Patrick Rabbitt · Christine Lowe

Patterns of cognitive ageing

Received: 31 March 1999 / Accepted: 23 July 1999

Abstract Neuroanatomical evidence suggests that normal ageing affects some brain areas, and the “local” functions they support, earlier and more severely than others. Changes appear to be especially marked in the hippocampus, temporal association and prefrontal cortex. Evidence from classical neuropsychological studies suggests that these brain areas are associated with memory and “executive” functions, respectively. We may, therefore, expect that tests purported to measure these functions may be disproportionately affected in old age and that there may be evidence for some separation of these functions even within neurologically normal populations. What we also know, however, is that measures reflecting general fluid ability also decline with increasing age, so any hypothesis relating to specific “local” deficits must acknowledge and account for any “globar” changes in performance. Volunteers (n = 162) aged between 60–80 years who had completed the Cattell and Cattell Culture Fair Intelligence Test (CCF) completed the Cambridge Automated Neuropsychological Test Battery (CANTAB). The CANTAB has been administered to several patient populations and tests from the battery have been shown to be sensitive to damage in both temporal and prefrontal areas (Owen et al., 1996). Results from the test battery showed that both the Paired Associate Learning and Spatial Recognition tests were the most sensitive to normal ageing even when CCF is accounted for. In contrast, this performance on the “executive” tests, shown to be sensitive to frontal lobe damage, was not related to age, and CCF scores predicted performance on these tests. These results are discussed in relation to theories of cognitive ageing and patterns of change and in relation to several important methodological and theoretical considerations for the study of executive function.

Introduction

Neuropsychological and neuroanatomical evidence from brain imaging techniques suggest that changes accompanying old age occur earlier, and progress more rapidly in the pre-frontal areas which are known to be implicated in planning and “executive” function (Brody, 1994; Fuster, 1989; Madden & Hoffman, 1997; Parashos & Coffey, 1994; Raz, Gunning-Dixon, Head, Dupuis & Acker, 1997) and in the temporal cortex, hippocampus and limbic system, which are known to be implicated in memory and learning (Nagahara, Nicolle & Gallagher, 1993; Eustache, Rioux, Desgranges, Marchal, Petit-Taboue, Dary, Lechavalier & Baron, 1995). We might expect that these patterns of brain ageing would be reflected in correspondingly different time courses of behavioural changes in mental abilities. It might seem that an obvious way to test this is to give elderly and younger people batteries of neuropsychological tests which have been developed specifically to detect frontal, pre-frontal temporal and limbic changes in neuropsychiatric patients. Results have been ambiguous because of particular logical and methodological difficulties in developing and using tests of “frontal” and “pre-frontal” functions.

A classical methodology in neuropsychology has been to search for patients who have anatomically delimited lesions that are associated with losses of very specific cognitive abilities. Associations between focal damage and loss of performance are especially convincing when a function reduced by damage in one area is spared by damage in another adjacent area which, to maximise the neatness of the classical “double dissociation”, can be associated with loss of a different and equally specific function. The success of this enterprise has depended on two closely related research skills: clinical sensitivity to patients with unusual and sharply delimited patterns of functional loss, and ingenuity in developing tests that make demands only on these functions and not others. A well-established methodol-
ogy exists for ensuring that tests are reliable, valid and specific to the particular functions under investigation. To be reliable tests should rank order a group of individuals in the same way when they are administered more than once. To be valid, tests should consistently identify individuals with the same neuroanatomical deficits and individuals’ scores should correlate highly with scores on other, different tests that have proved sensitive to the same functions. To be specific, tests must be sensitive to changes in some functions but not in others. When bodies of reliable and valid tests of different specific functions have been assembled they may be incorporated into “Neuropsychological Test Batteries” that assess the kinds and degrees of specificity of functional losses and so allow diagnoses of neuroanatomical loci of damage. We might, therefore, hope that if a well-designed test battery is given to a large number of people patterns of associations and disassociations between scores on the component tests will reflect corresponding overlaps and distinctions between the underlying functional abilities on which they make demands.

In practice this works well for batteries of tests assessing perceptual, motor and some language functions but not for the more diffuse “higher cognitive” abilities that are affected by damage to the frontal and prefrontal cortex. As Burgess (1998) points out “frontal tests often can only be used once” because, once they cease to be novel most patients can perform them as well as intact controls. A possible explanation is that when complex novel tasks are first encountered they make heavy demands on planning and control and, because participants do not yet know which items of rapidly changing information are relevant and which are incidental, they have to try to attend to and update all that goes on and so experience severe demands on “working memory” (Baddeley, 1986). However after some practice tasks become “automatised”, variations in information processing load have much less effect and they are performed in qualitatively different ways; (Schneider & Shiffrin, 1977). The characteristic sensitivity of “frontal” tasks to practice makes it less likely that they will rank order a group in the same way when they are given more than once. Lowe and Rabbitt (1998) have shown that this is indeed the case and that some widely used frontal and executive tests have levels of test/re-test reliability lower than are considered acceptable for meaningful psychometric comparisons.

Another methodological problem for measures of frontal tests is contravention of the assumption of validity: i.e. that a test, which detects a particular change in one person, will also detect the same change in others. As Burgess (1998) comments, clinical experience shows that patients with well-defined pre-frontal lesions who fail on one “frontal” or “executive” task may perform well within the normal range on “any number of others”. This raises the wider question as to whether frontal lobes functions are best described in terms of distinct subsystems which can be assessed by different tasks that demand different, specific, functional capabilities such as “inhibition” (Hasher, Zacks & Rypni, 1991), “attention switching” (Baddeley, 1996), “goal maintenance” (Duncan, Burgess & Emmslie, 1995) or “working memory management” (Baddeley, 1986), or whether each of these putatively separate functions actually only represents a common language description of a particular set of task demands which, in fact, do not require distinct and specialised neuroanatomical sub-systems because they can all be met by an identical system architecture (Kimberg & Farrah, 1993; Rabbitt, 1998). In this second model individual differences in the efficiency of all of these apparently distinct functions can best be assessed by overall performance across a variety of different tasks which make different demands. This idea that individual differences in efficiency in coping with a wide range of different task demands can be best expressed in terms of differences in the magnitude of a single higher order statistical construct is reminiscent of Spearman’s (1927) proposal that performance on a wide range of different intelligence tests is well described by a single factor, g. Duncan et al. (1995) have taken this analogy further and shown that, even in groups of patients with focal frontal lobe damage, correlations between levels of performance on different tests of pre-frontal or executive function are actually very modest and may entirely disappear when variance associated within individual differences in intelligence test scores is taken into consideration. He has also found that the performance of frontal patients on a particular test of “goal neglect” were well predicted by scores on an intelligence test highly loaded for g (Duncan et al., 1995). This suggests that tests of frontal functions are not only similar to tests of g, in that they are measures of global and diffuse rather than of specific and localised skills, but that measures of general fluid mental ability such as intelligence tests are actually the best empirical measures of “frontality”. If re-stated in functional terms, rather than as a neutral observation of high statistical commonality between tests, this may amount to a suggestion that intelligence tests are the most sensitive indices of frontal integrity because while individual frontal tests only make demands on particular functions mediated by the frontal lobes, intelligence tests include such a wide variety of problems that they make simultaneous demands on all of them.

The generality of this hypothesis makes it attractive, but hard to validate. Scores on intelligence tests predict levels of performance on all kinds of different tasks, even on very simple tests of reaction time (Jensen, 1982; Rabbitt & Goward, 1994), tachistoscopic recognition thresholds (Nettbeck, 1981), and memory (Rabbitt & Yang, 1996). Thus, if frontal functions are to be identified with g, we must conclude that they are involved in all kinds of tasks, including very simple tests of information processing speed. This makes it hard to empirically ask whether as people grow old they tend to experience decrements of frontal and executive capabilities before losses of other functions. Elders’ performance on all tasks is well predicted by their intelligence
test scores and, moreover, in large populations of older people, the variance in performance with age between 50 and 90 years is also almost entirely picked up by the concomitant age-related variance in intelligence test scores (Rabbitt, 1993). Another influential idea has been that age-related changes in performance in all tasks, including memory tasks, can be explained as the result of slowing of mental speed, and predicted from performance on very easy reaction time tasks (Salthouse, 1992, 1993, 1996). There is also an influential body of opinion that individual differences in intelligence test performance are directly determined by corresponding differences in information processing speed (Eysenck, 1986; Jensen, 1985; Vernon, 1985). If age-related changes in intelligence test scores are good measures of age-related changes in frontal performance almost any tests that we may give old people, including very easy reaction time tests, may be efficient tests of frontal function. It then becomes both difficult, and uninteresting, to argue that age affects these functions earlier and more severely than others.

However, the promiscuity with which intelligence test scores correlate with scores on all others in normal adults contrasts with their more clear-cut differentiation between loci of brain damage in neuropsychological patients. The classical diagnostic criterion for the amnesic syndrome associated with focal damage to the medial-temporal lobes has been that, while they cause severe losses of recall and recognition, they bring about little or no change in measures of g\textsubscript{F}. This makes it useful to ask whether age-related declines in intelligence test scores (ITSs) that predict corresponding declines in performance on frontal tasks also predict performance on memory tasks. This would be good circumstantial evidence that age-related declines in memory differ from the specific changes seen in amnesia and that, as Salthouse, Fristoe and Rhee (1996) argue, they are caused by global changes that affect all other mental abilities. Alternatively, if age-related changes in memory tests are, at least to some extent, independent of changes in g\textsubscript{F}, we may argue that they are brought about by specific age-related changes in the temporal lobes, limbic system and hippocampus which are not, necessarily, in step with more general changes in the entire central nervous system.

Robbins, James, Owen, Sahakian, McInnes and Rabbitt (1994, 1998) have already described data on the Cambridge Automated Neuropsychological Test Battery (CANTAB) obtained from a large population of elderly volunteers. Their results suggested that on the Intradimensional Shift task, a computerised analogue of the Wisconsin Card Sorting Test, their oldest group, aged from 74 to 79 years, performed as poorly as neurological patients with clear-cut frontal lesions. However, there were no equivalent changes on performance on other tests that involve not only planning but also memory, such as the CANTAB version of the “Tower of London” (TOL) task and a “Spatial Working Memory” task. There were also no changes on other tests involving only immediate and delayed recall and recognition memory. We made a further study to replicate these earlier experiments, to further investigate the relationships between scores on frontal and temporal tests in the CANTAB battery, and to check the relationships between scores on both frontal and temporal tests and scores on a well-established measure of g\textsubscript{F}, the Cattell and Cattell (1960) “Culture Fair” (CCF) intelligence test.

Method

Research participants

Participants (n = 162) were recruited from the volunteer panel at the Age and Cognitive Performance Research Centre. The age range was from 60 to 80 years, with a mean age of 70 years (standard deviation 4.7 years). Individuals were selected on the basis of general ability level, as indicated by scores achieved on CCF test; this ensured that there was a wide range of abilities sampled across the age range. Volunteers with impaired colour vision or uncorrected visual acuity were excluded, along with those who were taking concomitant medication which might have interfered with “normal” cognitive function.

Equipment and procedure

The program was run on an IBM PC with a touch-sensitive screen. Two of the tests required a response key, used to facilitate the recording of reaction times.

All participants were tested in the same room using the same equipment. Due to the lengthy duration of the testing session (2–3 h depending on performance level), participants were offered a short break half way through the tests. Instructions for all of the tests were given verbatim from the CANTAB test manual. All participants completed the ‘motor screening’ test before beginning the test batteries. In this task 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 centre point of each cross. This ensured that they were physically able to use the touch screen and familiarised them with the response procedure before the tests began. All participants performed this task accurately.

Visual memory tests

Pattern recognition

This test is presented in two phases. In the presentation phase participants are shown a series of 12 simple, abstract, coloured patterns which appear one at a time inside a white box located in the centre of the screen. Each of the “target” patterns appeared for 3 s, the screen is cleared and the next pattern appears. In the recognition phase, 12 pairs of coloured patterns appear on the screen (one pair at a time) and the participant is required to respond to each pair by touching the pattern seen during the presentation phase. Each of the target patterns is paired with a distractor pattern that differs from it in form but not colour. Each response is accompanied by an auditory tone, and visual feedback is provided in the form of a tick, for correct response and a cross for an incorrect response. This procedure is repeated with 12 new patterns and the participants total score (maximum = 24) is automatically recorded.

Spatial recognition

This task is also presented in two phases. In the presentation phase participants are shown a series of five unfilled 1 in. white squares, appearing one at a time, for 3 s, at different locations on the screen.
The screen is cleared between the presentation of each square. In the recognition phase, two squares appear simultaneously on the screen and the participant has to indicate which of these locations had been occupied by a square during the first phase. The target squares are presented and paired with distractor squares which appear in novel locations. The feedback is identical to that given in the pattern recognition test. The procedure is repeated three times using new target and distractor locations. The total score (maximum = 20) is automatically recorded.

**Simultaneous and delayed matching to sample**

At the beginning of each trial a sample pattern appears in the centre of the screen for 4.5 s. The stimuli are complex, abstract patterns consisting of four quadrants, each differing in colour and form. Participants are asked to study the pattern, so they can later identify it among three distractor patterns. One of the distractors is novel differing from the sample in both colour and form. The remaining two distractors are “partial”, one differing from the target pattern in colour but not in form and the other in form but not colour. Feedback is identical to that given in the recognition tests. After making any correct responses, participants have to continue to choose until the correct stimuli is touched. There are three delay conditions in which the choice stimuli appear either 0, 4 or 12 s after the sample disappears. Ten trials are given in each of the four conditions which are presented in pseudo random order. Participants are scored according to the number of trials correct on the first choice made in each condition (maximum = 40).

**Paired associate learning**

This task tests conditional learning of pattern-locations associations. The location of each concealed pattern is marked by a white box on the screen. Set sizes increase in difficulty, beginning with two locations, then four, followed by six and finally eight. The boxes present on the screen open up one at a time in a randomised sequence, remain open for 3 s and then close. Each box, when open, displays a different pattern. Participants are instructed to remember the spatial location at which each pattern appeared so that they can indicate the correct position when cued by the pattern at a later stage. Each level of difficulty (set size) must be completed to 100% accuracy before moving on to the next. Feedback is not provided immediately after each response but only when all choices had been recorded. If any incorrect choices are made, the boxes are successively reopened for 2 s and participants are given a second chance to locate the patterns correctly. On each trial the participants are allowed up to nine reminding phases, before the test is terminated. When all choices are correct, the words “All Correct” appear in the centre of the screen. The test yields two separate scores; first trial memory score which indicates the amount of learning on first presentation across all levels (mean score) and the total number of trials taken to success (summed across all levels).

**Attention battery**

**Compound discrimination – intradimensional and extradimensional (1D:ED) set-shifting**

In the test four boxes appear on the screen and shapes appear in only two of them. The shapes are simple geometric forms and in the same colour. The test begins with a simple discrimination stage in which participants are told that they must touch the ‘correct’ shape until the rule changes. Touching a shape results in feedback of either a tick or a cross to enable the participant to learn which shape is correct. After six trials the rule reverses and the alternate shape becomes the correct choice. When this phase is successfully completed the test moves on to the compound discrimination phase, where irrelevant lines appear, initially adjacent to the shapes and then superimposed upon them. Throughout this phase the participant must learn via feedback that the line is irrelevant and that the rule concerns only the shape. When the irrelevance of the lines has been established one of the two different types of ‘shift’ take place. Initially there is an “intradimensional” shift, whereby novel exemplars of the stimuli are used, but shape remains the relevant dimension. When discrimination of the new stimuli has been established an “extradimensional” shift occurs, in which new stimuli appear and the line present on each display eventually becomes irrelevant, while the previously trained shape becomes relevant. It is this stage which is formally equivalent to a category shift in the classical test of frontal lobe function, the Wisconsin Card Sorting Task.

**Working memory battery**

**Spatial span**

This is a computerised version of the Corsi Block Tapping task (Milner, 1971). The sequence to be remembered is communicated by a series of ‘boxes’ becoming successively highlighted. The participant’s task is to tap the highlighted boxes, in turn, in the same order. After each such trial, the number of boxes in the sequence is increased by one to a maximum span of nine. After any incorrect attempt at replicating a sequence an alternative sequence of the same length is presented. If a participant fails at the same span length on three successive attempts the test terminates. The score is the maximum span attained.

**Spatial working memory – strategy formation**

Participants are required to ‘search through’ an array of boxes presented on the screen, by touching each one so that it ‘opens up’, revealing what is inside. The boxes appear in random spatial locations across the screen. The object of the task is to collect ‘blue tokens’ hidden inside the boxes, and once each is found, to add it to fill an empty column at the side of the screen. At any one time, there is only a single token inside one of the boxes, and participants are required to search the entire set of boxes until they find it. When one token is found and placed in the column another is then hidden, and the search process begins again. The key instruction is that once a token has been taken from a box, that box will never be used again. Thus, two kinds of search errors can be made: participants may return to open a box in which a blue counter has already been placed (a ‘false search error’); or they may fail to open a box in which a blue counter has been placed (a ‘within search error’). Secondly, a participant may return to a box already opened and shown to be empty earlier in the same search sequence (a ‘within search error’). Arrays of six or eight boxes were used to provide two levels of search difficulty.

It is possible for participants to discover a highly efficient search strategy for completing the task. This involves always starting the search with the same box, always searching the set of boxes in a fixed order and as soon as a token has been found, always returning to start each new search sequence with the same box. A strategy score was taken by recording the total number of sequences started with the same box, within each of the six and eight box arrays.

**Tower of London – planning task**

This is a computerised version of the TOL task devised by Shallice and McCarthy (Shallice, 1982), which uses coloured beads and wooden pegs to create the ‘problems’ to be solved. There are two parts to the test; a passive movement copying stage and an active planning and execution task. For both stages the display is the same. The screen is split, with a target display at the top and the test display beneath it. In the passive movement time task, the discs in the top display move sequentially and the participants are told to copy the moves using the discs in the bottom display. In the planning stage, the top display represents the target arrangement and participants are asked to achieve the same arrangement by moving the discs in the bottom display. They are asked to plan the sequence to copy the target in the minimum number of moves
possible. This target number of moves is displayed as a number beneath the main display. If an optimum move sequence is not made, the participant continues to try to solve the problem until a failure criterion is exceeded. This is the optimum number of moves multiplied by two, plus two. The problems increased in difficulty from two to five move solutions.

This task yields several measures of ‘planning’ performance: the number of trials completed with the optimum solution, the average number of moves taken in the ‘four-move’ problems and the number of moves taken in the ‘five-move’ problems.

Results

Age and visual memory tests

Linear regression analyses were performed for all of the Visual Memory tests with age as the predictor variable. Age alone significantly predicted performance on all except one of the visual memory tests. However, the proportion of variance predicted by age was very small in all but the Paired Associates test. The predictions by age are summarised below:

Spatial span – $F = 7.445$, $p < 0.01$, $r^2 = 0.044$.
Pattern recognition – $F = 7.222$, $p < 0.01$, $r^2 = 0.043$.
Spatial recognition – $F = 12.475$, $p < 0.001$, $r^2 = 0.072$.
Delayed match to sample (DMTS) – 0 s delay – $F = 4$, $p < 0.01$, $r^2 = 0.029$; 4-s delay – not significant;
12-s delay – $F = 5.468$, $p < 0.05$, $r^2 = 0.033$.
Paired associates – 1st memory score – $F = 16.828$, $p < 0.001$, $r^2 = 0.10$; total trials score – $F = 20.545$, $P < 0.001$, $r^2 = 0.114$.

Age and frontal tests

Linear regression analyses were performed for the three executive tests with age as the predictor variable. Age alone significantly predicted only one measure, the between-trial errors on the Spatial Working Memory test, ($F = 5.329$, $p < 0.05$, $r^2 = 0.032$). None of the measures from the TOL and ID/ED shift tasks approached significance.

Fluid IQ score and frontal tests

The CCF test score significantly predicted variance in performance on all except one of the memory measures, 0 s delay on DMTS. The predictions by CCF are summarised below:

Spatial span – $F = 32.822$, $p < 0.001$, $r^2 = 0.17$.
Pattern recognition – $F = 19.264$, $p < 0.001$, $r^2 = 0.107$.
Spatial recognition – $F = 13.785$, $p < 0.001$, $r^2 = 0.079$.
DMTS – 4-s delay – $F = 12.612$, $p < 0.001$, $r^2 = 0.073$;
12-s delay – $F = 21.897$, $p < 0.001$, $r^2 = 0.12$
Paired associates – 1st memory score – $F = 32.512$, $p < 0.001$, $r^2 = 0.169$; total trials – $F = 37.153$, $p < 0.001$, $r^2 = 0.188$.

Fluid IQ score and frontal tests

CCF test score significantly predicted variance in all measures of the TOL and Spatial Working Memory tasks, but none of the measures from the ID/ED shift task. The predictions are summarised below:

Spatial working memory

Total errors – $F = 24.461$, $p < 0.001$, $r^2 = 0.133$.
Between trial errors $F = 36.064$, $p < 0.001$, $r^2 = 0.184$.
Tower of London – trials in optimum solution – $F = 19.267$, $p < 0.001$, $r = 0.107$.

Four-move problems – $F = 6.91$, $p < 0.05$, $r^2 = 0.037$.
Five-move problem – $F = 6.609$, $p < 0.05$, $r^2 = 0.04$.

Age and CCF multiple regression model

Visual memory tests

Multiple regression analyses were performed with age and CCF as predictor variables to examine the significance of each variable when combined. The combined age and CCF model significantly predicted all variables, with the exception of the ID/ED shift measures. The result of interest was the prediction made by age when CCF score was included in the regression model. Age remained a significant predictor of measures from two tests: spatial recognition ($t = 2.744$, $p < 0.01$); paired associates, both 1st trial memory score ($t = 2.96$, $p < 0.01$) and total number of trials to success ($t = 3.43$, $p < 0.001$).

Frontal tests

Age no longer significantly predicted any of the variance in the Spatial Working Memory test (between trial errors) when CCF score was included in the model.

Several multiple regression analyses were performed to ascertain whether CCF score was the only measure which would account for the age-related variance seen in the linear regression results. If the CCF score was unique in accounting for the age-related variance, then this may be because of the shared relationship with a g factor. This was not the case; spatial span, spatial recognition and paired associates performed in the same way as CCF score when included with age in the regression model, predicting spatial working memory (i.e. age was no longer a significant predictor of performance).

In addition, like the CCF score, other visual memory scores (span and pattern recognition) did not account for age-related variance for performance on both spatial recognition and paired associates.

Between test relationships

Linear regression analyses were performed, with every test variable as predictors of all other tests to compare
the relationships between the Visual Memory and Executive tests with those, ‘within’ construct relationships. Correlation analysis showed that the ID/ED shift measures were not related to any other measure and were, therefore, excluded from the regression analyses. None of the visual memory tests were significant predictors of the TOL measures. Only spatial working memory (total errors) significantly predicted variance in these measures, with the exception of the five-move problems (optimum solution no. $F = 29.651$, $p < 0.001$, four-move $F = 17.574$, $p < 0.001$). In contrast to this, the visual memory tests (except DMTS) significantly predicted spatial working memory between trial scores; span ($F = 15.343$, $p < 0.001$), pattern recognition ($F = 9.835$, $p < 0.01$), spatial recognition ($F = 9.345$, $p < 0.01$) and paired associates (1st trial $F = 13.108$, $p < 0.001$; total trials $F = 9.278$, $p < 0.05$). In all cases, however, the memory measures predicted less than 6% of the variance in the spatial working memory measure.

Relationships between tests when CCF scores are taken into consideration

Stepwise linear regression analyses were performed to examine the significance of the predictions between tests after variance associated with CCF was accounted for. Some of the previously significant predictions were removed following this procedure. Spatial working memory measures did not significantly predict the TOL measures after CCF was accounted for. None of the visual memory tests predicted performance on the spatial working memory measures when CCF was included in the regression model. This was not the case for several of the predictions within the visual memory tests. Paired associate learning measures significantly predicted variance in spatial recognition ($t = 3.1$, $p < 0.01$), pattern recognition ($t = 6.62$, $p < 0.001$) and spatial span ($t = 2.2$, $p < 0.05$) after CCF was accounted for. In both the spatial recognition and spatial span, CCF accounted for only 2% of the variance, with paired associates reliably accounting for 11% and 17%, respectively, and did not reliably predict any additional variance in the pattern recognition task. Pattern recognition also reliably predicted DMTS (4- and 12-s delay) performance after CCF-related variance was removed ($t = 2.14$, $p < 0.05$ and $t = 3.49$, $p < 0.05$, respectively).

Principal components analysis

A principal components analysis was performed to examine the factor structure in relation to CCF and age loadings and potential separation of executive and visual memory factors. Only one measure from each test was entered into the analysis to avoid the strong intercorrelations between within test measures forcing each test onto separate factors. The variables were chosen on the basis of the relative strength of associations with other test measures. The rotation converged in 11 iterations and there was a three-factor solution with eigenvalues $> 1.00$, which accounted for 57% of the variance. Factor loadings above 0.3 are reported. Factor 1 is strongly associated with age and tests of visual memory and learning. Factor 2 is associated with executive test measures. Factor 3 is also associated with visual memory performance which is not related to age. CCF does not load strongly with any one factor, but loads equally on all three factors.

### Discussion

The main question asked was whether, in the CANTAB, the “frontal” tests are more sensitive measures of age-related change than the “temporal” tests? How far can age-related declines in scores on all tests in the CANTAB, and particularly on the three frontal tests, ID/ED shift, the TOL and the Spatial Working Memory tasks, be explained by concomitant declines in general fluid mental ability as assessed by the CCF? Are the effects of age on performance of the temporal memory tasks similar to those of amnesia in the sense that they are independent of concomitant declines in CCF scores? Do the effects of age on memory tests differ from those of amnesia because they can be entirely accounted for by age-related declines in CCF scores and so may reflect a “global” rather than a “local” change in brain efficiency? Are there strong mutual correlations within, but not between, scores on members of the sub-sets of frontal and temporal tests which might support the idea that each of these sets of tests involves different sets of cognitive processes? Do mutual associations within and between both sets of tests simply reflect their common determination by general fluid mental ability?

### Table 1

<table>
<thead>
<tr>
<th></th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Factor 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>-0.717</td>
<td>0.376</td>
<td>0.363</td>
</tr>
<tr>
<td>CCF</td>
<td>0.376</td>
<td>0.338</td>
<td>0.783</td>
</tr>
<tr>
<td>DMTS (4 s)</td>
<td></td>
<td></td>
<td>-0.594</td>
</tr>
<tr>
<td>Paired associates</td>
<td>-0.437</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spatial recognition</td>
<td>0.732</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pattern recognition</td>
<td></td>
<td>0.726</td>
<td></td>
</tr>
<tr>
<td>Spatial span</td>
<td>0.492</td>
<td>0.700</td>
<td></td>
</tr>
<tr>
<td>SWM (between search)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOL (4 move)</td>
<td>0.862</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Which of these neuropsychological tests are the most age sensitive?

Both neurological and behavioural evidence led us to expect that we would replicate the findings of Robbins...
et al. (1998), notably that the frontal ID/ED shift task and the TOL are more sensitive to brain ageing than any of the other CANTAB measures. However, in this sample neither ID/ED shift scores nor TOL scores correlated with age between 60 and 80 years. The only frontal task that showed a significant correlation with age was spatial working memory. This, at best, accounted for only 3% of age-related variance in performance (search error scores). When variance in CCF scores was taken into consideration this relationship disappeared.

It is clear from these correlational analyses that the attentional set-shifting task, the ID/ED task, conceptually identical to the Wisconsin Card Sorting test, is segregated from all other tests in the battery. Robbins et al. (1994, 1998) reported a similar finding from a principal components analysis of CANTAB scores which found that ID/ED scores loaded on a different factor from all of the other CANTAB tasks. We do not, however, replicate other finding of Robbins et al. that ID/ED shift scores were especially age sensitive.

The most age-sensitive tests in the CANTAB battery were memory tasks that previous studies with the CANTAB have found to be effective measures of temporal lobe integrity. Correlations between memory task scores and age were, overall, very modest. The most age-sensitive memory performance indices were those from the Paired Associates task which accounted for more than 10% of age-related variance between individuals.

No other memory test accounted for more than 1% of age-related variance. The factor structure from the principal components analysis would support the finding that the memory tests are most strongly associated with age.

How far can the age declines in CANTAB scores be explained by an age-related decline in $g_f$?

After CCF scores had been taken into account, no measures of frontal performance (i.e. indices from the TOL and the Spatial Working Memory tasks) significantly correlated with age. This replicates previous analyses of data from other frontal tests reported by Burgess (1998) and Duncan et al. (1995) and is, of course, congenial for the hypothesis of Duncan et al. (1995) that $g_f$, as assessed by CCF scores, may be the best available index of frontal efficiency in people of any age. It follows that these results can offer no encouragement for the idea that age-related changes in performance on frontal tasks are independent of concomitant changes in measures of the more global changes in cognitive ability indexed by $g_c$. In this particular population we found no evidence that the functional abilities required to perform frontal tasks declined faster than all of the other mental abilities assessed by the CCF.

The only tasks in which age-related variance was to some extent independent of variance in CCF scores were two putatively temporal tasks from the visual memory battery, paired associate learning and spatial recognition memory. This hint that memory tasks are particularly age sensitive, and that age-related changes in memory competence are not completely explained by concurrent changes in general mental ability is consistent with other findings. Rabbitt and Yang (1994) found that age-related declines in scores on verbal memory tasks were independent of declines in IQ test scores. Rabbitt and Pendleton (in preparation) have recently replicated and extended earlier observations by Rabbitt (1993) that as populations age there is a steadily increasing incidence of individuals whose scores on tests of learning and memory are more than two standard deviations below expectation from their scores on well-validated indices of $g_f$ such as the CCF and the Heim (1970) AH 4 (1) and AH 4 (2).

Are associations more robust within than between sets of frontal and temporal tests?

Scores on the ID/ED shift did not correlate with scores on any other task. However scores on the TOL measures correlated significantly with scores on the Spatial Working Memory task. Scores on all of the memory tasks correlate robustly with each other.

There is weak evidence that scores on the two sets of measures may reflect different functional processes since scores on the TOL task were not predicted by scores on any of the visual memory tests. However, scores on the Spatial Working Memory task were correlated with scores on all of the visual memory tasks, except DMTS. The factor structure lends some support for a degree of separation between the executive and memory test measures with the two executive tests loading on a separate factor to the memory tests which are split across two factors.

Do associations within and between sub-sets of frontal and temporal tests reflect mutual associations with CCF scores?

The significant relationships between the visual memory tests (not DMTS) and spatial working memory (between test type) were removed by controlling for CCF. In contrast, the prediction from TOL (except for five-move problems) of spatial working memory measures (within test type) remained significant after controlling for CCF. Most of the relationships between the visual memory test scores remained significant after variance in CCF scores had been taken into consideration.

**Interpretation**

These results add to considerable earlier evidence that the frontal and temporal tests in the CANTAB are
sensitive to changes in different sub-sets of functional processes. This differential specificity has been found in previous studies of patients with localised brain damage (see Owen, Sahakian, Semple, Polkey & Robbins, 1995; Owen, Morris, Sahakian, Polkey & Robbins, 1996). Some separation of factors incorporating scores from frontal and temporal measures has also been reported from other principal component analyses of data from community-resident elderly populations (Robbins et al., 1994, 1998).

Taken at face value these findings also offer some support for the suggestion of Duncan et al. (1995) that performance on frontal tests is highly related to fluid intelligence level, and that associations between measures of frontal integrity disappear when variance in gₐ is taken into consideration. However, the analyses provide no convincing evidence that tests of executive functions correlate more strongly with each other than they do with specific tests of other mental abilities such as memory, that are both unrelated in functional models of cognition or by commonality of effects from local lesions. The findings that measures of frontal integrity correlate with measures of memory performance, and that the levels of these associations are determined by general cognitive ability measures is also consistent with recent observations made by Baddeley (1996), Lehto (1996) and Robbins et al. (1998) and with growing evidence that executive and frontal lobe functions are heterogeneous, served by a variety of different neural structures and dependent on the integrity of pathways connecting the frontal lobes with many other brain regions. The implication is that frontal tests, by their nature, are less specific than other tests, are likely to incorporate memory functions and are also likely to be well predicted by measures of global cognitive competence.

Scores on temporal tests are also predicted by CCF scores, and just as strongly as are scores on the frontal measures. This again suggests that, in a neurologically intact population, memory efficiency and so perhaps temporal lobe efficiency in general, is to a large extent determined by “global” efficiency of the cognitive system. However, measures of temporal integrity differ from measures of frontal integrity in two crucial ways. First, associations between these measures are not entirely determined by their mutual associations with CCF scores. This argues that, in addition to their mutual dependence on global cognitive efficiency, visual memory test scores correlate with each other because they all, also, commonly reflect the “local” efficiency of a particular cognitive subsystem. Second, unlike age-related changes in frontal measures, age-related changes in memory are not completely explained by concomitant age-related declines in gₐ. Taken together with earlier evidence for the separation of intelligence test scores and scores on verbal memory tasks discussed by Rabbitt and Yang (1996), this again argues that in ageing populations the well-documented age-related decline in memory test scores (e.g. see reviews by Cohen, 1996; Craik & Jennings, 1992; and Rabbitt, 1998) is, to some extent independent of concomitant declines in gₐ. This gross statistical relationship is consistent with further evidence that as a population ages we observe an increased incidence of individuals in whom scores on memory tests decline more than two standard deviations below expectation from their intelligence test scores (Rabbitt, 1993; Rabbitt and Pendleton, in preparation). This general picture of strong associations between test scores in young and middle-aged adults, contrasting with increasing disassociations between test scores in older groups is consistent with the idea that in most individuals biological ageing affects all parts of the brain, and so all mental abilities, to a similar extent. While this process of “global brain ageing” continues, both the levels of scores across all kinds of cognitive tests and the patterns of relationships between scores on different kinds of cognitive tests remain unaltered and are strongly mediated by global cognitive efficiency, which is well indexed by measures of gₐ. However, as a population ages the incidence of focal brain changes in its individual members increases. These changes bring about disassociations between global and specific abilities in increasing numbers of individuals which become discernible as weak, but robust, statistical trends in large cohorts.

References


