference between reading words and naming colors; it could reflect interference of color on word, or interference of word on color, or any mixture of these three effects. It is hard to know what is measured here. In other cases researchers compared reading color words in black (RCNb) and naming neutral stimuli (NC) (Killian et al., 1984). This comparison involves the two neutral conditions for the reading and the color naming tasks so that it may reflect differences between the two tasks. However, this cannot be taken as an indication of interference, as no interference is involved in any of the two conditions.

As an example, suppose that in order to measure interference one used the difference between the two incongruent conditions in Table 1. The differences would be 244 ms for the PD patients and 272 ms for the control subjects. This measure incorporates effects of color interference (for the incongruent word reading), word interference (for the incongruent color naming), and task differences as we compare color naming and word reading. Earlier I pointed out that word interference (i.e., Stroop interference) and color interference (i.e., reverse Stroop interference) should be measured by comparing the appropriate incongruent condition to the appropriate neutral condition.

Conclusion

The Stroop task has been employed by cognitive psychologists and neuropsychologists to study attention and related functions for many years. The robustness and reliability of the effect probably made it a favorable task. However, in order to pursue issues of automaticity and the like, there are certain conditions that should be employed and certain measures which should be derived and analyzed. Deviations from these may result in studies that provide no answer to the central issue, and in measures that are uninterpretable and incomparable with the literature. This may render the data much less useful than they might be. Whether one employs the same version of the task employed by Stroop (1935) or another version (MacLeod, 1991, for a review), it is advisable to follow Stroop's (1935) logic in order to benefit from the paradigm and effect bearing his name.

ACKNOWLEDGMENTS

This work was carried out while the author was on sabbatical leave at UC Davis Department of Neurology and the Center for Neuroscience. I thank Alan Kingstone, Robert. D. Rafal, Joseph Tzelgov, and an anonymous reviewer for their helpful comments on earlier versions of this work.

REFERENCES

Abramczyk, R.R., Jordan, D.E., & Hegel, M. (1983). "Reverse" Stroop effect in the performance of schizophrenics. *Perceptual and Motor Skills*, *56*, 99–106.

Bamford, K.A., Caine, E.D., Kido, D.K., Plassche, W.M., & Shoulson, I. (1989). Clinical-pathologic correlation in Huntington's disease: A neuropsychological and computed tomography study. *Neurology*, *39*, 796–801.

Buchanan, R.W., Strauss, M.E., Kirkpatrick, B., Holstein, C., Breier, A., & Carpenter, W.T.J. (1994). Neuropsychological impairments in deficit vs nondeficit forms of schizophrenia. *Archives of General Psychiatry*, *51*, 804–811.

Corcoran, R. & Upton, D. (1993). A role for the hippocampus in card sorting. *Cortex*, *29*, 293–304.

Dikmen, S.S., Machamer, J.E., Winn, H.R., & Temkin, N.R. (1995). Neuropsychological outcome at 1-year post head injury. *Neuropsychology*, *9*, 80–90.

Everett, J., Laplante, L., & Thomas, J. (1989). The selective attention deficit in schizophrenia: Limited resources or cognitive fatigue. *Journal of Nervous and Mental Disease*, *177*, 735–738.

Golden, C.J. (1976). Identification of brain disorders by the Stroop color and word test. *Journal of Clinical Psychology*, *32*, 654–658.

Henik, A., Singh, J., Beckley, D.J., & Rafal, R.D. (1993). Disinhibition of mental word reading in Parkinson's disease. *Cortex*, *29*, 589–599.

Hietanen, M. & Teravainen, H. (1986). Cognitive performance in early Parkinson's disease. *Acta Neurologica Scandinavica*, *73*, 151–159.

Kartsounis, L.D., Poynton, A., Bridges, P.K., & Bartlett, J.R. (1991). Neuropsychological correlates of stereotactic subcaudate tractotomy. *Brain*, *114*, 2657–2673.

Killian, G.A., Holzman, P.S., Davis, J.M., & Gibbons, R. (1984). Effects of psychotropic medication on selected cognitive and perceptual measures. *Journal of Abnormal Psychology*, *1*, 58–70.

Liddle, P.F. & Morris, D.L. (1991). Schizophrenic syndromes and frontal lobe performance. *British Journal of Psychiatry*, *158*, 340-345.

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

McLean, A.J., Temkin, N.R., Dikmen, S., & Wyler, A.R. (1983). The behavioral sequelae of head injury. *Journal of Clinical Neuropsychology*, *5*, 361–376.

Rush, A.J., Weissenburger, J., Vinson, D.B., & Giles, D.E. (1983). Neuropsychological dysfunctions in unipolar nonpsychotic major depressions. *Journal of Affective Disorders*, *5*, 281–287.

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

Van der Linden, M., Bruyer, R., Roland, J., & Schils, J.P. (1993). Proactive interference in patients with amnesia resulting from anterior communicating artery aneurysm. *Journal of Clinical and Experimental Neuropsychology*, *15*, 525–536.

Wapner, S. & Krus, D.M. (1960). Effects of lysergic acid diethylamide, and differences between normals and schizophrenics on the Stroop color-word test. *Journal of Neuropsychiatry*, *2*, 76–81.

Wolfe, N., Lynn, R., Babikian, V.L., Knoefel, J.E., & Albert, M.L. (1990). Frontal systems impairment following multiple lacunar infarcts. *Archives of Neurology*, *47*, 129–132.

Wysocki, J.J. & Sweet, J.J. (1985). Identification of brain-damaged, schizophrenic, and normal medical patients using a brief neuropsychological screening battery. *International Journal of Clinical Neuropsychology*, *7*, 40–49.