Use of the Stroop phenomenon as a diagnostic tool for malingering

Alicia Osimani, Ariel Alon, Andrea Berger, Jacob M Abarbanel

Abstract
Aims—To assess a computerised version of the Stroop test for detection of malingering of cognitive deficit.
Methods—Sixty subjects were assessed using this test. Twenty had cognitive deficits due to brain damage of miscellaneous aetiologies. Ten were healthy, not acquainted with the test, and were asked to simulate cognitive impairment. Another 10 simulators were psychology students trained in the use of the test. Twenty healthy subjects served as controls. Results were analysed for reaction time, error percentage, and the Stroop effect.
Results—There was a significant difference in reaction time among groups, showing a direct relation of age among control subjects, and also longer reaction time in patients with brain damage than in controls. Controls and patients with brain damage showed a clear Stroop effect. Simulators had a significantly prolonged reaction time, increased error percentage, and inverted or absent Stroop effect. This alteration of the Stroop effect is never present in organic cognitive deficits and seems to be a characteristic pattern of feigning, independently of knowledge of the test.
Conclusion—This technique is recommended as a valuable tool to detect feigned cognitive impairment.

(J Neurol Neurosurg Psychiatry 1997;62:617–621)

Keywords: malingering; Stroop effect; head injury; reaction time

The diagnostic and statistical manual-JV defines malingering as “the intentional production of false or grossly exaggerated physical or psychological symptoms, motivated by external incentives” and notes that it should be “strongly suspected” when there is a medicolegal referral, a large discrepancy between subjective and objective findings, a lack of cooperation with evaluation or treatment, or the antecedent of antisocial personality disorder. Mild head injuries usually occur in the context of road traffic, military, or occupational accidents. Neuropsychology consultation is usually required for medicolegal purposes. The discrepancy between subjective and objective findings is the rule in mild head injury, which is perhaps the most frequent situation in which the question of malingering arises, not only because of the medicolegal referral and the lack of objective symptoms, but also due to many misconceptions and prejudices arising from studies in the early 1960s, when the diagnosis of “compensation neurosis” became popular, establishing a mechanistic criterion “no objective lesion, no brain damage”. More recent literature has challenged many of these views. Even apparently mild head injury may sometimes cause severe neurobehavioural complications. The contention that patients become less symptomatic after a legal claim is settled has also been studied by several authors and has not found support. In addition, no relation has been found between the neurobehavioural syndrome and social or psychological factors.

In summary, many of the previous views tending to ignore the postconcussive syndrome have proved false. This does not mean that there are not many people who feign cognitive inability to obtain compensation, but the real frequency of malingering is not clear, ranging from 3% to 50% depending on the studies.

Several techniques have been proposed to identify malingers. The old clinical concept can also be applied in neuropsychology: symptoms must correspond to known syndromes. Amnesic patients, for instance, never show impairment for all memory tasks, although lay people tend to think that memory is a unitary process and therefore its impairment must result in failure of every mnemonic task. Several studies on malingers are based on this premise.

Lack of consistency between the results of examination and the patient’s everyday activities is another important clue. Binder reports a case of malingering who had a VIQ of 55 and PIQ of 59 but was able to give a detailed history of his problems and drive his car safely around the city. Discrepancy among subtests of a battery may also be an important clue.

A battery to discriminate between simulated and genuine cognitive impairment would be very useful. However, the increasing media interest on memory and cognitive problems may enable intelligent simulators to fake cognitive impairment and in some cases malingerers may be advised by professionals who are well acquainted with these tests.

Some authors have devised tests which are difficult to feign even when the subject knows the technique, such as the symptom validity test and the portland digit recognition test. We report the application of the Stroop test
as a diagnostic tool for identifying malingerers. It has the advantage that it can detect simulation of cognitive impairment, even when the subject is familiar with the test.

The Stroop test is based on the finding that for normal people, reading a word takes less time than naming the object or colour corresponding to that word. Basically, the test consists of presenting the names of different colours written in coloured ink. Part of the stimuli are presented in such a way that word and ink colour correspond—that is, the word “red” is written in red (congruent condition = CC), and part of the stimuli consist of words written in a different colour—that is, the word “red” is written in blue (incongruent condition = IC). As a baseline, a third situation (neutral condition = NC) is often included, in which a meaningless string of letters in different colours is presented. Subjects are asked to name the colour and ignore the written word. In NC, the meaningless string of letters does not interfere with colour naming. In CC not only is there no conflict between the two simultaneous stimuli but there is facilitation, so that reaction time is usually shorter than in NC. In IC, reaction time is the longest because the automatic reading of the word interferes with the naming of the colour and the subject must inhibit the automatic response. The number of errors is also higher in IC. The Stroop effect is the difference between reaction time in IC and NC or CC. The Stroop effect may be measured in terms of reaction time and terms of error percentage because this is usually higher in IC than NC or CC. Previous reports have shown that the inhibition effect (longer reaction time for IC) is relatively large and stable, whereas the facilitatory effect (RT difference between CC and NC) is small and may not reach significance.

In neuropsychology this technique has been applied in particular to the assessment of frontal lobe function and closed head injury. Depressed patients have been reported to show an increased Stroop effect. Loss of the Stroop effect has never been described in patients with brain damage. The accidental finding of an inverted Stroop effect (longer reaction time in CC than in IC) in three patients suspected of malingering from the history and the results of other tests led us to the present study. The objectives were: (1) to compare the results of patients with brain damage, in particular those with head injury, with those of healthy controls and simulators; (2) to assess the efficiency of the Stroop test as a tool for identifying malingerers, even when malingerers are acquainted with the test and able to plan a feigned performance.

Material and methods

SUBJECTS

Sixty subjects participated in this study, constituting six groups. In each group there was a similar number of subjects of both sexes. All of them had vision that was normal or corrected to normal.

NORMAL CONTROLS

Group I
Ten healthy subjects, age range from 20 to 38 (mean 25) years, participated as controls for groups IV, V, and VI.

Group II
Ten healthy subjects, age range from 60 to 75 (mean 67) years, were controls for group III.

To have a correct matching for age, education, and cultural background, patients from groups III and IV were asked to bring a healthy sibling, spouse, or close relative to be tested as a control.

PATIENTS WITH BRAIN DAMAGE

Group III
Ten patients with diagnosis of mild to moderate cognitive impairment due to degenerative or vascular disease were recruited at random from a dementia clinic. Their ages ranged from 60 to 75 (mean 68) years.

Group IV
Ten patients with closed head injury were recruited from a neurosurgery clinic. Their ages ranged from 17 to 38 (mean 27) years. Time from injury ranged from six to 10 months and they were diagnosed as having mild cognitive deficits. They were aware of participating in experimental research that had no value for any litigation process.

Cognitive deficits were diagnosed by a certified behavioural neurologist, using general clinical examination and a battery of neuropsychology tests. In group IV, the main deficits were in the sphere of concentration, as measured by serial 7’s test; and memory, as measured by Rey’s auditory verbal learning list and Rey’s complex figure. In group III, the deficits were varied depending on aetiology, but all patients had scores between 100 and 120 using the Mattis dementia rating scale.

SIMULATORS

Group V
Ten normal volunteers taken at random from hospital staff and from relatives of patients in hospital participated as naïve simulators. Ages ranged from 23 to 35 (mean 25) years. They were not acquainted with any neuropsychology test.

Group VI
Ten psychology undergraduates, age range from 23 to 33 (mean 25) years participated in the experiment. They were well acquainted with the Stroop test used in this study and had worked with it in different research projects.

METHODS

Stimuli

We employed four colours: blue, green, red, and yellow. These colours are all four letter words in Hebrew. For NC we used a repeated Hebrew letter (KHAYH). There were three types of stimuli: CC (for example, RED in red ink); IC (BLUE in red ink) and NC (KHAYH in the different colours). There were four different CCs, four different NCs, and 12 different
ICs. The total number of stimuli was 216 and they were presented randomly during one session.

**Apparatus**
Stimuli were presented on a 14 inch EGA Olivetti colour monitor. Stimuli presentation and data collection were controlled by an Olivetti M-240 IBM compatible computer. Vocal responses were recorded via a Gerbrand voice operated relay (model G1341T) interfaced to the computer.

**Design and procedure**
The independent variables were: “group” (six levels) and “congruency” (three levels, according to ink-word congruency condition: CC, NC, IC). NC was not included in the analysis for simplification purposes. The dependent variables were: reaction time (measured in ms from stimulus onset to the onset of the subject’s vocal reaction) and error percentage (for each subject, in each condition).

The experiment began with a central fixation “plus” sign for 100 ms. Right after fixation the stimulus appeared for 100 ms followed by a fixed pattern of white symbols. The subject responded through the microphone whereas the fixed pattern was in the screen and the experimenter keyed the subject’s response on four keys in the computer. The experimenter’s pressure of the key determined the end of the white fixed pattern and the appearance of the new stimulus on the screen.

Subjects were run individually in a 15 minute session. A session began with a practice block of 45 trials, composed of the same conditions and proportion of trials as the experimental block.

**Instructions**
The general instructions given to all groups were:

“In the screen you will see words or strings of letters written in different colours. The words you will see are: “blue”, “green”, “red”, or “yellow”. The strings of letters will be the Hebrew letter “shin” repeated several times. As soon as the coloured word appears you must say the colour you see, ignoring the written word or letters. Try to answer as quickly as possible and refrain from making errors”

In addition to these instructions, group V received the following explanation:

“As you probably know, people with head injury have difficulty in concentration, memory, and performing different mental tasks. When you perform this test, you must pretend that you have had a head injury and the insurance company will pay you 100 000 dollars if you can fake the results and prove that you are cognitively impaired. Use your imagination on how to do it”

Group VI received the following instructions:

“As you know, many people with mild head injury show some form of cognitive impairment mainly in the sphere of information processing. You will be given a Stroop test like the one you use in the laboratory. You must pretend that you had a mild head injury. You must devise a way to prove that you have cognitive impairment, but you must keep in mind that the Stroop effect is never lost.”

This last stress was added so that they would organise strategies, as it could be the case with malingers who are trained for this test.

Once the test was performed, simulators were interviewed and asked to explain the procedure they had applied to feign the test.

For each subject in each condition median reaction time in ms and the error percentage were computed. A separate analysis of variance (ANCOVA) was performed for mean reaction time and error percentage. In addition, planned comparisons were done between the following pairs of groups: young controls (I) and simulators (V-VI); simulators (V-VI) and patients with head injury (IV); young controls (I) and older controls (II); older controls (II) and demented patients (III).

**Results**

**POSTPROCEDURAL INTERVIEW**
Simulators used different strategies to feign impairment. Some of them, mainly the untrained simulators, tried to increase the number of errors, others, mainly those of group VI, tried to increase reaction time, and some of them tried both. Some of the students tried to work as metronomes, without paying attention to the stimuli, in an attempt to increase the number of errors. Most of them stated that the original plan was to increase reaction time and error percentage, but the speed of presentation of stimuli did not permit them to manipulate the answers properly. Two subjects from group V said that they were pulled by the dynamics of the test and could not fake results.

**ANALYSIS OF RESULTS**

**Reaction time analysis**
The table shows the means (SD) for each group. The main findings were the following:

1. Age effects: young controls showed significantly shorter reaction time than older controls \(F(1,54) = 3.8, P < 0.05\).
2. Brain damage effects: comparison between young controls (I) and patients with head injury (IV) showed a significantly shorter reaction time for group I \(F(1,54) = 4.63, P < 0.03\). The difference between older controls (II) and dementia (III) had the same trend but did not reach significance.
3. Simulators’ reaction time: groups V and VI showed significantly longer reaction time than young controls and patients with head injury.
injury \((F(1,54) = 21.68, P < 0.01\) and \(F(1,54) = 4.71, P < 0.03\) respectively).

Figure 1 presents the results of the Stroop effect (IC-CC), in reaction times, for each group. The Stroop effect was significant for the six groups, as reflected by significant main effects of the congruency variable in the six cases \((F(1,54) = 4.88, P < 0.03; F(1,54) = 15.7, P < 0.01; F(1,54) = 9.65, P < 0.03; F(1,54) = 8.93, P < 0.04; F(1,54) = 4.41, P < 0.04; F(1,54) = 11.7, P < 0.01)\). The same pattern of Stroop effect was found for the first four groups, and an inverted Stroop effect was found for simulators (V–VI), as reflected in the significant interaction between group and congruency \((F(5,54) = 9.55, P < 0.01)\). The inversion of the Stroop effect is mainly due to the increased reaction time in CC.

**Error analysis**

Simulators (V–VI) showed significantly more total errors than the other groups (I, II, III, IV). \((F(1,54) = 9.74, P < 0.02)\). Group V showed the highest error percentage.

Figure 2 is a plot for error percentage for each congruency condition, for each group, showing the Stroop effect as measured by error percentage in the different groups. There was a significant interaction between simulators (V–VI) versus the other groups (I–IV) and congruency \((F(1,54) = 15.3, P < 0.01)\). Further analysis yielded a significant Stroop effect for error percentage for group I, II, and III \((F(1,53) = 4.74, P < 0.03; F(1,53) = 21.4, P < 0.01; F(1,53) = 9.97, P < 0.02\) respectively. Group IV showed a similar trend, without reaching significance. Simulators, on the contrary, showed no significant effect in the analysis of error percentage. Group V showed a trend to an inverted Stroop effect; group VI showed practically no difference between CC and IG.

**Discussion**

It may be claimed that differences exist between the performance of malingering patients and simulators. However, the same method has been applied in many previous studies suggesting that the results obtained may be generalised.\(^{10}\) In the present study, on the other hand, the participation of the students in group VI was important because one of the aims was to show that previous training or knowledge of the test did not help in feigning the results.

We shall not discuss the findings referring to the direct relation between age and reaction time seen in the control groups, or the longer reaction time shown by patients with head injury. These results have been repeatedly reported and discussed.\(^{16, 20, 22}\) However, the extremely long reaction time of simulators compared with patients with head injury is worthy of comment. After our initial three patients we have studied two more subjects with a high suspicion of malingering. Although their number is too small to include them in the study, in all cases we have found not only loss or inversion of the Stroop effect but also an exaggerately prolonged reaction time, in the range of 2000 or 3000 ms. This reaction time is much longer than the one shown by our simulators. By contrast, two patients with hysterical amnesia showed completely normal results.

The most important finding in this study is the inversion of the Stroop effect in simulators, whereas in the other groups the Stroop effect was preserved or exaggerated. Such results, such as this has never been described before.

The results of this study confirm our chance findings in patients with a high suspicion of malingering, who had very long reaction times and a markedly inverted Stroop effect. Naive as well as trained simulators showed a pattern incompatible with either normal subjects or patients with brain damage. The situation created in this experiment, by using subjects who were trained in the test, is extreme, and probably not a common occurrence. However, questions always arise suggesting that many malingers are "trained to feign". Our results show that training does not help to feign cognitive impairment in this kind of test.

Three possible patterns of performance need to be considered: (1) the subject shows a normal performance, which does not confirm or deny any diagnosis; (2) the subject shows longer reaction time, or increased error percentage, or both, and maintains the Stroop effect, which strongly suggests organic brain damage; and (3) the subject shows a very long reaction time or increased error percentage, or both, and the Stroop effect is lost or inverted, which strongly suggests malingering. In the first case, although the result does not imply any diagnosis, normal reaction time indicates...
that there is no serious brain damage, especially when dealing with closed head injury, which almost always causes a prolonged reaction time. These results must be compared with those of other tests (a very poor score in tests of attention and concentration in the presence of a normal Stroop test is suggestive of malingering).

Use of the Stroop phenomenon as a diagnostic tool for malingering.

A Osimani, A Alon, A Berger and J M Abarbanel

J Neurol Neurosurg Psychiatry 1997 62: 617-621
doi: 10.1136/jnnp.62.6.617