A MODIFIED CARD SORTING TEST SENSITIVE TO FRONTAL LOBE DEFECTS

Hazel E. Nelson

(The National Hospital for Nervous Diseases, London)

Introduction

Following Weigl's (1941) original report of impaired performance in brain-injured patients on a simple colour/form sorting task, further studies with sorting tasks produced somewhat conflicting evidence as to the localization of these effects; for example, Teuber, Battersby and Bender's (1951) results suggested greater impairment following posterior than anterior lesions, whereas McFie and Piercy's (1952) results indicated that the important factor was laterality of lesion (dominant hemisphere lesions producing greater impairments than non-dominant hemisphere lesions) rather than site of lesion within the hemisphere. In order to resolve some of these apparent contradictions Milner (1963) conducted an extensive study of 94 patients undergoing surgery for the relief of epilepsy. Using the Wisconsin Card Sorting Task (WCST) of Grant and Berg (1948) she found no significant difference between the effects of anterior and posterior lesions; indeed she was able to implicate specifically the dorsolateral aspects of the frontal lobes. Patients with dorsolateral frontal lobe lesions completed significantly fewer categories than those with orbitofrontal or more posterior lesions, the difficulties in shifting from one category to another apparently being due to the perseverative interference of responses made according to the previously correct category. The recent study by Drewe (1974), using the WCST on 91 patients with unilateral lesions, confirmed that patients with frontal lobe lesions completed fewer categories and tended to make more perseverative errors than patients with lesions elsewhere, though her results do suggest that lesions in the medial rather than the dorsolateral aspects of the frontal lobes may produce the greatest overall impairments in performance. Drewe also found a laterality effect, the left frontal group completing fewer categories and making more perseverative errors than the right frontal group, and she suggested that performance of these different groups of patients might be impaired for different reasons. More recently at the International Symposium

Cortex (1976) *12*, 313-324.

of Neuropsychology, Milner has reported the results from a long-term follow-up study which clearly indicated more lasting and more consistent deficits on the WCST after left frontal lesions than after right, these deficits after left frontal lobectomy being dissociable from language functions (Ettlinger, Teuber and Milner, 1975).

If performance on the sorting task can be impaired for different reasons, it is obviously important to know what strategies (if any) the patient is employing, to be able to determine what category concept he has in mind with each response card that he sorts. This is not always possible with the WCST. Indeed, well over half of the test cards share two or more attributes with a stimulus card, so that correct sorting of these cards does not necessarily determine which category is being applied by the subject; similarly it is not possible to determine exactly what information a 'correct' response is conveying to the patient since he could interpret this reinforcement in terms of any one, or combination, of the shared attributes. Modification of the test by the removal of test cards sharing two or more attributes with a stimulus card would remove much of this ambiguity, and in this respect might provide a more useful tool for research purposes.

Modification of the test in the way suggested would also make the test simpler for the patient. Drewe (1974) reported that six of her 91 patients refused to complete the WCST, and at the National Hospital a higher refusal rate was found when the test was being administered as a routine clinical procedure. A failure to grasp what was required in the test situation was particularly apparent in the older age groups and it is probably relevant here to note that Milner's patient population was a much younger one, with an average age in the late twenties, and even Drewe's patients were comparatively young with an average age in the mid-thirties. Having failed to grasp the test requirements, bewilderment frequently turns to distress as responses are negatively reinforced on a basis that to the patient appears totally arbitrary (if not perverse).

In order to avoid stressing patients unnecessarily, to enable more of them to attempt the test, and to provide a less ambiguous research tool, the efficiency of a simplified version of the Wisconsin Card Sorting Test was investigated. Clinical experience also suggested that an alternative method of measuring perseveration might provide better discrimination between patient groups.

Material and Method

Subjects

The experimental subjects were an unselected group of 64 in-patients at NHQS with unilateral cerebral lesions: 10 patients were excluded because their

lesions extended from the frontal lobes into adjoining areas and therefore could not be classified as either purely frontal or as purely non-frontal. Only one patient refused to complete the modified sorting test (he had a left parietal glioma, with severe dysphasia and more generally impaired intellect), so the final analysis was undertaken on the results of the remaining 53 patients. The patients were a heterogeneous group as regards aetiology, most had gliomas but a few cases of haematomas and meningiomas were also included, these additional cases being randomly distributed amongst the various subgroups.

Patients were allocated to subgroups according to site of lesion. Initially there were eight subgroups, i.e., with left frontal (LF), right frontal (RF) and left and right temporal, parietal and mixed temporo/parieto/occipital lesions, but since analysis showed that non-frontal subgroups were indistinguishable on all measures they were combined and reclassified as left non-frontal (LNF) and right non-frontal (RNF).

The control subjects were a consecutive series of 32 in-patients at NHQS with extra-cerebral lesions, most had spinal lesions but five patients with peripheral neuropathy and two patients with carpal tunnel syndrome were included (WAIS results on these patients showed a mean Full Scale IQ of 106 points, with S.D. of 11 points). To these results were added those of 8 friends/relatives accompanying out-patients attending the psychology department and 6 non-medical hospital personnel who visited the department during this period.

The numbers and mean ages of the experimental patients in each subgroup and of the controls are given in Table I. These figures show that the subgroups

TABLE I

Number and Age of Experimental Subjects According to Site of Lesion, and of Control Subjects

	Number	Age (in years)	
		Mean	S.D.
Left frontal (LF)	9	41	13
Left non-frontal (LNF)	16	42	14
Right frontal (RF)	16	45	12
Right non-frontal (RNF)	12	45	14
Total frontal (TF)	25	43	13
Total non-frontal (TNF)	28	43	14
Control (C)	46	45	14

and controls were well matched for age, so this factor could not account for other inter-group differences.

Procedure

The Modified Card Sorting Test (MCST)

The Wisconsin Card Sorting Test (described more fully in Milner, 1963) consists of four stimulus cards each of which is unique in terms of its colour (one of four, i.e., red, green, blue or yellow), shape (one of four, i.e., triangle, star, cross or circle) and number of items (one of four, i.e., one, two, three or four). The

response cards consists of 2 sets of 64 cards comprising all possible combinations of the colour, shape and number attributes. The subject is instructed to place each response card below a stimulus card, wherever he thinks it should go, and after each response he is told whether this was "right" or "wrong," no further enlightenment is given at any stage of the test. Initially the subject is required to match according to colour, but following ten consecutive correct responses this requirements is changed without warning to shape, and thence to number.

The Modified Card Sorting Test has the same set of four unique stimulus cards, but for the response cards all the cards of the WCST which shared more than one attribute with a stimulus card (e.g., the same in colour *and* the same in shape) were eliminated. This left a set of 24 cards, each of which shared *one and only one* attribute with each of three stimulus cards (e.g., the same colour as the first, the same shape as the second and the same number as the third) and shared no attribute with the fourth card (i.e., dissimilar on all three accounts). Thus if a single category concept were being used (that is colour, or shape, or number) a response card could be sorted under a particular stimulus card for that one reason only, and the right/wrong reinforcement from the examiner conveyed unambiguous information to the patient as to the correctness of the category concept being applied. This avoided the ambiguity which could arise with the WCST if for example the response card had the same colour *and* shape as a stimulus card, in which case one could not be certain which category the patient was using or how he interpreted the reinforcement of "correct." Placement under the fourth stimulus card, where no attributes were shared, indicated either an attempt to use some more complex concept (which in practice was rare) or random sorting. A pilot study showed that testing with a total pack of 2 sets of 24 cards provided essentially the same results as using a longer test.

The instructions given to the patient were: "Here we have four key cards, I want you to sort these cards (indicating the response cards) under the key cards according to certain rules: but the whole point of the test is that I shall not tell you what the rule is. I want you to find that out by trying out different rules and each time I shall tell you whether it's right or wrong. Now, go ahead and try to find out the rule."

Following these instructions, whichever category was chosen by the patient was scored "correct" and subsequent responses scored accordingly. After six correct responses in a row the patient was told "the rules have now changed, I want you to find another rule." This instruction was repeated each time the rule was changed. A pilot study had shown that explicit instruction at this point did not affect the tendency of perseverating patients to perseverate with the previous response, yet greatly reduced the distress to some patients which was caused by the sudden negative reinforcement which occurs with the Milner procedure. Whichever new category was chosen by the patient was then scored "correct" and the same procedure followed, i.e., after six consecutive correct responses the third category had to be attained; following this the three categories were repeated in the original order. The test was discontinued following the completion of the six categories, or when the 48 cards were exhausted.

Responses were recorded according to the attribute shared with the stimulus card, or as a "random" sort. Two types of assessment were then possible, the first an assessment of overall proficiency at the task, as given by the number of categories achieved and the total number of errors made, and the second an assessment of the types of errors made. The errors made were subdivided into perseverative and nonperseverative errors in two ways. Using Milner's method, an error was scored

as perseverative (PM) if it followed the category concept which had previously been correct. Clinical experience in our hospital suggested that some patients quickly forgot which category concept had been correct so that a sudden return to this concept was not necessarily indicative of "perseveration." Observation of these patients and others suggested that a more genuine measure of tendency to perseverate might be to score an error as perseverative (PN) if it followed the same category concept as the immediately preceding response, i.e. if the patient was persisting with a category despite the information that this was incorrect. Thus, so long as a patient was trying a different rule with each card he sorted then no perseverative errors were scored. The total error score (TE) was then considered in terms of the percentage of perseverative errors made, i.e. $\frac{P}{TE} \times 100\%$.

Results

Number of categories obtained

Distributions of the mean number of categories obtained by each group of subjects are given by the histograms in Figure 1. These distributions were compared using the Mann-Whitney U Test. Comparison of the individual subgroups of the experimental subjects produced no significant differences, though the comparison of Total Frontal with Total Non-Frontal patients almost reached the 5% significance level ($U = 437$; $Z = 1.59$, $p = 0.056$ on one-tailed test) in the direction of better overall performances by the Non-Frontal patients. The control subjects obtained significantly more categories than both the TF patients ($U = 908$; $Z = 4.30$, $p < 0.001$) and the TNF patients ($U = 885$, $Z = 2.94$, $p < .002$).

Total number of errors

The means and standard deviations of the total error scores are given in Table II. In keeping with the results on the number of categories obtained,

TABLE II

Means and Standard Deviations of Total Error Scores for All Subjects

	Total No. errors (TE)	
	Mean	S.D.
LF	20.9	9.4
LNF	20.9	10.8
RF	13.8	12.7
RNF	14.1	11.3
TF	20.9	11.6
TNF	13.9	11.0
C	9.2	8.5

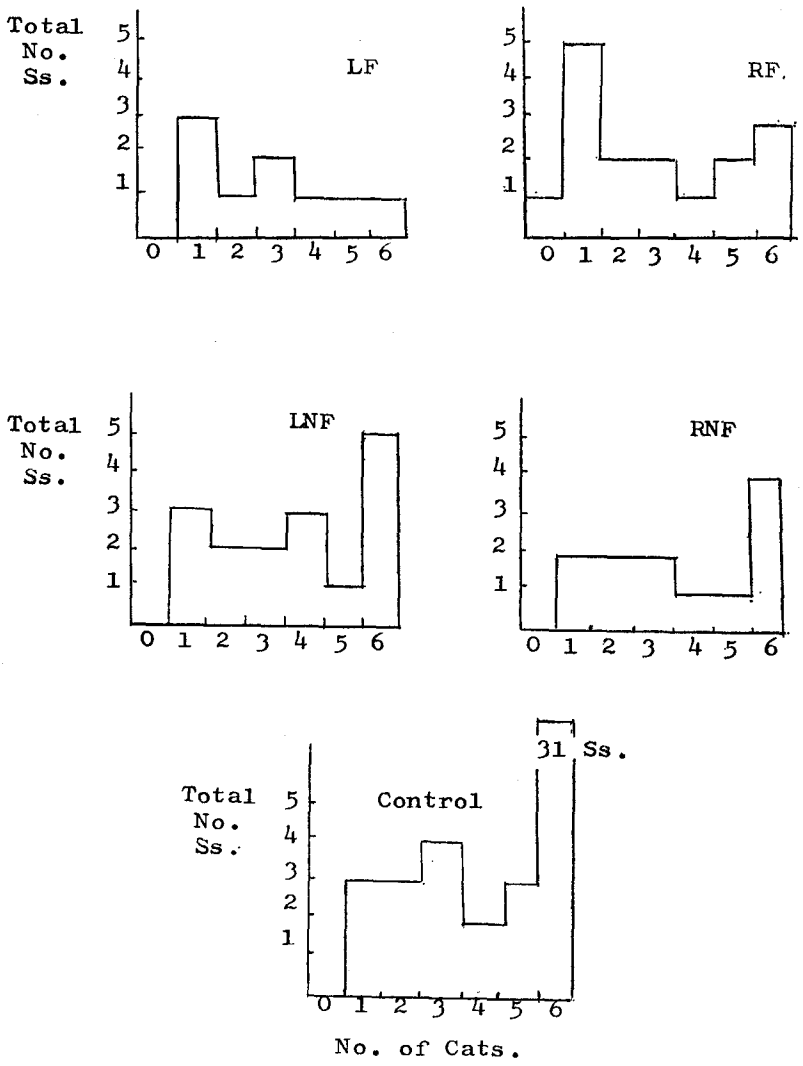

Fig. 1 — *Histograms of numbers of categories obtained by experimental and control subjects.*

the TF group had the highest mean total error score, followed by the TNF group and then the control group.

Thus the patients with frontal lobe lesions tended to perform less well generally than the patients with non-frontal lobe lesions, but the overlap between these groups was quite considerable: there were no differences at all between comparable left and right sided lesions.

Types of error made

Analysis of quality of errors is considered to be inappropriate in those cases where no difficulties are experienced and few errors made, so the results of the thirteen patients who successfully obtained the six categories within the 48 cards were not included in this part of the study. The results of the remaining forty experimental and fifteen control subjects are given in Table III.

TABLE III

Types of Error Made in Subjects Obtaining Five or Fewer Categories

				Milner classification				New classification			
		Total		Perseverative		$\frac{PM}{TE} \times 100$		Perseverative		$\frac{PN}{TE} \times 100$	
	No.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
LF	8	23.1	7.4	13.4	5.3	58	16	9.9	6.7	41	24
RF	13	24.8	10.7	12.6	10.9	47	27	11.8	9.9	42	25
TF	21	24.1	9.7	12.9	9.2	51	24	11.0	9.0	42	24
LNF	11	19.4	8.3	7.9	4.7	43	18	5.1	2.3	29	11
RNF	8	20.4	8.4	7.4	2.8	39	13	6.3	4.5	27	10
TNF	19	19.8	8.3	7.7	4.0	42	16	5.6	3.5	28	11
C	15	19.6	6.9	7.9	5.7	38	18	5.2	3.9	23	14

Inter-group comparisons of absolute numbers of perseverative errors made are essentially similar whether the Milner or New method of classification is used. Thus the total frontal group produce more perseverative errors than the total non-frontal group ($t = 2.22$, $p < 5\%$, for Milner method; and $t = 2.42$, $p < 5\%$, for New method) and more than the control group ($t = 1.81$, $p < 10\%$, for Milner method; and $t = 2.28$, $p < 5\%$, for New method), but there is no significant difference between the left and right frontal subgroups, or between the left and right non-frontal subgroups. There was no difference between the non-frontal and control groups on absolute numbers of perseverative errors made. However, since the scores of the most successful subjects were excluded from this part of the analysis, comparison of absolute numbers of perseverative errors in the remaining subjects is of very limited value.

More important than the absolute number of perseverative errors made is the proportion of errors that are perseverative. Using the Milner method of classifying perseverative errors (PM) the total frontal group make a higher proportion of perseverative errors than the total non-frontal group, but the mean difference does not reach an acceptable level of significance ($t = 1.34$, n.s.): the total frontal group also make a higher proportion of

perseverative errors than the control group, the mean difference just reaching the 10% significance level on a one-tailed test ($t = 1.72$). Using the New method of classifying perseverative errors (PN) the mean difference between the total frontal and total non-frontal groups (proportion of PN is 42% for frontals as compared to 28% for non-frontals) does reach significance level ($t = 2.10$, $p < 5\%$): the mean difference between the total frontal and control groups is also significant ($t = 2.67$, $p < 2\%$).

The non-frontal group do not differ from the control group in the proportion of perseverative errors made, whichever method of classifying these is used.

There is no difference between left and right hemisphere lesion patients in proportion of perseverative errors made.

Cut-off scores

A significant difference between groups may be obtained despite considerable overlap of scores, but if a test is to be used for diagnostic work with the individual patient the probability of misclassification must be low, so at this point the extent of inter-group overlap becomes very important. Considering the percentage of perseverative errors there was a greater overlap between the frontal and the non-frontal and control groups with the Milner method of classification than with the New method of classification.

The chances of correct identification and misclassification of patients using a range of possible cut-off scores were calculated, and those most relevant for clinical work are given in Table IV. In order to retain the test's usefulness in detecting frontal lobe lesions whilst keeping misclassification of non-frontal lesions to a minimum, the cut-off score of $\frac{PN}{TE} \times 100 = 50$ is recommended (which, in the present series of patients, correctly classified 38% of patients with frontal lobe lesions and misclassified only 5% of patients with cerebral lesions elsewhere, and none of the control subjects).

These results demonstrate the advantage of using the New method of scoring perseverative errors for diagnostic work.

Age effects

In order to study the effects of age on test performance, the Spearman rho correlations between scores and ages of patients with non-frontal lesions and of controls were calculated: scores of patients with frontal lobe lesions were not included, since any age effect in this group would be confounded by lesion effect. The results are given in Table V and show that age does have a deleterious effect on overall efficiency on the MCST, but that it has no effect on the types of error made.

TABLE IV

Comparison of Cut-Off Scores Using the Milner and New Methods of Classification

Milner method				New method			
Ia				Ib			
Cut-off score: $\frac{PM}{TE} \times 100 = 70$				Cut-off score: $\frac{PN}{TE} \times 100 = 50$			
	Site of lesion				Site of lesion		
	F	NF	C		F	NF	C
No. of subjects at or above cut-off	4 (19%)	1 (5%)	0 (0%)	No. of subjects at or above cut-off	8 (38%)	1 (5%)	0 (0%)
below cut-off	17 (81%)	18 (95%)	15 (100%)	below cut-off	13 (62%)	18 (95%)	15 (100%)
IIa				IIb			
Cut-off score: $\frac{PM}{TE} \times 100 = 50$				Cut-off score: $\frac{PN}{TE} \times 100 = 35$			
	Site of lesion				Site of lesion		
	F	NF	C		F	NF	C
No. of subjects at or above cut-off	11 (52%)	4 (21%)	5 (33%)	No. of subjects at or above cut-off	11 (52%)	3 (16%)	4 (27%)
below cut-off	10 (48%)	15 (79%)	10 (67%)	below cut-off	10 (48%)	16 (84%)	11 (73%)

Ia and Ib are those cut-off scores that keep the proportion of false positives within acceptable low limits.

IIa and IIb are those cut-off scores that keep the overall proportion of misclassifications to a minimum, but without regard to whether these are false positives or negatives.

Discussion

To summarize briefly, testing with the MCST demonstrated that patients with frontal lobe lesions tend to obtain fewer categories and make a significantly higher proportion of perseverative errors than patients with lesions elsewhere in the brain, and there is no difference between the effects of right and left frontal lobe lesions in these respects.

Consideration of the nature of card sorting tasks suggests that two

TABLE V

Age Effects on MCST Performance: Spearman rho Correlations

	Group	
	TNF	Control
Age VS. No. categories	—0.57, $p < .02$	—0.79, $p < .01$
Age VS. per cent perseverative errors ($\frac{PN}{TE} \times 100$)	0.30, p = n.s.	0.11, p = n.s.

main components are involved, first the formation of concepts on the basis of information concerning class membership and secondly the inhibition of one mode of response in favour of another when this becomes appropriate. The precise role played by the frontal areas of the brain in cognitive functioning is still far from being understood, but it has long been supposed that the frontal lobes are particularly involved in both of the above aspects of cognition, so one would predict that patients with lesions in these areas might have particular difficulty with a card sorting test. Following Milner's (1963) and Drewe's (1974) confirmation of this prediction using the Wisconsin Card Sorting Test, the present study demonstrates that the deficit is not merely due to the complex and often confusing nature of the WCST, but that it can also be elicited with greatly simplified and less ambiguous material. This finding is consistent with the occasional but quite striking example found in clinical practice of the patient with a frontal lobe lesion who can verbally demonstrate that he fully understands what he ought to be doing in the sorting test, but who nevertheless continues to perseverate with a wrong category response, often much to his own frustration.

In the present study there were no significant differences between the effects of left and right frontal lesions, which is in contrast to the significantly poorer performance of the left frontal group in Drewe's study and the more lasting deficits following left frontal lesions in Milner's follow-up study. Verbal mediation is probably involved to a much greater extent in the more complicated WCST than it is in the MCST, so although these former studies indicate that language deficits alone cannot account for the poor performances on the WCST, the results of the present study would suggest that these laterality effects on the WCST may be due to some more complicated interactive effects of the language plus other factors.

Although touching on the more theoretical issues concerning the functions of the frontal lobes, the present study was primarily concerned to develop a tool which would be useful for diagnostic clinical work, and the results clearly show that the Modified Card Sorting Test, with its advantages of being less stressful than the WCST and more appropriate in difficulty level

for our hospital population, can be used in this way. It was noted in the Introduction that it was our older patients who seemed to have particular difficulty in grasping what was required in the WCST and who, therefore, were often unable to make even an attempt at the test. There is also an age effect with the MCST, but nevertheless this did not prevent the older patients and controls from attempting and completing the test. There is no age effect on the quality of errors made, so that the greater difficulty found by older subjects is not due to an increased tendency to perseverate, as is found amongst those patients whose performance is impaired as the result of a frontal lobe lesion. Hence, although the older non-frontal patients and controls tend to resemble the frontal patients in their overall poorer performances, the effects of age and site of lesion can be distinguished if the quality of performance is taken into consideration.

The percentage of perseverative errors made by a patient may be used in the all-or-none fashion of a diagnostic "sign." The rigorous application of cut-off scores in clinical work may be criticised in so far as it involves needless loss of information concerning severity of deficit, etc., but there are situations in which a cut-off score can serve as a useful guide. The actual cut-off score chosen will depend not only on the nature of the population for which it is intended (in particular the base-rate of the "target" population) but also on the purposes for which the test is being used, since this latter affects the types and proportions of misclassifications which are acceptable. If the test is being used as a "screening" procedure then the most important consideration is to avoid "missing" one of the target population; this may involve including many of the non-target population as "possibles", but these can be eliminated by the more detailed follow-up investigations. In our clinical setting the assessment of cognitive deficits forms only one part of a much more extensive investigation, including physical examinations and special tests, and in these circumstances the possibility of falsely indicating a frontal lobe lesion where none exists is less acceptable than the possibility of failing to produce evidence of a frontal lobe lesion where one does exist. For such circumstances a cut-off score of $\frac{PN}{TE} \times 100 = 50$ is recommended for use with the MCST, since this figure is low enough to enable the detection of about one-third of the patients with frontal lobe lesions whilst keeping to a minimum the false classification of patients who do not have a frontal lobe lesion. Experienced clinicians will probably prefer to use the percentage of perseverative errors in a quantitative fashion in their assessment of the presence and strength of perseverative errors; the likely significance of these findings would then be evaluated and taken into account with other test results and information before an attempt at localisation were made. But, whichever way the test is

used, the results of the present study do indicate that the simpler modified version of the card sorting test, and the new method of scoring perseverative and non-perseverative errors, provide a more accurate tool for clinical work than the original WCST on which it was based.

Summary

Milner's (1963) report of impaired performance on the Wisconsin Card Sorting Test (WCST) in a group of patients with frontal lobe lesions suggested that this test might be a useful one in the investigation of individual patients with suspected brain lesions. However, for many of our older hospital population the WCST was found to be too difficult and distressing, and also the inherent ambiguities associated with certain responses limited the test's usefulness for research purposes. Therefore, a simpler and less ambiguous modification was devised (MCS) and a new method of measuring perseverative errors proposed. In a group of 53 patients with unilateral cerebral lesions, those with frontal lobe lesions performed less well with the MCST and made a higher proportion of perseverative errors than those with lesions elsewhere: there were no laterality effects in either frontal or non-frontal groups. The usefulness of the MCST for detecting frontal lobe lesions in individual patients was established, and the use of cut-off scores briefly discussed.

Acknowledgements. I should like to thank Dr. E.K. Warrington for all her help throughout this project, and to thank Dr. R.T.C. Pratt for his assistance in the final preparation of the manuscript. My thanks go also to the patients who took part in this investigation and to their physicians at the National Hospital.

References

Drewe, E.A. (1974) *The effect of type and area of brain lesion on Wisconsin Card Sorting Test performance*, "Neuropsychologia," *12*, 159-170.

Ettlinger, G., Teuber, H., and Milner, B. (1975) *The seventeenth International Symposium of Neuropsychology*, "Neuropsychologia," *13*, 131.

Grant, D.A., and Berg, E.A. (1948) *A behavioural analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card sorting problem*, "J. Exp. Psychol.," *38*, 404-411.

McFie, J., and Piercy, M.F. (1952) *The relation of laterality of lesion to performance on Weigl's Sorting Test*, "J. Ment. Sci.," *98*, 299-308.

Milner, B. (1963) *Effects of different brain lesions on card sorting: the role of the frontal lobes*, "Arch. Neurol.," *9*, 100-110.

Teuber, H.L., Battersby, W.S., and Bender, M.B. (1951) *Performance of complex visual tasks after cerebral lesion*, "J. Nerv. Ment. Dis.," *114*, 413-429.

Weigl, E. (1941) *On the psychology of so-called processes of abstraction*, "J. Abnorm. Soc. Psychol.," *36*, 3-33.

Hazel E. Nelson, Psychology Department, The National Hospital, Queen Square, London WC1N 3BG.