

# A study of performance on tests from the CANTAB battery sensitive to frontal lobe dysfunction in a large sample of normal volunteers: Implications for theories of executive functioning and cognitive aging

TREVOR W. ROBBINS,<sup>1</sup> MERLE JAMES,<sup>2</sup> ADRIAN M. OWEN,<sup>3</sup> BARBARA J. SAHAKIAN,<sup>3</sup>  
ANDREW D. LAWRENCE,<sup>1</sup> LYNN MCINNES,<sup>4</sup> AND PATRICK M.A. RABBITT<sup>5</sup>

<sup>1</sup>Department of Experimental Psychology, University of Cambridge, Cambridge, U.K.

<sup>2</sup>Department of Psychology, National Hospital for Neurology and Neurosurgery, London, U.K.

<sup>3</sup>Department of Psychiatry, University of Cambridge, Cambridge, U.K.

<sup>4</sup>Northeast Age Research, Department of Psychology, The University of Newcastle-upon-Tyne, Newcastle, U.K.

<sup>5</sup>Age and Cognitive Performance Research Centre, University of Manchester, Manchester, U.K.

(RECEIVED May 12, 1997; REVISED December 24, 1997; ACCEPTED January 30, 1998)

## Abstract

Several tests from the CANTAB neuropsychological test battery previously shown to be sensitive to frontal lobe dysfunction were administered to a large group of normal volunteers ( $N = 341$ ) ranging in age from 21 to 79 years. The main tests included a computerized form of the Tower of London test of planning, a self-ordered spatial working memory task, and a test of attentional set formation and shifting. A computerized form of the Corsi spatial span task was also given. Age-related graded declines in performance were seen, sometimes in a discontinuous manner, especially for the attentional set shifting task (at the extradimensional shift stage). Patterns of deficits reminiscent of frontal lobe or basal ganglia damage were observed in the oldest age group (74–79). However, overall the data were only partially consistent with the hypothesis that frontal lobe functions are the most sensitive to effects of aging. Factor analyses showed that performance in the executive tests was not simply related to a measure of fluid intelligence, and their performance had a factor loading structure distinct from that for the CANTAB tests of visual memory and learning previously administered to the same sample. Finally, only limited support was found for the hypothesis that cognitive aging depends on slowed information processing. (*JINS*, 1998, 4, 474–490.)

**Keywords:** Executive function, Prefrontal cortex, Aging, Set-shifting, Working memory, Planning

## INTRODUCTION

There is currently much interest in the definition and measurement of executive functioning both in normal individuals and in patients with brain damage, particularly those in whom there is frontal lobe or closed head injury (Burgess, 1997; Tranel et al., 1994). The term “executive functioning” refers to those processes by which an individual optimizes his performance in multicomponent tasks. These different processes include the ability to respond flexibly

and appropriately in altered circumstances, efficient scheduling of behavior and attentional resources, as well as the suppression of inappropriate responding, the use of strategies to enhance mnemonic function, and the formulation of new plans of action.

A number of neuropsychological tests have been developed to measure frontal lobe dysfunction, such as the Wisconsin Card Sorting Test (Milner, 1963), the FAS and category tests of verbal fluency (Benton & Hamsher, 1976), the Stroop test (Vendrell et al., 1995), and the test of cognitive estimates (Shallice & Evans, 1978). These tap various aspects of executive functioning, although it is important to realize that executive and frontal lobe functions are not synonymous. A major difficulty, however, is the heterogeneity of frontal lobe impairments and executive forms

Reprint requests to: Trevor W. Robbins, Department of Experimental Psychology, University of Cambridge, Downing St., Cambridge, CB2 3EB, U.K. E-mail: twr2@cus.cam.ac.uk

of dysfunction. Consequently, novel means of testing and measuring such functions are constantly being devised and evaluated.

The CANTAB (Cambridge Neuropsychological Test Automated Battery) was originally devised for the assessment of cognitive function in elderly and dementing subjects (Robbins et al., 1994). It is a battery of computerized tests administered with the aid of a touch-sensitive screen. The main guiding principles have been to use some tests that can be related to the extensive neuropsychological literature in animals and to employ tests that can be broken down into their cognitive components in order to define more readily which functions are impaired and which are spared. The battery has now been used quite extensively in the testing of patients with Alzheimer's disease and other forms of dementia (Sahakian et al., 1988, 1990; Sahgal et al., 1991, 1992), patients with basal ganglia disorders such as Parkinson's disease (Downes et al., 1989; Owen et al., 1992, 1993), Korsakoff's syndrome (Joyce & Robbins, 1991), depression (Abas et al., 1990; Beats et al., 1996), schizophrenia (Elliott et al., 1995), HIV-positive patients (Sahakian et al., 1995), and children with minimal learning disabilities or autism (Hughes et al., 1994). The tests have also been validated in neurosurgical patients with excisions of the temporal or frontal lobes and amygdalohippocampotomy (Owen et al., 1995). Results from these studies have confirmed that some of the tests are sensitive to temporal lobe damage and others to frontal lobe damage.

A detailed study of some of the tests from the visual memory battery of CANTAB was made on a large population of elderly volunteers, for the purposes of standardization (Robbins et al., 1994). This study showed that, whereas performance on some of the tests showed a gradual and continuous decline with age, performance on other tests showed significant decline within particular age ranges. A factor analysis of data from almost 800 participants provided a solution with four main factors, which were tentatively associated with visual memory and learning, speed of processing, executive functioning, and perception. However, relatively few of the tests employed are sensitive to frontal lobe dysfunction or have obvious executive components. For example, the CANTAB test of self-ordered, spatial working memory has been shown to be sensitive to frontal lobe damage in human patients, but this impairment depends in part on the formulation of an effective strategy for carrying out this self-ordered memory task (Owen et al., 1990, 1996c). The study did not include an assessment of strategy in this large elderly sample. Furthermore, two other prominent tests sensitive to frontal lobe damage were not included in this analysis: a computerized form of the Tower of London test of planning (Owen et al., 1990; Shallice, 1982), which allows measurement of both the speed and latency to solve the test problems, and a test of attentional set formation and shifting, modeled after the Wisconsin Card Sorting Test but simplified and presented in a graded format suitable for the testing of experimental animals (Dias et al., 1996; Roberts et al., 1988).

A main purpose of the present study is to report data from a large population of normal, healthy older people on the three tests sensitive to frontal lobe dysfunction: self-ordered, spatial working memory; the Tower of London; and attentional set formation and set shifting. We were able to test a large sample of individuals between the ages of 55 and 79 years, together with a sample of participants younger than 55 years of age. The majority of the older sample had also received testing with the CANTAB Visual Memory battery, which includes tests of recognition memory and visuospatial learning (Robbins et al., 1994). Thus, it was possible to perform a factor analysis which examined the factor structure of the entire CANTAB battery as well as to establish the factor structure of the tests with executive components. Both of these analyses are important. First it is necessary to test the hypothesis that certain functions, probably mediated by specialized regions of the posterior neocortex, are correlated to only a minor extent with performance on tests with large executive components. Second, it is important to define possible interrelationships among the various aspects of executive functioning, potentially in order to relate different aspects of executive functioning to different regions of the prefrontal cortex (Robbins, 1996; West, 1996). There is already some evidence from psychometric and functional neuroimaging perspectives that some of the CANTAB tests of executive function are closely interrelated (Owen et al., 1996a, 1996b; Robbins, 1996).

Ancillary aims of the study were to examine the relationship between test performance, age, and intelligence. The data could be analyzed within the context of one theory of cognitive aging—that speed of information-processing underlies all age-related declines in cognitive function (Salt-house, 1985, 1996)—by using one set of variables from the CANTAB battery to provide indices of speed of processing (reaction times) (Nettelbeck & Rabbitt, 1992).

As it has recently been suggested that tests of executive function require high levels of fluid intelligence (Duncan et al., 1995), we also used a factor analytic method to examine the relationship between performance on this battery and scores on the AH 42 (Heim, 1968), a set of tests of visuospatial reasoning that requires fluid intelligence. This analysis enabled us to determine to what extent the sensitivity of some of these tests to frontal lobe dysfunction could be predicted from their loading for *g* (or general intelligence).

## METHOD

### Research participants

Participants were drawn from a sample of 341 normal healthy volunteers, aged 21 to 79 years. Those over 50 were drawn from the North East Age Research panel based in Newcastle-upon-Tyne. Participants younger than 55 were drawn from a variety of sources in the Cambridgeshire area. Details of their numbers by age, sex, and verbal IQ (as estimated by the NART; Nelson, 1982) are shown in Table 1 in terms of

**Table 1.** Characteristics of population by age

Age group	<i>N</i>	Age ( <i>M, SD</i> )	NART score ( <i>M, SD</i> )
Under 55	77	34.2 (10.0)	114.9 (7.6)
Male	34	37.4 (10.8)	112.9 (8.9)
Female	43	31.7 (8.7)	116.5 (6.0)
55–59	49	57.8 (1.1)	118.0 (6.9)
Male	9	57.0 (1.0)	118.1 (4.6)
Female	40	57.9 (1.1)	117.9 (7.4)
60–64	36	62.1 (1.5)	119.6 (6.4)
Male	9	62.3 (1.3)	120.1 (4.9)
Female	27	62.0 (1.5)	119.4 (6.9)
65–69	61	67.3 (1.4)	118.3 (7.2)
Male	26	67.4 (1.4)	120.3 (5.4)
Female	35	67.3 (1.4)	116.8 (8.0)
70–74	83	71.9 (1.4)	119.5 (6.4)
Male	40	72.4 (1.4)	119.5 (6.2)
Female	43	71.5 (1.4)	119.4 (6.7)
75–79	35	76.4 (1.4)	116.1 (6.5)
Male	15	76.5 (1.5)	116.4 (6.6)
Female	20	76.3 (1.3)	116.0 (6.6)

5-year age bands or “quincades.” It was not possible to obtain data on all tests or variables from all participants; therefore, the precise numbers of participants used for each dependent variable are indicated throughout the text. The full sample was used unless otherwise indicated. For the participants older than 50, it was possible to obtain scores on a test of visuospatial fluid intelligence, AH 4 2 (Heim, 1968; see Nettelbeck & Rabbitt, 1992), as well as the National Adult Reading Test, which can be used to estimate verbal IQ (Nelson, 1982).

### Equipment and Procedure

The main tests were taken from the Cambridge Neuropsychological Test Automated Battery (CANTAB). This series of computerized tasks was run on an Acorn BBC Master 128 microcomputer with a high resolution Microvitec (Bradford, U.K.) 12-inch VDU and a Microvitec Touchtec 501 touch sensitive screen. Participants sat at a comfortable height approximately 0.5 m from the monitor. It was explained that they would have to respond to stimuli by touching the screen. They were introduced to the apparatus by way of a motor screening task in which they were asked to respond to a series of flashing crosses on the screen by placing the index finger of their preferred hand on the center point of each cross as soon as possible after it appeared. All volunteers passed this introductory screening task and were then given the following test batteries over two visits within a few weeks of one another. On the first visit they completed the tests of spatial span, spatial working memory, and the Tower of London from the CANTAB Planning and Working Memory battery and the attentional set shifting paradigm from the CANTAB Attentional battery. On the second

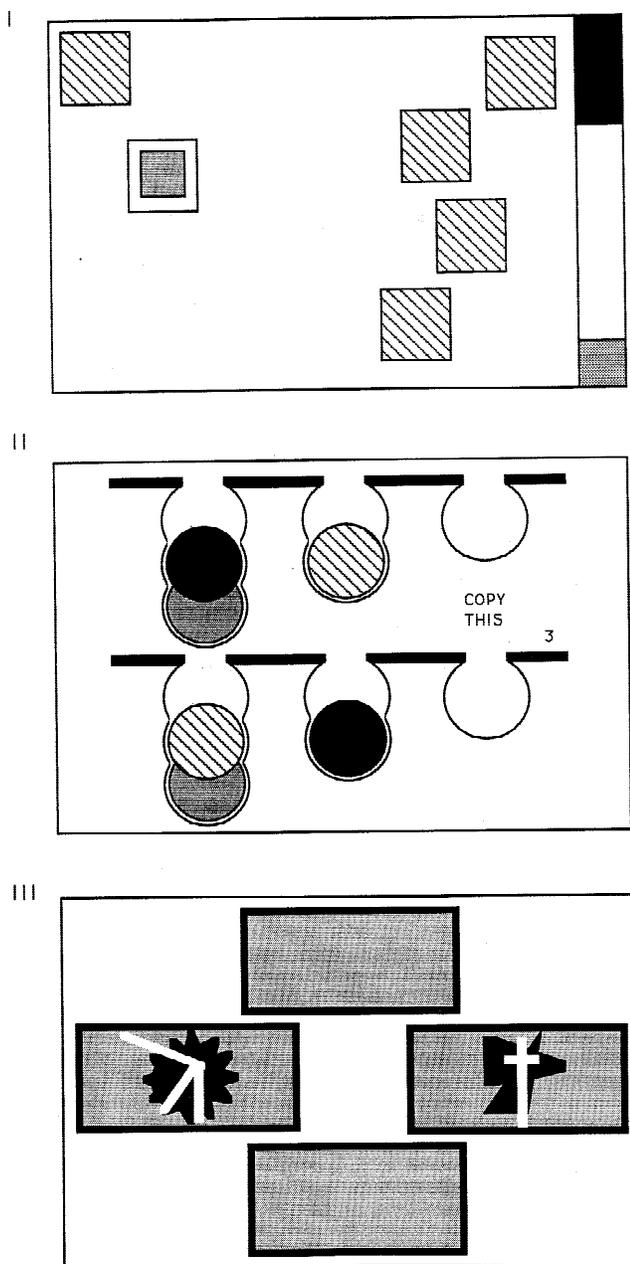
visit they did the CANTAB Visual Memory battery as described previously (Robbins et al., 1994), including tests of pattern recognition memory, spatial recognition memory, simultaneous and delayed matching to sample and a test of visuospatial paired associate learning. They also completed a matching to sample, visual search test from the CANTAB Attentional battery, also previously described (Robbins et al., 1994). Data from the tests completed on the second visit have already been published using this sample (Robbins et al., 1994), but were also used here to compare with tests sensitive to frontal lobe damage.

### *Spatial short-term span*

This computerized version of the Corsi Block Tapping task (Milner, 1971) was used to determine the ability of participants to remember a sequence of squares highlighted on the screen. Precise details are provided in Owen et al. (1990). After each successful trial, the number of boxes changing in the next sequence was increased by one to a maximum of nine boxes. After an incorrect attempt at any particular level, an alternative sequence of the same length was presented. This continued until the participant had failed three consecutive trials at any one level. The spatial short term memory span was calculated as the highest level at which the participant had successfully recalled at least one sequence of boxes.

### *Spatial working memory task*

The rationale for this task and its implementation has been described previously in some detail (Owen et al., 1990). The participant was required to search through a number of colored 3-cm boxes presented on the screen by touching each one in order to “open it” and thus revealing its contents (see Figure 1–D). The goal was to collect blue tokens hidden inside the boxes and, once found, to use them to fill an empty column at the side of the screen. The participants were instructed that at any one time there would be a single token hidden inside one of the boxes. Their task was to search until they found it, at which point the next token would be hidden. The key instruction was that, once a blue token had been found within a particular box, that box would not be used again to hide a token for that particular trial. Since every box was used once, the total number of blue tokens to be found on each trial corresponded to the number of boxes on the screen. In order to analyze performance, it is convenient to consider each sequence of responses that terminates in finding a token as a “search.” Two main types of search error are possible. First, a participant may return to open a box in which a blue counter has already been found in a previous search (a between-search error). Second, a participant may return to a box already opened in the same search sequence (a within-search error). It is of course possible for an error to be counted under both categories, and such examples were counted under both categories (they generally form a small proportion of the total errors).



**Fig. 1.** Three tests from the CANTAB battery: (I) Spatial working memory; (II) Tower of London; (III) Attentional set shifting.

Participants could search the boxes in any order, but for control purposes the number of boxes visited (excluding errors) before a token was found was determined by the computer. Thus, each participant received the same degree of feedback prior to the first error. After two practice trials with three boxes, there were four test trials, each with four, six, and finally eight boxes. The task was scored according to the total number of between- and within-search errors.

A previous investigation has shown that impaired performance on this task in neurosurgical patients with frontal lobe excisions may be related to the inefficient use of a particu-

lar search strategy associated with superior performance in normals (Owen et al., 1990). The extent to which each group used this repetitive searching pattern as a strategy for approaching the problem was estimated from the number of search sequences starting with the same box within each of the more difficult six and eight box problems. The total of these scores provided a single measure of strategy for each participant, with a high score (many sequences beginning with a different box) representing low use of the strategy and a low score (many sequences starting with the same box) representing more extensive usage.

### Planning task

This task is closely related to one developed by Shallice and McCarthy, the Tower of London, which provides a simpler version of the Tower of Hanoi problem (Shallice, 1982). The computerized version of the task is described in detail by Owen et al. (1990). Two sets of three colored balls were presented, one in the top half of the screen and one in the bottom half. These were described to the subject as snooker (or pool) balls since they appear to be hanging in pockets (or socks) (see Figure 1–II). On each trial a red ball, a blue ball, and a green ball were placed in predetermined positions in the pockets of each of the two displays. The participant was asked to rearrange the balls in the bottom display such that their positions matched the goal arrangement in the top half of the screen. A ball could be moved by first touching it and then touching an empty position in one of the other pockets. Once selected, a tone sounded, and the rim of the ball began to flash, indicating that it was ready to be moved. At any time, the participant could cancel a selected ball by touching it a second time. Illegal moves, such as trying to remove a ball while there was another sitting above it in the same pocket, were carefully explained to the participant, and if attempted they were registered but evoked no response from the computer.

The starting position of the balls was varied such that in any particular trial the solution could only be reached after a minimum of two, three, four, or five moves. Participants were instructed to examine the position of the balls at the beginning of each problem and to solve it in the minimum possible number of moves. This instruction was given to them verbally and displayed on the screen throughout each trial. They were encouraged not to make the first move until they were confident that they could execute the entire sequence needed to solve the problem. The maximum moves allowed corresponded to twice the minimum number possible plus one or, in the case of five-move problems, plus two. The program stored the number of moves required by the participant to rearrange the balls and measured the selection and movement latencies for both the first and the subsequent moves. After six practice problems with one and two moves, the participant was given two each of two- and three-move problems and four each of four- and five-move problems. These test problems corresponded exactly to those used in the original Tower of London test.

For each test problem a yoked control condition was employed to provide baseline measures of the times taken to initiate and execute single movements defined as motor initiation and execution times. On each trial of this control condition, the participant was required to follow a sequence of single moves executed by the computer in the top half of the screen by moving the corresponding ball in the lower arrangement. Thus, initially the two arrangements differed by just one ball. Once the subject had made the appropriate move, the top arrangement changed again so that the participant had to make another single move. The test was yoked to the main test in the sense that in each trial the movement of the balls was an exact replication of those moved by the participant in the corresponding test trial. The measurement of selection and execution latencies in this control condition provided baseline estimates of motor initiation and execution times.

Test trials and yoked control trials were arranged in four blocks of six problems each. The first six test trials were given (two problems at each of two, three, and four moves), followed by their corresponding yoked control trials. Then the remaining six test problems were presented (two at four moves and four at five moves), followed by their yoked control trials. Between each block change there were two practice trials to ensure that the requirements of the current set of tasks had been fully understood.

*Accuracy of performance.* Across the 12 test problems, the main measure was the proportion of problems solved in the minimum number of moves specified (“perfect solutions”).

*Latency measures of performance.* Baseline measures of motor initiation and motor execution times were extracted from the 12 yoked control trials. In all cases, latencies were recorded in hundredths of seconds and converted to seconds for the purpose of presentation. The motor initiation time was the mean time between the onset of each problem and the completion of the first move (i.e., a correct touch of the required ball). The motor execution time was the time between touching the first ball and completing the sequence of single moves that comprise the whole problem. Since these control problems were yoked to the test problems, the total execution time was divided by the number of moves to provide an estimate of the average movement time per move.

The motor initiation and execution times were used to derive estimates of planning or “thinking” time in the main task. Two separate estimates were calculated. In each problem, the initial thinking time was the time between the presentation of the problem and the first touch minus the corresponding latency to make the same response on the yoked control task (i.e., subtracting motor initiation time). The subsequent thinking time was the time between the selection of the first ball and the completion of the problem minus the total motor execution time summed over all moves when made separately, from the corresponding control problem. Since this measure clearly varied with problem length,

subsequent thinking time scores were divided by the number of moves to give an estimate of the average thinking time per move. In this way, pure estimates of initial and subsequent thinking times were derived, unconfounded by motor initiation or execution times.

### *Attentional set formation and shifting*

This includes a series of visual discrimination tasks (see Downes et al., 1989). The test set-up and stimuli were shown to the participants. Four rectangular boxes—to the top and bottom and to the right and left of center—appeared on the screen. Two of these contained the test stimuli, but the boxes used changed from trial to trial. Participants were instructed in the following way:

*Now you can see two patterns. One of the patterns is correct. You must point to the one you think is correct. There is a rule you can follow to make sure you make the correct choice each time. The computer will be keeping track of how well you are doing and when it is clear that you know the rule the computer will change it, but this will not happen very often. To begin with, there is nothing on the screen to tell you which of the two patterns is correct, so your first choice will be a simple guess. However, the computer will give a message after each attempt to tell you whether you are right or wrong. You can start now.*

The test then proceeded through a number of stages, each with a different contingency, up to a maximum of nine (see Figure 1–III). For each, continuation to the next stage was dependent on a criterion of six successive correct discriminations being reached. If criterion was not reached at the fiftieth trial of a stage, the test was discontinued, and the participant did not proceed to the following stage. The order of discrimination was fixed so that the EDS always followed the IDS. However, previous work has established that comparable effects are found when the alternative ordering is used (Roberts et al., 1988). More detailed rationales for the exact design of the test can be found in previously published articles (Downes et al., 1989; Roberts et al., 1988).

To begin with, participants were given a simple simultaneous discrimination in which the stimuli varied along only one of the two dimensions for deriving the stimuli. These dimensions were purple-filled shapes or white lines. The starting dimension was balanced across participants and groups. Feedback for responses was in the form of the words correct and wrong, presented respectively in green and red lettering above the middle two boxes.

Following the initial simple discrimination (SD), the remaining eight stages were as follows. For the second stage (SDR), the discriminanda remained the same, but the previously incorrect choice became the correct one and vice versa (i.e., the contingencies were reversed). At the third stage (C-D), the second dimension was introduced with one exemplar of each dimension paired together to form a compound stimulus in two of the response boxes. To succeed, a participant had to continue to respond to the correct exemplar of the previous stage. For this and all subsequent stages, exemplars of the different dimensions were paired in a

pseudorandom fashion so that all four possible compound stimuli were used, with the constraint that runs of no more than three trials with the same pairings were allowed. The stimuli for the fourth stage (CD) and subsequent stages were also compounds, but the two exemplars from the different dimensions were superimposed, with the white line always in the foreground. The contingencies were again unchanged from those for the previous two stages. A reversal then occurred at the fifth stage (CDR). New exemplars for both dimensions were introduced at the sixth stage, the intradimensional shift (IDS), but the relevant dimension (i.e., shapes or lines) was unchanged from Stage 1. This was followed by a further reversal at the seventh stage (IDR). For the penultimate stage, the extradimensional shift (EDS), new exemplars were again introduced, but success at this point depended on the participant shifting response set to the exemplars of the previously irrelevant dimension. Finally, contingencies were reversed to the previously incorrect exemplar of the new dimension (EDR). One of the measures of performance on this task was the stage successfully attained. Another was the number of errors made summed over the stages up to but not including the EDS stage (Total errors – EDS). This provides a measure of set formation and maintenance. A specific measure of set shifting is provided by the errors made at the EDS stage. For participants failing the test at earlier stages, 25 errors was substituted for their score.

### Statistical analysis

Most of the dependent variables described above were subjected to one-way analysis of variance (ANOVA), with age group as a factor, using the SPSS package (Norusis, 1990). Following the finding of a significant effect of age group, pair-wise comparisons between means were made using the Newman–Keuls test. In the case of repeated measures designs (e.g., the Tower of London test), multivariate analysis of variance (MANOVA; SPSS) was used, and the two factors were age group and difficulty. For the attentional set shifting task, the numbers of participants succeeding and failing to reach criterion at each of the nine stages of the test were cast into contingency tables and analyzed using a likelihood ratio analysis (Kullback, 1968; Robbins, 1977) which allows the use of small cell frequencies, the resulting statistic,  $2i$ , being distributed as chi squared. In order to compute the relationship between the strategy score on the spatial working memory task and the main index of memory performance, the between-search errors score, Pearson's correlation coefficient,  $r$ , was used.

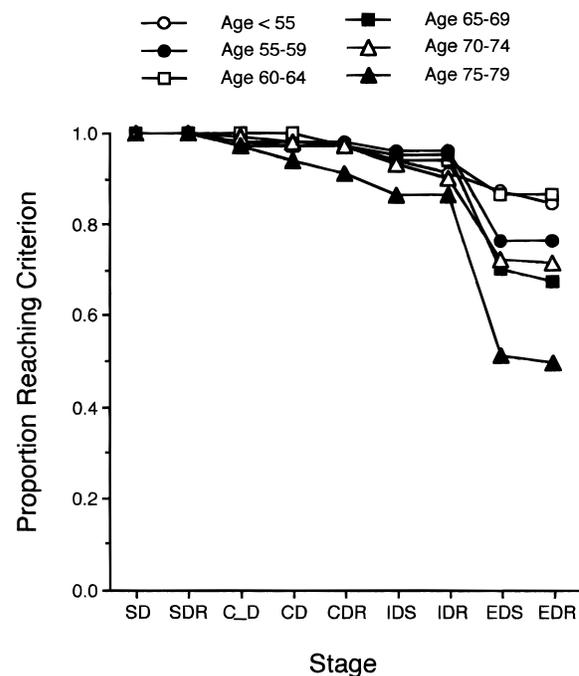
The data for the main variables reported in this article were subjected to factor analysis. The method employed was a varimax rotation with Kaiser normalization (SPSS). As complete data sets were not obtained for all participants, the method of pair-wise comparisons was employed (Norusis, 1990, p. B46), which uses all of the available data in the analysis. However, separate analyses using the sample with the complete data set ( $N = 101$ ) gave qualitatively similar

results, providing evidence for the consistency of the factor structure obtained. An additional factor analysis was used in order to relate the data from the tests used in this study to data obtained using other tests of memory and perception (see above for descriptions) previously reported for the same population (Robbins et al., 1994). Finally, we also constructed an intercorrelation matrix for participants with complete data on a slightly smaller subset of cognitive tests ( $N = 215$ ) for Pearson's  $r$  and also computed partial correlations where appropriate (see Results).

## RESULTS

### Attentional Set Shifting

Figure 2 shows the proportion of participants from each of the five age groups reaching criterion at each stage of the attentional set shifting paradigm. Likelihood ratio analysis of the numbers of participants passing or failing all stages of the test indicated a significant age group difference [ $\chi^2(5) = 20.25, p < .01$ ], which *post-hoc* contrasts showed to be wholly attributable to the differences between the youngest and oldest age groups and the rest [ $\chi^2(1) = 10.20, p < .01$ ; and  $\chi^2(1) = 11.60, p < .01$ , respectively]. Further anal-



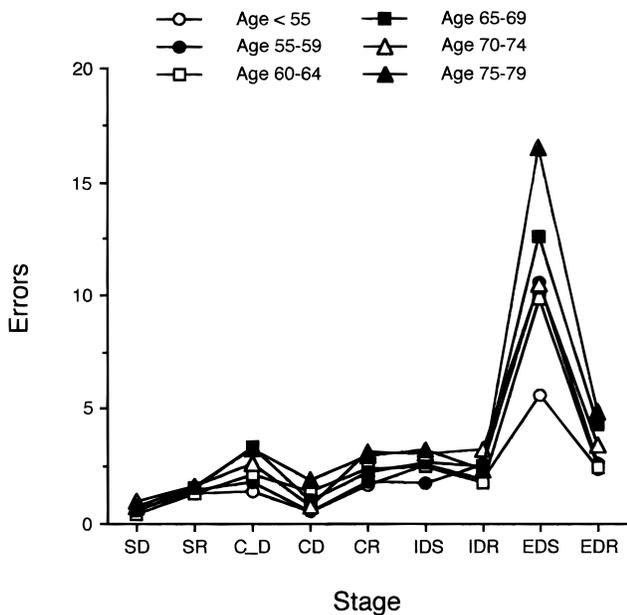
**Fig. 2.** Proportion of participants passing the criterion (6/6 correct) for each stage of the attentional set shifting paradigm (full sample, see Table 1). Abbreviations: SD = simple discrimination; SR = simple reversal; C-D = compound discrimination, with separated elements; CD = compound discrimination, superimposed elements (see Figure 1C); CDR = compound discrimination reversal; IDS = intradimensional shift; IDR = intradimensional reversal; EDS = extradimensional shift; EDR = extradimensional shift reversal.

ysis was undertaken to determine which stage of the test most distinguished these groups from the rest. For this purpose, we considered the performance of participants up to and including the IDS stage and those participants only attempting the EDS stage. This showed no significant differences in the former case [ $\chi^2(4) = 4.34, p > .05$ ], but the pattern of results was similar at the EDS stage to that for the test as a whole; that is, both the youngest [ $\chi^2(1) = 13.24, p < .001$ ] and the oldest [ $\chi^2(1) = 9.00, p < .01$ ] differed significantly from the rest, combined.

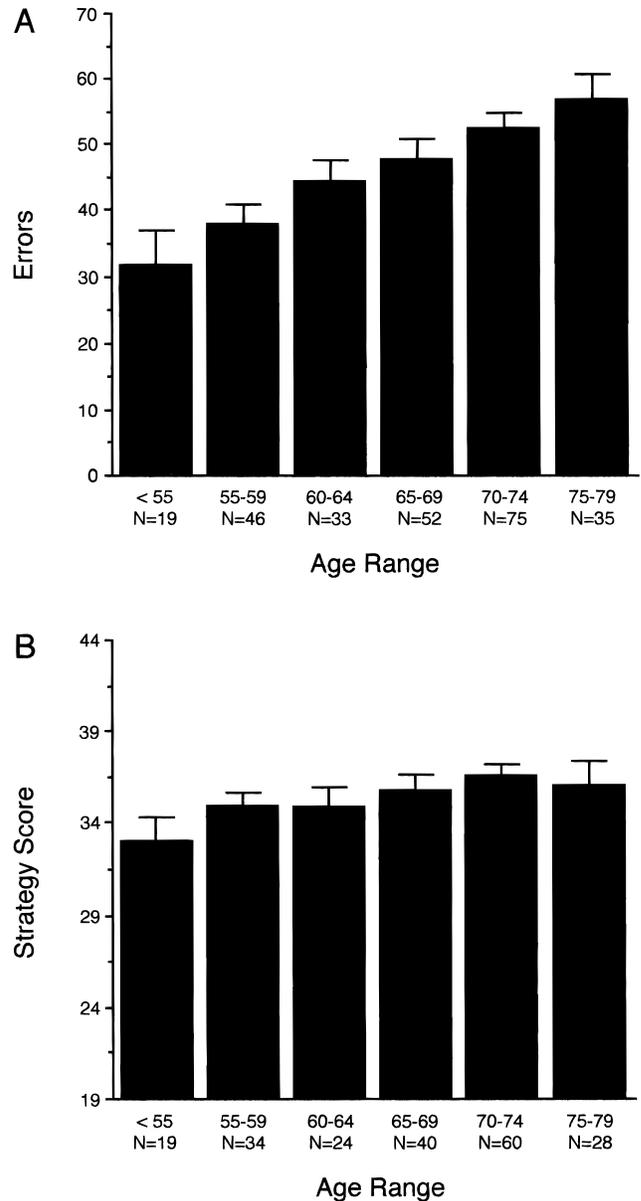
Figure 3 shows data for the total errors measure summed across stages of the test for the five age groups. It is obvious that the major differences between the groups are present at the EDS stage. One-way ANOVA of the data at that stage showed a significant effect of age group [ $F(5, 335) = 5.33, p < .0001$ ], and *post-hoc* comparisons revealed the same pattern of significant effects as found for likelihood analysis at the EDS stage: that is, the youngest group made significantly fewer errors, and the oldest group made significantly more errors than each of the other groups.

**Spatial Working Memory**

Figure 4a and 4b shows the effects of age group on the between-search errors and strategy measures. The strategy measure was only available for a limited subset of participants, whose data are shown in Figure 4. For between-search errors there was a significant age group effect [ $F(5, 254) = 6.64, p < .0001$ ]. Further analysis showed that the 75 to 79 year old group committed significantly more errors than the three youngest groups, the 70 to 74 year old



**Fig. 3.** Attentional set shifting paradigm: Mean number of errors made at each stage. Participants not attempting a stage were allotted 25 errors for that stage. See Figure 2 for abbreviations.



**Fig. 4.** Spatial working memory: (a) Mean number of between search errors as a function of age. *N* = number of participants tested for each age group. (b) Mean strategy score for each age; lower scores indicate greater use of strategy.

group made significantly more errors than the two youngest groups, and the 65 to 69 year old group made significantly more errors than the youngest group. No other significant differences were present.

The significant differences for the between-search errors measure contrasts with the lack of effect of age group on the within-search errors measure [ $F(5, 254) = 0.79, p = .56$ ]. Mean values were 5.95, 8.33, 7.27, 10.25, 7.97, 10.29 across the six quincades (youngest first).

For the strategy score there was no significant age group effect [ $F(5, 199) = 1.78, p = .119$ ]. However, there was a significant age-related difference when the relationship be-

tween the between-search error score and the strategy score was computed using Pearson's correlation coefficient  $r$ . Figure 5 shows that there were highly significant correlations between these two measures (confirming previous findings for a much smaller sample; Owen et al., 1990) for all age groups, except the eldest (75–79), even though strategy scores were not available for all participants and were quite low in certain cases. The failure of the eldest age group to show a significant relationship is unlikely to have been due to the small sample size ( $N = 28$ ), as the youngest group showed a significant value for Pearson's  $r$  with an even smaller  $N (= 19)$ .

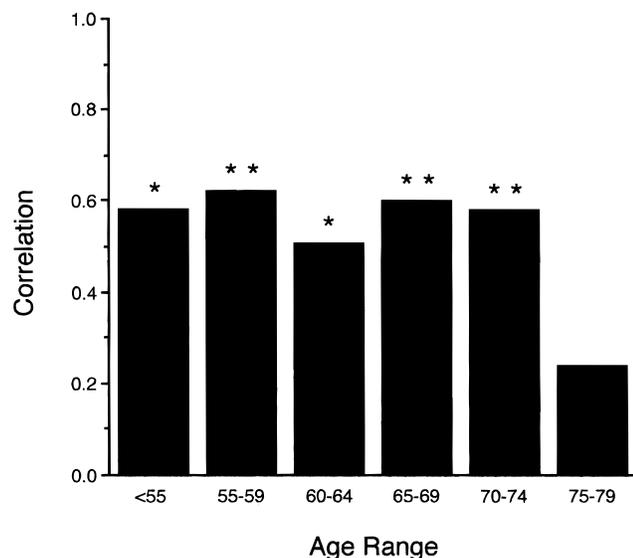
### Spatial Span

Figure 6a shows that spatial span held relatively constant across quincades at about the level of 5. Nevertheless there was a significant decline over age groups [ $F(5, 210) = 3.23, p = .008$ ]. However, the Newman-Keuls procedure failed to find significant differences among subgroups.

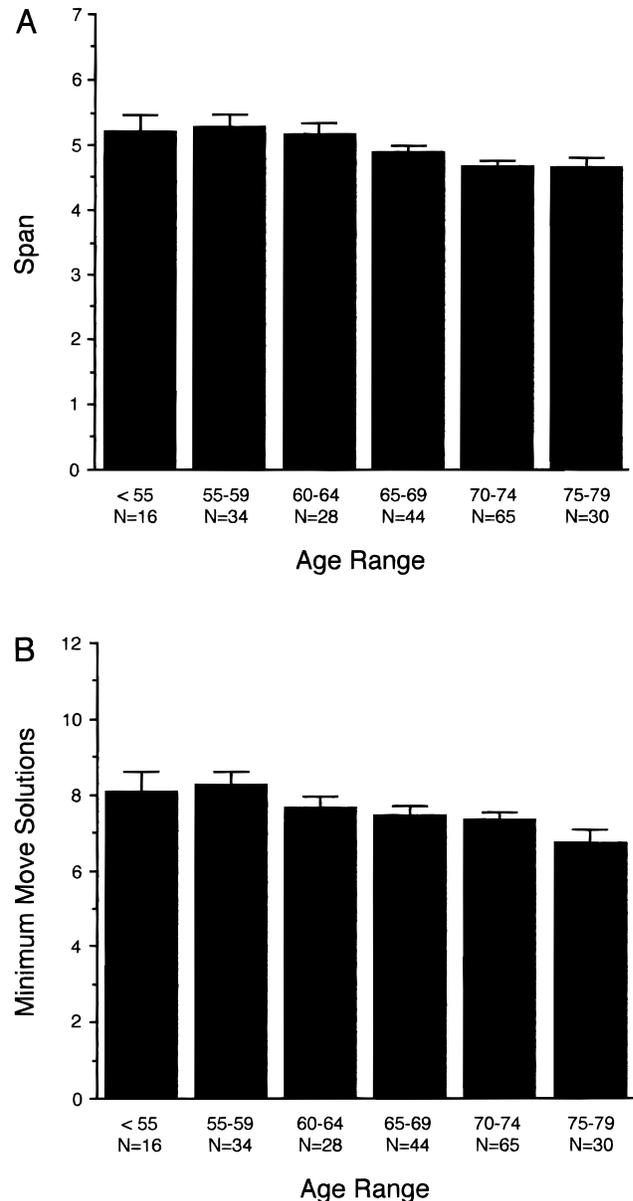
### Tower of London Planning Task

Figure 6b shows that the number of problems solved in the minimum number of moves did not decline greatly in overall terms (from 8.1/12 [youngest group] to 6.8/12), but declined significantly across age group [ $F(5, 209) = 2.78, p = .019$ ]. Further analysis showed that the two youngest groups solved significantly more problems than did the oldest group.

For the latency measures, Figure 7a shows that initial thinking time varied both by age group and by problem difficulty. The ANOVA revealed a significant main effect of



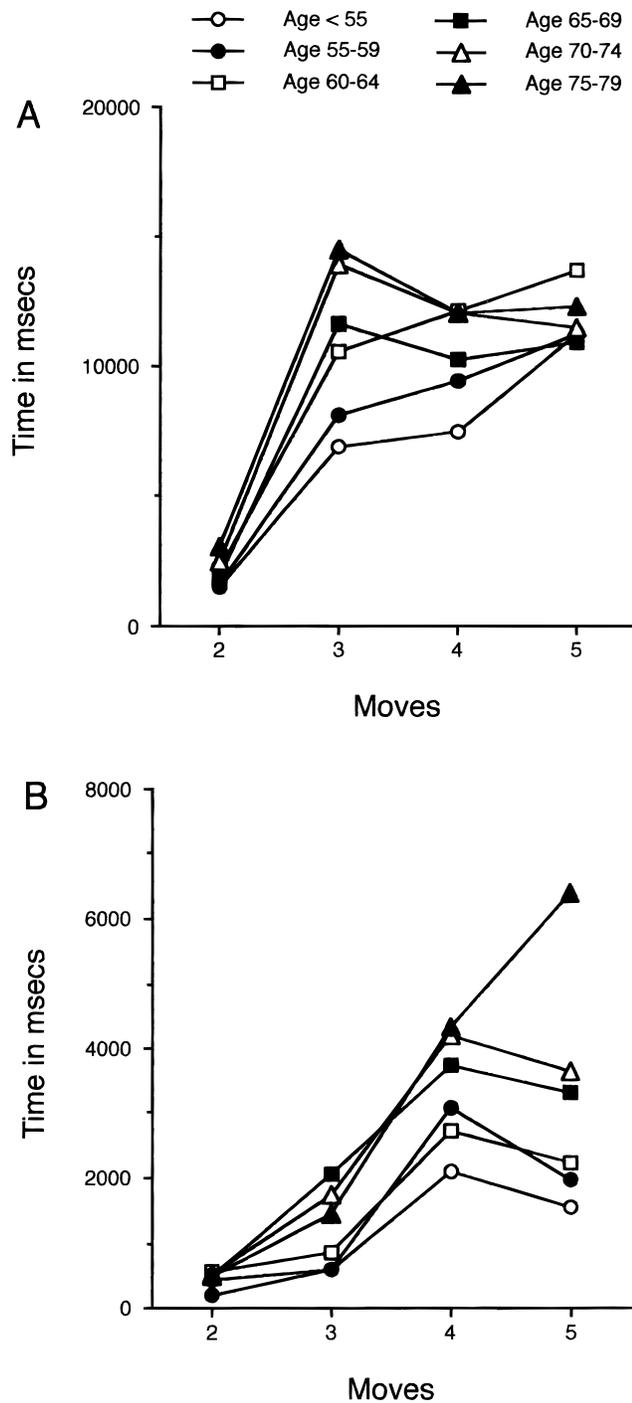
**Fig. 5.** Spatial working memory: Relationship between strategy score and memory performance (6 + 8 box stage) for each age group, as measured by Pearson's correlation coefficient  $r$ . \*\* $p < .01$ , \* $p < .05$ .



**Fig. 6.** (a) Maximum spatial span score as a function of age. (b) Tower of London test of planning, minimum-move solutions. The data are from the full sample (see Table 1).

age group [ $F(5, 209) = 2.63, p = .025$ ]. There was also a significant interaction between age groups and difficulty [ $F(15, 627) = 1.73, p = .041$ ]. When each level of difficulty was analyzed separately in order to determine simple main effects of groups, significant effects were found only for three move problems [ $F(5, 209) = 5.83, p < .001$ ]. *Post-hoc* tests applied at this stage confirmed that the oldest (75–79) age group had significantly longer latencies than the three youngest groups, and that the 70 to 74 age group was significantly slower than the two youngest groups, no other significant differences being evident.

The subsequent thinking time measure is shown in Figure 7b. Again there was a significant main effect of age group



**Fig. 7.** Tower of London test of planning. (a) Initial thinking time. (b) Subsequent thinking time. The data are from the full sample (see Table 1).

[ $F(5, 209) = 3.97, p = .002$ ]. However, there was also a significant interaction of Age Group  $\times$  Difficulty [ $F(15, 627) = 2.26, p = .004$ ]. When each level of difficulty was analyzed separately, there was a significant main effect of age group only for the most difficult (five move) problems. The oldest age group was significantly slower than each of the other groups, there being no other significant effects.

The movement time measures will not be shown or analyzed in detail. However, there was a highly significant main effect of Age Group  $\times$  Initial Movement Time [ $F(5, 194) = 10.44, p < .00001$ ]. Mean values for the six age groups, from youngest to oldest, were 1921, 2582, 2431, 2881, 3761, and 3529 ms, respectively. For subsequent movement times, the analysis again showed a significant main effect of age group [ $F(5, 194) = 8.32, p < .00001$ ]. Mean values were 2211, 2220, 2119, 2534, 2860, and 3002 ms for the six age groups in ascending order.

### Factor Analyses

The results of the varimax rotation for the main tests of executive function are shown in Table 2 (total  $N = 341$ ). A four factor solution was derived based on eigenvalues of 3.82, 1.42, 1.23, and 1.03, which cumulatively accounted for 62.5% of the variance. Only factor loadings greater than .40 are shown, as none of the others exceeded .30. The four factors probably correspond to the following constructs on the basis of the tests and measures which contribute loadings. Factor 1 includes measures of accuracy and of subsequent thinking time on the Tower of London test, as well as the between errors score on spatial working memory, the strategy measure, and spatial span. This factor may correspond to a planning/spatial working memory construct. Factor 2 includes contributions from the set formation and shifting paradigm and the yoked control task for the Tower of London, which has elements of a continuous performance test. Note the loadings of AH 4 2 both with Factor 1 and Factor 2. Factor 3 captures both of the main latency measures on the Tower of London. Presumably it does not simply reflect response speed, as neither initial nor subsequent movement time from the yoked control task load to any significant degree. Factor 4 (spatial memory) incorporates the between- and within-error scores on the self-ordered spatial working memory task and presumably reflects the more mnemonic aspects of this task.

The second factor analysis incorporated data not only from the tests of executive function described here, but also those of tests of visual memory and learning, together with visual search, that had been analyzed for a large ( $N = 771$ ) factor analysis in a previous publication (Robbins et al., 1994). The resultant analysis for  $N = 215$  is shown in Table 3 with only loadings greater than .4 listed. Six factors were generated, with eigenvalues of 5.93, 1.98, 1.60, 1.32, 1.16, and 1.07, accounting cumulatively for 62.2% of the variance. Factor 1, as in the previous study, represents primarily visual memory and learning functions, but with a contribution from the perceptual encoding of visual material (MTS visual search task). Factor 2 captures some of the variance shown under Factor 1 (planning and executive function) in the analysis shown in Table 2. Factor 3 loads mainly on latency measures for the Tower of London, MTS visual search, and delayed matching to sample and thus appears to represent a speed of response factor. Factor 4 captures other aspects of the executive and planning functions represented

**Table 2.** Summary of loadings for CANTAB tests on Factors 1–4 following factor analysis

Test	Factor 1	Factor 2	Factor 3	Factor 4
Spatial Working Memory				
Between-search errors	.70			.57
Strategy score	.70			
Tower of London				
Minimum moves	-.63			
AH 42	-.58	-.46		
Tower of London				
Initial movement time	.56	.44		
Subsequent movement time	.55	.46		
Spatial Span	-.52			
Attentional Shift				
Total errors (– ED stage errors)		.78		
ED stage errors		.77		
Tower of London				
Initial thinking time			.89	
Subsequent thinking time	.41		.73	
Spatial Working Memory				
Within-search errors				.91

by Factor 1 in Table 2. Factor 5 represents attentional set formation and maintenance (total errors) and shifting (EDS errors). Factor 6 seems equivalent to Factor 4 from the analysis shown in Table 2. It should be noted that one of the variables, spatial recognition score, did not load to any great extent on any of the six factors, though it was highest for Factor 1 (.29) and Factor 2 (–.30). The measure of fluid intelligence (AH 42) loaded most highly on Factors 1 (visual memory and learning) and 2 (planning and executive function). Loadings on Factors 3 to 6 were as follows: –.22, –.21, –.18, and –.09. When age was added as a variable to this factor analysis, it loaded as follows across Factors 1 to 6: –.30, .06, .67, .32, –.11, and .18. Thus, by far the largest loadings for age were with Factor 3, which loads highly on latency measures such as initial movement time and matching to sample, visual search reaction time.

### Correlational Analyses

Two main analyses were conducted. Analysis 1 focused on the interrelationships between the main test variables, age, and an index of speed of information-processing (choice reaction time for matching to sample between two visual stimuli) (Downes et al., 1989; Robbins et al., 1994). This enabled a test of two predictions derived from Salthouse's (1985, 1996) theory of cognitive aging (see Nettelbeck & Rabbitt, 1992). We also analyzed the contribution of fluid intelligence (AH 42 scores) to the age-related correlations with specific cognitive measures. Analysis 2 focused on the interrelationships of the various test variables after the effects of age and intelligence had been partialled out to test the hypothesis of Duncan et al. (1995) concerning the equivalence

of tests of fluid intelligence to those sensitive to frontal lobe dysfunction.

### Analysis 1

Table 4 shows the intercorrelation matrix for various test battery variables and measures of age, speed (i.e., choice reaction time on the matching to sample task), and intelligence (AH 42 and NART scores). In order to reduce the number of comparisons, other variables relating to the matching to sample, visual search and delayed matching to sample tasks were excluded.

Also shown in Table 4 are the effects of speed when age-related variance is controlled by partial correlation and the effects of age and when speed-related variance is similarly partialled out. According to Salthouse (1985, 1996) and Nettelbeck and Rabbitt (1992):

*If age only affects cognition by reducing information-processing speed, then correlations between mental speed and performance on cognitive tasks should be unaffected by controlling for (i.e., partialling out) the effects of chronological age. . . . If information-processing speed is the sole factor that mediates correlations between chronological age and cognitive performance on a particular skill, then partialling out the contributions of individual differences in speed should eliminate the correlations between age and performance. (Nettelbeck and Rabbitt, 1992, p. 191)*

The first prediction is upheld to some extent, except for the tests of spatial working memory and attentional set shifting. The second is also upheld, as age-related variance was attenuated considerably by partialling out the effects of speed. Considering the  $r^2$  values, this attenuation of age-related variance was by over 80% for spatial working memory, ex-

**Table 3.** Summary of loadings for CANTAB tests on Factors 1–6 following factor analysis

Test	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
Paired Associate Learning						
Trials	-.76					
Memory score	.76					
DMTS (del) accuracy	.73					
Pattern Recognition	.61					
DMTS (sim) accuracy	.54					
MTS (visual search) accuracy	-.48					
Spatial Working Memory						
Between-search errors		.67				.58
Tower of London						
Initial movement time		.67				
Spatial Working Memory						
Strategy score		.62				
Spatial Span		-.57				
Tower of London						
Subsequent movement time		.54				
AH 4 2	.43	-.51				
Spatial Recognition						
MTS (visual search) latency			.76			
DMTS latency			.75			
Tower of London						
Subsequent thinking time				.78		
Initial thinking time			.46	.60		
Minimum moves				-.54		
Attentional Shift						
Total errors (- ED stage errors)					.81	
ED errors					.81	
Spatial Working Memory						
Within-search errors						.87

DMTS, delayed matching to sample; del, delay trials; sim, simultaneous condition trials; MTS, matching to sample; ED, extra-dimensional shift.

tradimensional shifting, and paired associates. However, significant correlations remained for between search errors, minimum move solutions, and initial thinking time on the Tower of London, as well as on paired associates learning.

The parallel analysis partialled out effects of AH 4 2 from the correlations between age and cognitive performance and found a globally similar picture as for speed: that the values of  $r$  were significantly diminished, but remained significant in many instances (Table 4). Finally, when effects of both speed and fluid intelligence (AH 4 2) were partialled out together from the correlations of test performance with age, there were substantial residual correlations with test measures from the spatial working memory test (between search errors), the Tower of London task (minimum move solutions), and paired associates learning (see last column of Table 4).

### Analysis 2

As Table 4 shows, there were few major correlations between the cognitive test variables and either AH 4 2, the measure of fluid intelligence, or the NART score, which estimates verbal intelligence. The only obvious exception was pattern recognition memory, which correlated ( $p < .001$ ) with both.

However, when the effects of age were partialled out of these correlations, both AH 4 2 and NART scores, in slightly different ways, were significantly related to performance on many of the test measures. The NART and AH 4 2 scores were highly correlated, as would be expected. Table 5 shows the intercorrelation matrix for most of the main task variables after partialing out the effects of fluid intelligence (AH 4 2). Substantially the same results were obtained after controlling for the effects of NART or of NART and AH 4 2 combined. As can be seen, there were a substantial residual number of significant correlations, where  $r$  was significant beyond the .001 level when the effects of intelligence were controlled in this way.

## DISCUSSION

This study has provided some information on the performance of healthy normal volunteers on a subset of tests from the CANTAB battery, which has been used in the neuropsychological assessment of patients with neurological and neuropsychiatric disorders. The subset, which included tests of self-ordered working memory, planning, and attentional set shifting, has been shown to be particularly sensitive to frontostriatal syndromes ranging from neurosurgical exci-

**Table 4.** Correlations with speed, age, and intelligence

Test	Speed	Speed/ Age partialed	Age	Age/ Speed partialed	AH42	NART	Age/ AH42 partialed	Age/ AH42 & speed partialed	AH42/ Age partialed	NART/ Age partialed
Spatial Working Memory										
Between-search errors	31**	15	43**	19*	−05	03	25**	21*	−29**	−29**
Within-search errors	17*	13	12	−04	−08	03	00	−04	−08	−19*
Strategy score	12	05	35**	10	−08	11	11	11	−22*	06
Tower of London										
Minimum moves	0	10	−31**	−27**	11	−04	−23	−29**	20*	13
Initial thinking time	26**	19*	19*	21**	04	01	14	−05	05	03
Subsequent thinking time	25**	25**	27**	11	−15	00	17*	09	−29**	−19*
Attentional set shifting										
Total errors (− ED)	21*	07	17*	11	−05	00	13	12	−23**	−05
ED errors	15	07	25**	03	−07	−17*	03	03	−30**	−23**
Paired Associates										
Total trials	22*	17*	42**	20*	−15	−16*	26**	25**	−33**	−27**
Spatial Recognition	−27**	−20*	−23**	−13	16*	13	−21*	−13	29**	16
Spatial Span	−22*	−18*	−33**	−14	15	−12	−20*	−16	28**	19*
Pattern Recognition	−21*	−15	−18*	−11	29**	−27**	−15	−12	42**	32**
AH42	−13	−30**	19*	01	—	45**	—	—	—	50**
NART	−14	−11	12	−07	45**	—	−07	−09	50**	—

\* $p < .01$ ; \*\* $p < .001$ ; two-tailed probability.  $N = 215$ . Total errors (− ED) = errors up to but not including EDS + EDR stages. Other measures defined in text. Values shown are  $r$  values with decimal points omitted.

sions for tumors or epilepsy (e.g., Owen et al., 1990) to Parkinson’s (e.g., Owen et al., 1992) and Huntington’s diseases (Lawrence et al., 1996). The present data set is derived from the performance of 341 normal participants ranging in age

from 21 to 79 years. This sample includes a subset of a larger population ( $N = 787$ ) who also received other tests from the CANTAB battery involving tests of visual memory, learning, and perception (Robbins et al., 1994, in press). It ex-

**Table 5.** Correlation matrix for main task variables with AH 4 2 scores partialed out

Test	BS	WS	St	Min	IT	ST	Total errors (− ED)	EDS	PA	SR	SS	PR
Spatial Working Memory												
Between-search errors	—	53**	53**	−36**	−04	25**	12	20**	27**	−33**	−40**	−24**
Within-search errors		—	12	−10	−02	08	20**	17*	11	−08	−12	−27**
Strategy score			—	−31**	−03	25**	11	16**	15	−20**	−22**	−10
Tower of London												
Minimum moves				—	06	−46**	−13	−22**	−35**	21**	18**	32**
Initial thinking time					—	45**	09	05	03	−05	01	−12
Subsequent thinking time						—	21**	28**	32**	−24**	−16*	−31**
Attentional set shifting												
Total errors (− ED)							—	35**	20**	−17*	−04	−21**
ED errors								—	27**	−33**	−20**	−26**
Paired Associates												
Total trials									—	−33**	−25**	−56**
Spatial Recognition												
Spatial Span										—	−27**	33**
Pattern Recognition												—

\* $p < .01$ ; \*\* $p < .001$ ; two-tailed probabilities.  $N = 215$ . Total errors (− ED) = errors up to but not including EDS + EDR stages. Values shown are  $r$  values with decimal points omitted.

BS, between search error; WS, within search errors; St, strategy score; Min, minimum move solutions; IT, initial thinking time; ST, subsequent thinking time; EDS, extradimensional shift; PA, paired associates; SR, spatial recognition; SS, spatial span; PR, pattern recognition.

tends that study by presenting data from the tests of executive function (in the case of the spatial working memory test, in considerably more detail than in the earlier study) and considers the structural relationship of these tests to the CANTAB battery as a whole.

The results also address two main questions. First, whether tests sensitive to frontal lobe and basal ganglia dysfunction are especially sensitive to the effects of aging, as compared, for example, to tests in the CANTAB battery such as pattern recognition, which are more sensitive to damage of the temporal lobes (cf. West, 1996). Second, whether any relationship between particular tests of executive function and age can be attributed to general factors, such as information-processing speed (Nettelbeck & Rabbitt, 1992; Salthouse, 1985, 1996) or fluid intelligence (Duncan et al., 1995).

### Sensitivity of Cognitive Tests to Aging

Considering first the performance of the young group (< 55 age group) *versus* the other groups, the largest differences were found on the test of attentional set shifting, specifically at the extradimensional shift stage. This greatly modified analogue of the WCST significantly differentiated the under 55 group from the 55 to 59 subgroup, and it is noteworthy that it has been shown previously to be particularly sensitive to basal ganglia dysfunction (Downes et al., 1989; Lawrence et al., 1996; Owen & Robbins, 1993) and excisions of the frontal, though not temporal, lobe (Owen et al., 1991). The data also accord very well with analogous analyses reported by Daigneault et al. (1992) for the relatively early decline in performance on the WCST in normal individuals age 50 to 59 years. By contrast, there were no significant differences between the youngest group and the other age groups on tests of spatial span, spatial working memory, or spatial planning, as indexed by the Tower of London test. The latter two tests have been found sensitive to frontal lobe damage in a group study (Owen et al., 1990), as well as basal ganglia damage, although in qualitatively different ways from the effects of frontal lobe excisions (Owen & Robbins, 1993). In a previous comparison of the larger population ( $N = 787$ ), subsuming the present participants (Robbins et al., 1994, 1996), we found with the aid of the greater statistical power afforded by the larger sample that there was a significant difference between the under 55 and the 55 to 59 subgroups on the spatial working memory test and also for two of the other tests from the present battery (paired associates learning and delayed matching to sample). The spatial working memory test is susceptible to damage to the temporal, as well as the frontal, lobes, although the deficits are milder and are not related to impairments in the use of strategy for temporal patients (Owen et al., 1995, 1996c). The earlier declines in performance on the spatial working memory test in this study were also not clearly related to impaired use of strategy. An analogous result has been reported by Daigneault et al. (1992) using the Petrides and Milner (1982) self-ordered memory task. Moreover, while the paired associates task is also sensitive to both

forms of neurosurgical excision, the delayed matching to sample test is much more sensitive to temporal lobe than frontal lobe damage (Owen et al., 1995). Thus, it would be difficult to claim that frontal functions are the only ones susceptible to decline in 55 year old normal individuals, although the present extradimensional shift deficit indicates that aspects of cognitive flexibility are already compromised by late middle age.

Some aspects of performance (e.g., spatial span; minimum move solutions on the Tower of London) held up remarkably well in the eldest (75–79) group, but other aspects resembled the performance of individuals with frontal lobe damage. On the attentional set shifting test, there was a trend for some of the participants from this group to fail on the earlier stages preceding the critical intradimensional and extradimensional shift stages (Figure 2), but this failed to reach significance. The 75 to 79 age group, however, made significantly more errors than did the 70 to 74 year group at the critical extradimensional shift stage, suggesting that this test was especially sensitive to age-related decline. This is also consistent with evidence of prefrontal alterations in a functional activation study in the elderly of Wisconsin Card Sort performance (Esposito et al., 1995).

This apparently selective decline in performance in the most elderly group was also evident in two other executive measures from the tests sensitive to frontal lobe dysfunction. Thus, on the spatial working memory test this group selectively lost the benefit for overall performance normally provided by the use of a defined strategy (Figure 5). Specifically, although they showed the same tendency to employ the strategy as for younger groups, the normally high correlation with memory performance, as measured by between-search errors was lost. This could be interpreted as a failure to implement strategy to the same degree as normal. There was also a significant, selective lengthening in the subsequent thinking time measure of the Tower of London Test of planning relative to all other age groups, which is normally seen in patients with frontal lobe damage and interpreted to reflect the greater need in this group to monitor and correct incompletely formulated plans (Owen et al., 1990). These signs of frontal lobe dysfunction may reflect impairments in the executive monitoring of performance of relatively late onset in the normal elderly population. The analysis would, of course, be strengthened by evidence from repeated testing of the sample in a longitudinal design, which would avoid possible cohort effects; this is planned in further work.

### Relation to “Speed of Processing”

The results do not support a strong version of Salthouse’s (1985, 1996) theory of cognitive aging that age-related declines in performance can be reduced to a single underlying deficit in speed of information-processing. Specific tests of this hypothesis were made by considering the relationships between measures of speed of information-processing, age, and test performance. Consistent with Salthouse’s hypoth-

esis, partialing out the effects of age (Table 4) failed to remove existing correlations between speed of processing and certain CANTAB test measures. Moreover, controlling for the effects of speed (Table 4) significantly reduced age-related variance in virtually all of the computerized test measures. However, some of the data were inconsistent with a strong version of the Salthouse hypothesis. Specifically, controlling for age did render nonsignificant the correlations between speed and either spatial working memory or attentional set-shifting performance, suggesting that variance associated with information-processing rate was not primarily associated with these tests. In addition, significant residual correlations were evident for spatial working memory, certain measures of Tower of London performance, and paired associates learning after partialing out the effects of speed by itself or in conjunction with a measure of fluid intelligence (AH 4 2). Thus, the relationship of age to performance on these tests was also not wholly secondary to changes in information-processing rate. The persistent relationship between age and the test of visuospatial paired associates learning is consistent with data reported by Nettelbeck and Rabbitt (1992) and by Nettelbeck et al. (1996), even though they employed somewhat different indices of speed of information-processing. Our results are also compatible with those reported by Rabbitt and Maylor (1991) and Baddeley (1996), which also suggest that certain executive functions may be related to age, over and above any relationships with speed of processing and IQ.

Some caveats are, however, in order. A study by Roznowski (1993) made it clear that the assumption of linear relationships between speed of processing and cognitive performance in the predictions of the Nettelbeck and Rabbitt (1992) study may not always hold, especially when performance reaches asymptotic levels. The second problem is the possible lack of purity of some of the measures, for example, of speed of processing, fluid intelligence, and indeed chronological age itself (which may better be determined by an hitherto undiscovered index of biological aging). Nevertheless, the apparent consistency of findings across different tests and measures of speed of processing is encouraging.

### Relevance to Neural Theories of Aging

With its emphasis on the rate of information-processing, Salthouse's theory is relevant to the neuropsychological concept of subcortical (or frontosubcortical) dementia (Cummings, 1986) and appears to be consistent with the burgeoning evidence for accelerated aging in frontostriatal structures (see West, 1996). However, while some of our psychological evidence is consistent with this biological perspective, we have no specific evidence to support the hypothesis that aging selectively implicates subcortical or frontosubcortical circuitry. Thus, of the other tests for which marked declines in performance occur in the most elderly of the two age groups (Robbins et al., 1994), the delayed matching to sample test is particularly sensitive to tempo-

ral, but not frontal, lobe damage, and the paired associates task is sensitive to damage to both the temporal and the frontal lobes, as well as the basal ganglia (Owen et al., 1993, 1995).

Overall, our findings are compatible with the results of studies using a variety of neuroimaging modalities that aging is especially associated with structural, metabolic, and neurochemical changes in regions of the prefrontal cortex (Pantano et al., 1984; Riege et al., 1985; Waldemar et al., 1991), the striatum (Murphy et al., 1992; Riege et al., 1985; Rinne et al., 1993; Sawle et al., 1990; van Dyck et al., 1995), and also the temporal lobes (Arrigada et al., 1992; Loessner et al., 1995; Pantano et al., 1984; Price et al., 1991). This position is not very different from that reached by West (1996), after a comprehensive review of the available neurobiological evidence.

### Relationship of Executive Test Performance To General (Fluid) Intelligence

An important part of the present report is the principal component analyses relating different aspects of performance on the three tests of executive function to one another, to a measure of fluid intelligence (performance on the AH 4 2 intelligence test), and to performance on other tests from the CANTAB battery. A solution was found for the tests sensitive to frontal lobe dysfunction, suggesting that at least four factors are necessary to account for individual differences in performance. The first factor, which captured loadings from both the Tower of London task and the test of self ordered spatial working memory, confirm the finding from neuropsychological and functional neuroimaging studies that these tests have many common cognitive requirements (Owen et al., 1996a,b; Robbins, 1996). However, there were differences between the two tests on certain measures. Thus, some of the variance associated with the latency measures loaded with the test of attentional set shifting and some (initial and subsequent thinking time) on a separate factor. This confirmed the impression, also gained from both neuropsychological and neuroimaging studies (Owen et al., 1996b), that the Tower of London test has several cognitive components.

The measures of strategy and performance accuracy (between-search errors) for the spatial working memory test were closely related, as expected from previous studies (Owen et al., 1991). However, they also covaried with the minimum moves accuracy measure of the Tower of London test of planning, confirming that this factor loads highly on strategic aspects of executive function. Some of the memory requirements on the spatial working memory test itself loaded separately (Factor 4), as would be expected of a test having separable mnemonic and executive components (Owen et al., 1996c; Robbins, 1996).

Performance on the test of attentional set shifting loaded mainly on an independent factor (Factor 2), only covarying with the measures of movement time. The relationship between these seemingly very different variables may have

been observed because the movement-time measures presumably index, to some degree, the speed of processing processes that do seem to account for much of the age-related performance on the attentional set shifting test (Table 4). Nevertheless, it is striking that the test of attentional set shifting segregates so completely from the others in this battery. Consideration of the relationships with the overall CANTAB battery (Table 3) shows a similar independence from other tests. In a related analysis (Robbins et al., 1997), we found that the Hayling test of frontal lobe functioning (Burgess & Shallice, 1996a) is also independent from the present tests of executive function based on a factor analysis of patients performing tests in addition to the CANTAB battery. This analysis showed that the traditional tests of frontal lobe functioning, such as cognitive estimates and verbal fluency, were more related to the CANTAB Visual Memory tests than to any of the tests sensitive to frontal lobe dysfunction. However, the Brixton test of spatial anticipation, a spatial set formation and shifting task (Burgess & Shallice, 1996b), did load with accuracy measures from the Tower of London probably because of the contribution of such set shifting functions to efficient problem solving (Robbins et al., 1997).

These patterns of intercorrelations of measures, from executive and nonexecutive tests, which show some clustering but also considerable independence among the different executive test measures, are not dissimilar to the observations of other investigators who employed different batteries (Baddeley, 1996; Burgess, in press; Lehto, 1996). For example, the EDS errors score correlated with Tower of London performance to a lesser extent than did pattern recognition performance and also loaded on measures other than those related to planning on the factor analyses shown in Tables 2 and 3. On the other hand, spatial working memory scores were quite closely related to Tower of London performance. Overall, these patterns support the conclusion that executive functions are diverse, being probably governed by different psychological, as well as neural, structures—an “executive committee” rather than an “homunculus,” as Baddeley (1996) put it. Their common dependence on the integrity of frontal lobe functioning does not contradict this conclusion, as the prefrontal cortex itself is an extensive and heterogeneous anatomical structure, with specific connections to many other brain structures, both cortical and subcortical (Pandya & Yeterian, 1995; see Robbins, 1996).

This heterogeneity is gradually being related to discrete and different aspects of executive function using evidence of selective lesions of different parts of the prefrontal cortex in monkeys (Dias et al., 1996); evidence of different stages of memory processing mediated by different portions of the prefrontal cortex based on findings from functional neuroimaging studies (Owen et al., 1996a; Petrides, 1989, 1996); and evidence of lateralization of prefrontal cortical function in humans (Shallice et al., 1994; Tulving et al., 1994).

The hypothesis that executive processes can be distinguished from processes that may depend on dedicated posterior cortical “modules” is also supported by other aspects

of the factor analyses conducted here. Thus, the measure of fluid intelligence, AH 4 2 performance, loaded on both the planning/spatial working-memory factor and the attentional set shifting factor, but was not related uniquely to either. When the entire battery was analyzed, the AH 4 2 variable also showed considerable loadings on the factor tapping visual memory and learning (Factor 1, Table 3), again confirming its utility as an index of general intelligence but further showing the distinctive nature of some of the measures of the executive test battery. This is substantiated by the demonstration that, when considered separately, performance on the AH4 2 correlated significantly with a test insensitive to frontal lobe damage, pattern recognition, and also paired associates learning. However, as Table 4 makes clear, when the effect of age is partialled out of the relationship between performance on the computerized tests and the NART and AH 4 2 scores, both of the latter are significantly related to a number of the measures, consistent with the factor analyses discussed above.

Overall, this study has provided detailed information from a large sample of normal volunteers on a battery of neuropsychological tests that has already been shown to be useful in the assessment of brain damaged and psychiatric patients. The data support evidence from other neuropsychological and neuroimaging studies of the specificity of some of these tests in measuring different aspects of executive function and also contribute to cognitive theories of aging by providing an important neuropsychological perspective.

## ACKNOWLEDGMENTS

The work was supported by a Programme Grant from the Wellcome Trust to T.W. Robbins, B.J. Everett, A.C. Roberts, B.J. Sahakian, and by the Nuffield Foundation. We thank C. Brennard, J. Hunter, J. Iddon, M. Robson, and J. Swain for assistance and for the volunteers taking part. We also thank C. van Dyck for discussion. A.D.L. held an MRC Studentship.

## REFERENCES

- Abas, M.A., Sahakian, B.J., & Levy, R. (1990). Neuropsychological deficits and CT scan changes in elderly depressives. *Psychological Medicine*, *20*, 507–520.
- Arrigada, P.V., Marzloff, B.A., & Hyman, B.T. (1992). Distribution of Alzheimer-type pathologic changes in nondemented elderly individuals matches the pattern in Alzheimer’s disease. *Neurology*, *42*, 1681–1688.
- Baddeley, A.D. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology*, *49A*, 5–28.
- Beats, B.C., Sahakian, B.J., & Levy, R. (1996). Cognitive performance in tests sensitive to frontal lobe dysfunction in the elderly depressed. *Psychological Medicine*, *26*, 591–603.
- Benton, A.L. & Hamsher, K. deS. (1976). *Multilingual aphasia examination*. Iowa City: University of Iowa.
- Burgess, P.W. (1997). Theory and methodology in executive function research. In P. Rabbitt (Ed.), *Methodology of frontal and executive functions*, (pp. 81–111). Hove: Psychology Press.

- Burgess, P.W. & Shallice, T. (1996a). Response suppression, initiative and strategy use following frontal lobe lesions. *Neuropsychologia*, *34*, 263–272.
- Burgess, P.W. & Shallice, T. (1996b). Bizarre responses, rule detection and frontal lobe lesions. *Cortex*, *32*, 241–259.
- Cummings, J.L. (1986). Subcortical dementia: Neuropsychology, neuropsychiatry and pathophysiology. *British Journal of Psychiatry*, *149*, 682–697.
- Daigneault S., Braun, C.M.J., & Whitaker, H. (1992). Early effects of normal aging on perseverative and non-perseverative prefrontal measures. *Developmental Neuropsychology*, *8*, 99–114.
- Dias, R., Roberts, A., & Robbins, T.W. (1996). Dissociation in prefrontal cortex of affective and attentional shifts. *Nature*, *380*, 69–72.
- Downes, J.J., Roberts, A.C., Sahakian, B.J., Evenden, J.L., & Robbins, T.W. (1989). Impaired extra-dimensional shift performance in medicated and unmedicated Parkinson's disease: Evidence for a specific attentional dysfunction. *Neuropsychologia*, *27*, 1329–1344.
- Duncan, J., Burgess, P., & Emslie, H. (1995). Fluid intelligence after frontal lobe lesions. *Neuropsychologia*, *33*, 261–268.
- Elliott, R., McKenna, P.J., Robbins, T.W., & Sahakian, B.J. (1995). Neuropsychological evidence for fronto-striatal dysfunction in schizophrenia. *Psychological Medicine*, *25*, 619–630.
- Esposito, G., Kirkby, B.S., Van-Horn, J.D., Weinberger, D.R., & Berman, K.F. (1995). Different pathophysiological mechanisms of altered Wisconsin card sort performance in schizophrenics and elderly normal subjects. *Society for Neuroscience Abstracts*, *297*, 16.
- Heim, A. (1968). *The AH-4 IQ Test*. Slough, NFER-Nelson.
- Hughes, C., Russell, J., & Robbins, T.W. (1994). Evidence for executive dysfunction in autism. *Neuropsychologia*, *32*, 477–492.
- Joyce, E.M. & Robbins, T.W. (1991). Frontal lobe function in Korsakoff and non-korsakoff alcoholics: Planning and spatial working memory. *Neuropsychologia*, *29*, 709–723.
- Kullback, S. (1968). *Information theory and statistics*. New York: Dover.
- Lawrence, A.D., Sahakian, B.J., Hodges, J.R., Rosser, A.E., Lange, K.W., & Robbins, T.W. (1996). Executive and mnemonic functions in early Huntington's disease. *Brain*, *119*, 1633–1645.
- Lehto, J. (1996). Are executive function tests dependent on working memory capacity? *Quarterly Journal of Experimental Psychology*, *49A*, 29–50.
- Loessner, A., Alavi, A., Lewandrowski, K.-U., Mozeley, D., Souder, E., & Gur, R.E. (1995). Regional cerebral function determined by FDG-PET in healthy volunteers: Normal patterns and changes with age. *Journal of Nuclear Medicine*, *36*, 1141–1149.
- Milner, B. (1963). Effects of different brain lesions on card sorting: The role of the frontal lobes. *Archives of Neurology*, *9*, 100–110.
- Milner, B. (1971). Interhemispheric differences in the localisation of psychological processes in man. *British Medical Bulletin*, *27*, 272–277.
- Murphy, D.G.M., DeCarli, C., Schapiro, M.B., Rapoport, S.I., & Horwitz, B. (1992). Age-related differences in volumes of subcortical nuclei, brain matter, and cerebrospinal fluid in healthy men as measured with magnetic resonance imaging. *Archives of Neurology*, *49*, 839–844.
- Nelson, H.E. (1982). *National Adult Reading Test (NART) test manual*. Windsor: NFER-Nelson.
- Nettelbeck, T. & Rabbitt, P.M.A. (1992). Aging, cognitive performance and mental speed. *Intelligence*, *16*, 189–205.
- Nettelbeck, T., Rabbitt, P.M.A., Wilson, C., & Batt, R. (1996). Uncoupling learning from initial recall: The relationship between speed and memory deficits in old age. *British Journal of Psychology*, *87*, 593–607.
- Norusis, M.J. (1990). *SPSS/PC + Statistics 4.0*. Chicago: SPSS, Inc.
- Owen, A., Downes, J.J., Sahakian, B.J., Polkey, C.E., & Robbins, T.W. (1990). Planning and spatial working memory following frontal lobe lesions in man. *Neuropsychologia*, *28*, 1021–1034.
- Owen, A.M., Beksinska, M., James, M., Leigh, P.N., Summers, B.A., Quinn, N.P., Sahakian, B.J., & Robbins, T.W. (1993). Visuospatial memory deficits at different stages of Parkinson's disease. *Neuropsychologia*, *31*, 627–644.
- Owen, A.M., Doyon, J., Petrides, M., & Evans, A.C. (1996b). Planning and spatial working memory: A positron emission tomography study in humans. *European Journal of Neuroscience*, *8*, 353–364.
- Owen, A.M., Evans, A.C., & Petrides, M. (1996a). Evidence for a two-stage model of spatial working memory processing within the lateral frontal cortex: A positron emission tomography study. *Cerebral Cortex*, *6*, 31–38.
- Owen, A.M., James, M., Leigh, P.H., Summers, B.A., Marsden, C.D., Quinn, N.P., Lange, K.W., & Robbins, T.W. (1992). Fronto-striatal cognitive deficits at different stages of Parkinson's disease. *Brain*, *115*, 1727–1751.
- Owen, A.M., Morris, R.G., Sahakian, B.J., Polkey, C.E., & Robbins, T.W. (1996c). Double dissociations of memory and executive functions in working memory task following frontal lobe excision, temporal lobe excisions or amygdala-hippocampectomy in man. *Brain*, *119*, 1597–1615.
- Owen, A.M. & Robbins, T.W. (1993). Comparative neuropsychology of parkinsonian syndromes. In P. Wolters & E. Ch. Scheltens (Eds.), *Mental dysfunction in Parkinson's disease* (pp. 221–242). Amsterdam: Vrije University Press.
- Owen, A.M., Roberts, A.C., Polkey, C.E., Sahakian, B.J., & Robbins, T.W. (1991). Extra-dimensional versus intradimensional set shifting performance following frontal lobe excision, temporal lobe excision or amygdalo-hippocampectomy in man. *Neuropsychologia*, *29*, 993–1006.
- Owen, A.M., Sahakian, B.J., Semple, J., Polkey, C.E., & Robbins, T.W. (1995). Visuospatial short term recognition memory and learning after temporal lobe excisions, frontal lobe excisions or amygdala-hippocampectomy in man. *Neuropsychologia*, *33*, 1–24.
- Pandya, D.P. & Yeterian, E.H. (1995). Morphological correlations of human and monkey frontal lobe. In A.E. Damasio, H. Damasio, & Y. Christen (Eds.), *Neurobiology of human decision-making* (pp. 13–46). New York: Springer.
- Pantano, P., Baron, J.-C., Lebrun-Grandié, P., Duquesnoy, N., Bousser, M.-G., & Comar, D. (1984). Regional cerebral blood flow and oxygen consumption in human aging. *Stroke*, *15*, 635–641.
- Petrides, M. (1989). Frontal lobes and memory. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (Vol. 3, pp. 75–90). Amsterdam: Elsevier.
- Petrides, M. (1996). Specialized system for the processing of mnemonic information within the primate prefrontal cortex. *Proceedings of the Royal Society of London B*, *351*, 1455–1462.
- Petrides, M. & Milner, B. (1982). Deficits on subject-ordered tasks after frontal- and temporal lobe lesions in man. *Neuropsychologia*, *20*, 249–262.

- Price, J.L., Davies, P.B., Morris, J.C., & White, D.L. (1991). The distribution of tangles, plaques, and related immunohistochemical markers in healthy aging and Alzheimer's disease. *Neurobiology of Aging*, *12*, 295–312.
- Rabbitt, P.M.A. & Maylor, E. (1991). Investigating models of human performance. *British Journal of Psychology*, *82*, 259–290.
- Riege W.H., Metter, E.J., Kuhl, D.E., & Phelps, M.E. (1985). Brain glucose metabolism and memory functions: Age decrease in factor scores. *Journal of Gerontology*, *40*, 459–467.
- Rinne, J.O., Hietala, J., Ruotsalainen, U., Sako, E., Laihininen, A., Nagren, K., Lehtikoinen, P., Oikonen, V., & Syvalahti, E. (1993). Decrease in human striatal dopamine D2 receptor density with age: A PET study with [<sup>11</sup>C] raclopride. *Journal of Cerebral Blood Flow Metabolism*, *13*, 310–314.
- Robbins, T.W. (1977). A critique of the methods available for the measurement of locomotor activity. In L.L. Iversen, S.D. Iversen, & S.H. Snyder (Eds.), *Handbook of psychopharmacology* (Vol. 7, pp. 37–82). New York: Plenum.
- Robbins, T.W. (1996). Dissociating executive functions of the prefrontal cortex. *Philosophical Transactions of the Royal Society of London B*, *351*, 1463–1471.
- Robbins, T.W., James, M., Owen, A., Sahakian, B.J., McInnes, L., & Rabbitt, P.M. (1994). Cambridge Neuropsychological Test Automated Battery (CANTAB): A factor analytic study of a large sample of normal elderly volunteers. *Dementia*, *5*, 266–281.
- Robbins, T.W., James, M., Owen, A.M., Sahakian, B.J., McInnes, L., & Rabbitt, P. (1997). A neural systems approach to the cognitive psychology of ageing using the CANTAB battery. In P. Rabbitt (Ed.), *Methodology of frontal and executive function* (pp. 215–235). Hove: Erlbaum.
- Roberts, A., Everitt, B.J., & Robbins, T.W. (1988). Extra- and intra-dimensional shifts in man and marmoset. *Quarterly Journal of Experimental Psychology*, *40B*, 321–342.
- Roznowski, M. (1993). Measures of cognitive processes: The stability of other psychometric and measurement properties. *Intelligence*, *17*, 361–388.
- Sahakian, B.J., Downes, J.J., Roberts, A.C., Philpot, M., Levy, R., & Robbins, T.W. (1990). Preserved attentional function and impaired mnemonic function in dementia of the Alzheimer type. *Neuropsychologia*, *28*, 1197–1213.
- Sahakian, B.J., Elliott, R., Low, N., Mehta, M., Clark, R.T., & Pozniak, A.L. (1995). Neuropsychological deficits in tests of executive function in asymptomatic and symptomatic HIV-1 seropositive men. *Psychological Medicine*, *25*, 1233–1246.
- Sahakian, B.J., Morris, R.G., Evenden, J.L., Heald, A., Levy, R., Philpot, M., & Robbins, T.W. (1988). A comparative study of visuo-spatial memory and learning in Alzheimer-type dementia and Parkinson's disease. *Brain*, *111*, 695–718.
- Sahgal, A., Sahakian, B.J., Robbins, T.W., Wright, C., Lloyd, S., Cook, J.H., McKeith, I., Disley, J.C.A., Eagger, S., Boddington, S., & Edwardson, J.A. (1991). Detection of memory and learning deficits in Alzheimer's disease using the Cambridge Neuropsychological Test Automated Battery. *Dementia*, *2*, 150–158.
- Sahgal, A., Lloyd, S., Wray, C.J., Gallaway, P.H., Robbins, T.W., Sahakian, B.J., McKeith, I.G., Cook, J.H., Disley, J.C.A., & Edwardson, J.A. (1992). Does visuospatial memory in Alzheimer's disease depend on the severity of the disorder? *International Journal of Geriatric Psychiatry*, *7*, 427–436.
- Salthouse, T.L. (1985). *Theoretical perspectives on cognitive aging*. Hillsdale, NJ: Erlbaum.
- Salthouse, T.L. (1996). Constraints on theories of cognitive aging. *Psychonomic Bulletin & Review*, *3*, 287–299.
- Sawle, G.V., Colebatch, J.G., Shah, A., Brooks, D.J., Marsden, C.D., & Frackowiak, R.J.S. (1990). Striatal function in normal aging—implications for Parkinson's disease. *Annals of Neurology*, *28*, 799–804.
- Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London B*, *298*, 199–209.
- Shallice, T. & Evans, M.E. (1978). The involvement of the frontal lobe in cognitive estimation. *Cortex*, *14*, 294–303.
- Shallice, T., Fletcher, P., Frith, C.D., Grasby, P., Frackowiak, R.S.J., & Dolan, R. (1994). Brain regions associated with acquisition and retrieval of verbal episodic memory. *Nature*, *368*, 633–635.
- Tranel, D., Anderson, S.W., & Benton, A. (1994). Development of the concept of “executive function” and its relationships to the frontal lobes. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (Vol. 9, pp. 125–148). Amsterdam: Elsevier.
- Tulving, E., Kapur, S., Markowitsch, H., Craik, F.I.M., Habib, S., & Houle, S. (1994). Hemispheric coding/retrieval asymmetry in episodic memory: Positron emission tomography findings. *Proceedings of the National Academy of Sciences U.S.A.*, *91*, 2012–2025.
- van Dyck, C.H., Seibyl, J.P., Malison, R.T., Laruelle, M., Wallace, E., Zoghbi, S.S., Zea-Ponce, Y., Baldwin, R.M., Charney, D.S., Hoffer, P.B., & Innis, R.B. (1995). Age-related decline in striatal dopamine transporter binding with iodine-123-b-CIT SPECT. *Journal of Nuclear Medicine*, *36*, 1175–1181.
- Vendrell, P., Junqué, C., Pujol, J., Angles Jurado, M.A., Molet, J., & Grafman, J. (1995). The role of prefrontal regions in the Stroop task. *Neuropsychologia*, *33*, 341–352.
- Waldemar, G., Hasselbalch, S.G., Anderson, A.R., Delecluse, F., Petersen, P., Johnsen, A., & Paulson, O.B. (1991). <sup>99m</sup>Tc-*d,l*-HMPAO and SPECT of the brain in normal aging. *Journal of Cerebral Blood Flow and Metabolism*, *11*, 508–521.
- West, R.L. (1996). An application of prefrontal cortex function theory to cognitive aging. *Psychological Bulletin*, *120*, 272–292.