Development of an Aspect of Executive Control: Development of the Abilities to Remember What I Said and to "Do as I Say, Not as I Do"

ADELE DIAMOND
COLLEEN TAYLOR
Department of Psychology
University of Pennsylvania
Philadelphia, Pennsylvania

Luria's tapping test (tap once when E taps twice, tap twice when E taps once) was administered to 160 children (80 males, 80 females) between 3½ to 7 years old. Older children were faster and more accurate than younger children, with most of the improvement occurring by the age of 6. All children tested demonstrated understanding of the instructions during the pretest, and most started out performing well, but younger subjects could not sustain this. Over the 16 trials, percentage of correct responses decreased, especially among younger subjects. Performance here was compared with performance on the day–night Stroop-like task. The most common error on both tasks was to comply with only one of the two rules. Other errors included tapping many times regardless of what the experimenter did and doing the same thing as the experimenter, rather than the opposite. It is suggested that the tapping task requires both the ability to hold two rules in mind and the ability to inhibit a strong response tendency, that these abilities improve between 3–6 years of age, and that this improvement may reflect important changes within frontal cortex during this period of life.

Introduction

The ability to hold two or more pieces of information in mind and at the same time inhibit a strong response tendency has been hypothesized to depend on dorsolateral prefrontal cortex (e.g., Diamond, 1988, 1991). For example, rhesus monkeys with lesions of dorsolateral prefrontal cortex have difficulty focusing on both the goal object and the box opening in a transparent barrier detour task termed "object retrieval"; they tend to ignore the opening, focusing exclusively on the goal ob-

Reprint requests should be sent to Adele Diamond, Department of Brain & Cognitive Sciences, MIT, Building E10-044, 77 Massachusetts Avenue, Cambridge, MA 02139, U.S.A. E-mail: diamond@psyche.mit.edu.

Received for publication 11 August 1995
Accepted for publication 22 December 1995

Developmental Psychobiology 29(4):315–334 (1996)
 CCC 0012-1630/96/040315-20

ject. They also have great difficulty inhibiting the pull to reach straight for the desired goal (e.g., Diamond, 1990). When the inhibitory requirement of the task is reduced, by using an opaque barrier, prefrontally lesioned macaques perform much better.

Patients with damage to frontal cortex often have difficulty when asked to compute the answer to a two-step problem in their heads, despite their ability to solve each of the steps individually (Barbizet, 1970; Luria, 1973). They can have difficulty when asked to do two things (such as clean the windshield and change the oil). They are inclined to focus on only one aspect of a story instead of on the story as a whole. Indeed, Goldstein (1936, 1944) considered the fundamental disorder caused by damage to the frontal lobe to be an "inability to grasp the entirety of a complex situation." Unilateral neglect and simultaneous extinction can also be observed in patients with frontal cortex damage; i.e., the patients notice only one stimulus when two stimuli are presented simultaneously and bilaterally (e.g., Damasio, Damasio, & Chui, 1980; Heilman & Valenstein, 1972).

Patients with damage to frontal cortex also have difficulty inhibiting a strong response tendency. For example, such a patient may automatically reach for a presented object, even if instructed not to do so, and even if the patient doesn't want the object: "Taking a pack of cigarettes, he hesitated a moment, then opened it and drew out a cigarette. He looked puzzled at it, being a nonsmoker" (L'Hermitte, 1983, p. 246). Multiple areas within the frontal lobe have been shown to have important inhibitory functions. For example, "disinhibited" social behavior (lacking the normal restraints dictated by social norms) is often seen in patients with damage to the orbital region of frontal cortex (e.g., Stuss & Benson, 1986). The ability to inhibit the normal tendency to say the word when one is reading (on the Stroop task), and instead to say the color of the ink in which the word appears, has been linked to the anterior cingulate cortex (Casey, Cohen, Noll, Forman, & Rapoport, 1993; Pardo, Pardo, Janer, & Raichle, 1990). The frontal eye fields and supplementary motor area appear to be critical for inhibiting the normal tendency to look toward a cue on the antisaccade task (Guitton, Buchtel, & Douglas, 1985; O'Driscoll et al., 1995). Patients with damage to premotor cortex have difficulty inhibiting an action once it has begun (e.g., when asked to squeeze an object once or twice, they repeat the squeezing action over and over again; Luria, 1966). Patients with damage to "nonmedial" frontal cortex have difficulty inhibiting a response on the go/no-go task, and are more impaired when required to make a response opposite to their natural inclination (press a blue key when a red light appears, and press a red key when shown a blue light) relative to their performance when asked to make the congruent response (press blue for blue, and red for red) than are other patient groups (Drewe, 1975).

Luria devised several tasks that require both the ability to hold two things in mind and the ability to exercise inhibitory control over one's behavior. One such task requires subjects to tap once when the experimenter has tapped twice and to tap twice when the experimenter has tapped once. Here, the subject must remember two rules and inhibit the natural tendency to mimic what the experimenter does. Not surprisingly, Luria (1966) found that patients with frontal-lobe damage were impaired on the task. Similarly, patients with damage to the frontal lobe have difficulty acting in accord with the instruction to raise their finger when the experimenter makes a fist and to make a fist when the experimenter raises his finger (Luria, 1966).

There is evidence that children improve in both the ability to attend to two things at the same time and the ability to exercise inhibitory control over their behavior during the preschool years. The period preceding mature acquisition of these abilities was termed "preoperational" by Piaget (Piaget & Inhelder, 1969). Evidence that children 3 or 4 years old have difficulty keeping two things in mind at the same time, or that they tend to focus on only one aspect of a problem, can be seen in (a) their failure on tests of liquid conservation (they fail to attend to both height and width, attending only to height), (b) their inability to compare an old idea with a new one and hence (c) their inability to work through a two-step problem without losing track of what they are doing, and (d) their seeming failure to understand the appearance–reality distinction (Flavell, 1993). By 5 or 6 years of age, children are capable of doing all these things. Also, children of 3 or 4 years, like patients with frontal cortex damage, often fail to perceive both touches if their face and hand are touched simultaneously. By 6–7 years, these sorts of errors occur only about half as often (Fink & Bender, 1953).

There is also evidence that preschool children have difficulty exercising inhibitory control over what they do. For example, in the delay of gratification paradigm, when faced with the choice of a smaller, immediate reward or a later, larger reward, children of 4 years are unable to inhibit going for the immediate reward although they would prefer the larger one. By 5–6 years of age, children are much better at waiting for the bigger reward (Mischel & Mischel, 1983). Similarly, on the windows task, where children are rewarded for pointing to a box which is visibly empty, and are *not* rewarded for pointing to a box in which they can see candy, 3-year-olds fail to inhibit the tendency to point to the baited box (Russell, Mauthner, Sharpe, & Tidswell, 1991). Children 3–4 years of age also tend to fail go/no-go tasks because they cannot inhibit responding. They appear to understand and remember the task instructions (e.g., they can verbalize the instructions), but they cannot get themselves to act accordingly. By 5–6 years, they succeed on these go/no-go tasks (Bell & Livesey, 1985; Livesey & Morgan, 1991). Instructed to sort a deck of cards first by one criterion and then by another, children of 3 years seem unable to inhibit sorting by the criterion that had previously been correct (Zelazo, Frye, Reznick, Schuster, & Argitis, 1995). As in the go/no-go paradigm, the problem does not appear to be forgetting, for Zelazo has done this with the experimenter reminding the child of the new sorting criterion at the outset of each trial, and indeed with the child explaining the new criterion to the experimenter at the outset of each trial, yet the child persists in acting in accord with the earlier rule (Zelazo, Frye, & Rapus, 1996). Children of roughly 3 years are slower to extinguish a response when the reinforcement has stopped than are children of roughly 5 years (Gladstone, 1969). Similarly, Luria (1959, 1961) reports that when children of 3–4 years are asked to press a balloon twice when a light appears, the children are unable to inhibit the motor program once it has begun and press the balloon many times in succession, whereas children about 5 years old have no difficulty complying with this instruction.

The fact that the ability to hold two things in mind and the ability to inhibit prepotent responses seems to improve between 3–6 years of age led us to predict that we would find an improvement between these ages in children's performance on Luria's tapping task, which requires both abilities. Since Luria first introduced this task over 30 years ago, it has been widely used in neurological assessments of frontal lobe damage in patients. It is our hope that by investigating the normal developmental progression of

Table 1
Demographic Information

Mean Number of Siblings	1.54
Mean Birth Weight [in lb (oz)]	7(13)
Percent of Subjects who were European Caucasian	95.60
Mean Age of Mother at Child's Birth (in years)	29.91
Mean Education Level of Mother (in years)	15.09
Percent of Mothers Working Since Child's Birth	46.88
Mean Age of Father at Child's Birth (in years)	32.65
Mean Education Level of Father (in years)	15.70

the performance of healthy children on this task, we are laying the groundwork for future neuropsychological studies of the maturation and integrity of the frontal lobes during early childhood.

Method

Subjects

The subjects were 160 full-term, healthy children from middle to upper-middle class homes. See Table 1 for information on their family background. Twenty children (10 male, 10 female) were tested at each age: 3½, 4, 4½, 5, 5½, 6, 6½, and 7 years (see Table 2). The range in age for children in the 3½-years age group was 3 years, 4 months to 3 years, 9 months. The age range for children in the 4-year group was 3 years, 10 months to 4 years, 2 months. Similar ranges were used for all age groups.

We had hoped to include children of 3 years, but discontinued testing children of this age because the task appeared to be too difficult for them. Most failed the pretest or could not be engaged in doing the task. Of the 15 children of 3 years we tried to test on the task, none provided usable sessions. In addition to the 160 children included in our analyses, we tried to test 30 others between 3½–7 years but were unable to use

Table 2
Subjects' Ages

Age (in years)	Mean Age [in weeks (days)]
3½	186 (3)
4	211 (3)
4½	238 (3)
5	263 (3)
5½	289 (3)
6	316 (3)
6½	342 (2)
7	361 (3)

Table 3
Unusable Subjects

Age	Number	Reason
3½	7 (6 male, 1 female)	wouldn't play[a]
3½	7 (2 male, 5 female)	did not pass pretest
4	5 (2 male, 3 female)	wouldn't play
4	5 (5 male, 0 female)	did not pass pretest
4½	3 (3 male, 0 female)	did not pass pretest
5	1 (0 male, 1 female)	wouldn't play
5	1 (1 male, 0 female)	did not pass pretest
5½	1 (0 male, 1 female)	wouldn't play

[a] The category "wouldn't play" includes those who tapped many times regardless of what the experimenter did as well as those who refused to tap at all.

their sessions. Fourteen of the subjects seemed to be playing their own game, rather than doing our task. Many of these 14 subjects tapped correctly in the beginning of the session, but then began tapping many times, regardless of what the experimenter did. Others refused to continue after the pretest. Sixteen of the subjects did not pass the pretest. All subjects whose sessions were unusable were under 6 years of age and most were under 5 (see Table 3). Significantly more of them were boys than girls, z scores, $p < .01$.

Ninety-three of the subjects on the tapping task were also tested on the day–night Stroop-like task earlier in the same session [at age 3½: 6 subjects (2 male, 4 female); at age 4: 12 subjects (5 male, 7 female); at age 4½: 7 subjects (4 male, 3 female); at age 5: 14 subjects (6 male, 8 female); at age 5½: 13 subjects (7 male, 6 female); at age 6: 12 subjects (8 male, 4 female); at age 6½: 13 subjects (9 male, 4 female); at age 7: 16 subjects (8 male, 8 female)].

Procedure

Each child was tested individually in a quiet room either at the laboratory of the University of Pennsylvania Infant and Early Childhood Study, or at the child's school or preschool. The experimenter and a back-up person sat across the table from the subject. Each session was recorded on videotape for detailed analysis.

The rules for the task were as follows: Immediately after the experimenter tapped once with a wooden dowel (22.5 cm long, 2.5 cm in diameter), the child was to tap twice with the dowel. Immediately after the experimenter tapped twice, the child was to tap once.

The experimenter instructed the child in the first rule thusly: "When I tap one time like this (Experimenter tapped once), I want you to tap two times like this (Experimenter tapped twice). Let's try that. When I tap one time (Experimenter tapped once), you tap. . . ." The experimenter handed the dowel to the child. If the child responded correctly, the experimenter praised the child and proceeded on to the second rule. If the child's response was incorrect, the experimenter explained and demonstrated the first rule again. If the child was correct the first time, or regardless of how the child performed the second time, the experimenter went on to explain and demonstrate the second rule in the same way. Again, the child received enthusiastic praise if correct, and correction if incorrect.

The experimenter then continued the pretest by tapping once and handing the dowel to the child for a response. If the child was correct, the experimenter praised the child, tapped twice, and handed the dowel to the child for a response. If the child was correct again, the child was praised, and these two trials counted as the first two trials of testing. If the child responded incorrectly on either of these trials, after the two trials were over, the experimenter reminded the child of both rules again, explaining first the rule the child had executed incorrectly. Then testing began.

Each session consisted of a pseudorandom series of 16 trials; each trial was composed of the experimenter's tap(s) and the subject's response. The experimenter tapped, then handed the dowel to the child for a response, after which the child returned the dowel to the experimenter. Experimenters were carefully trained to avoid influencing the child's response by reaching for the dowel too early or by leaving it with the child too long. For example, experimenters were trained not to reach too quickly after a response of one tap in case the child might tap again. We used only one dowel for both the experimenter and child so that neither the child nor experimenter would begin tapping before the other had finished. The series of the experimenter's tap was as follows: 1,2,2,1,2,2,1,1,1,2,1,2,2,1,1,2. No feedback was given during testing.

In order for testing to continue beyond the first two trials, the child had to be correct on each of the rules at least once over the course of practice plus Trials 1 and 2. That is, we needed evidence early on that the child understood what we were asking him or her to do in order for the session to count. We counted early practice trials, if answered correctly, as part of testing because children who readily understood the instructions became bored if given too much practice.

Response latency was coded from the videotape records. Latency was initially timed from the point at which the child had control of the wooden dowel to the point at which the child made his or her response. Latency was also calculated for half of the subjects (females only) from the point at which the experimenter finished tapping to the point at which the subject made his or her response. This second measure included the time taken to transfer the dowel. Three research technicians, blind to the experimental hypotheses, coded the videotapes for the initial latency measure. Intercoder reliability was $\alpha = .87$. One person did all of the videocoding for the second latency measure; her intracoder reliability was .91.

Our tapping task is similar in several ways to the day–night Stroop-like task (for a detailed description of procedure for the day–night task see Gerstadt, Hong, and Diamond, 1994). In the day–night task, subjects are shown a black card with stars and are asked to say "day," and a white card with a bright sun and asked to say "night." This is comparable to asking the subjects to tap twice when the experimenter taps once and to tap once when the experimenter taps twice. Both tasks require subjects to remember two rules and to inhibit their natural response. The day–night task, however, requires a verbal response to a visual stimulus, whereas the tapping task requires a motor response to a visual and auditory stimulus. Both tasks consisted of 16 trials presented in pseudorandom order. Response latency for the day–night task was timed from the point at which the child saw the card to the point at which the child responded. The second measure of response latency calculated for the tapping task is the most comparable to this in that here latency was timed from the point at which the experimenter finished tapping to the point when the child responded.

We predicted that the percentage of correct responses would increase over age, and that response latency would decrease over age. We expected to see deterioration in performance over the course of a session for the younger children but less of a change in performance over trials for the older children. More precisely, based on the results of Gerstadt et al. (1994) with the day–night Stroop-like test, we predicted that younger children would make more errors on later trials and that this would be accompanied by quicker responding on later trials by the younger subjects. The dependent measures were whether a response was correct or not, the percentage of correct responses over the course of a session, response latency for each trial, and the mean response latency over a session. The independent variables included between subject variables (age and sex), and a within subject variable (trial number).

Results

Performance on the Tapping Task

Older children performed significantly better on the tapping task than did younger children; linear regression of percentage of correct responses on age: $F(1,159) = 10.56$, $p < .0001$; see Table 4. The largest improvement in percentage of correct responses occurred between 3½ and 4 years of age, a difference of 17%; orthogonal contrast for performance at 3½ years versus 4 years: $F = 6.28$, $p < .02$; see Figure 1. No significant sex differences in percentage of correct responses were found at any age or overall, nor were there any significant Age × Sex interactions.

A significant decrease in response latency was found across age using both the original and the second measure of latency (regression: $F(1,159) = 2.98$, $p < .01$; $F(1,78) = 4.08$, $p < .001$; respectively; see Figures 2a and 2b).[1] The largest decrease in response time using Measure 1 occurred between 5½ and 6 years of age, 110 ms; orthogonal contrast: $F = 3.70$, $p = .06$. The largest decrease using Measure 2 occurred between the ages of 4½ and 5 years, 300 ms; orthogonal contrast: $F = 6.5$, $p = .02$. There were no significant sex differences in response latency at any age, nor were there any significant Age × Sex interactions.

Performance deteriorated over the course of a session. That is, children gave more incorrect responses on later trials than on earlier ones, repeated measures regression of accuracy of response on Trial Number (1–16): $F(1,2550) = 31.92$, $p < .0001$. This can also be seen in the higher percentage of correct responses on the first four trials than on the last four, 91% versus 83%; paired $t(1,159) = 4.57$, $p < .0001$; see Figure 3. There were no significant interactions between age and trial number on the percentage of correct responses; i.e., a similar pattern of deterioration in performance over trials was found at all ages. However, excluding the 3½-year-old subjects who performed poorly throughout, there was a significant decrease in the difference between the percentage of correct responses on the first four and last four trials with increasing age (repeated measures regression: $F(1,139) = 2.12$, $p < .05$). Thus, by this index, the decline in performance over the course of a session was more evident in younger children; older children were better able to sustain a high level of performance.

Over the course of a session, children began responding more quickly (repeated measures regression of response latency on trial number: $F(1,2550) = 7.48$, $p < .01$). There was no interaction between age and trial number on response latency and no sex difference in the main effect or interaction. Response time was significantly slower on the first four trials (mean latency$_1$ = 740 ms, mean latency$_2$ = 1760 ms) than on the last four trials (mean latency$_1$ = 660 ms, mean latency$_2$ = 1550 ms; paired $t_1(158) =$

Table 4

Performance on the Tapping Task by Age

Age in years	Mean Percent Correct	Mean Response $Latency_1$	Mean Response $Latency_2$	Mean Percent Correct on First Four Trials	Mean Percent Correct on Last Four Trials	Difference Between Percent Correct on the First and Last Four Trials	Response $Latency_1$ on the First Four Trials	Response $Latency_1$ on the Last Four Trials	Difference Between $Latency_1$ on the First and Last Four Trials	Number of Trials to Pass Pretest
3.5	64 (24.70)	0.80 (0.35)	2.11 (0.52)	72.50 (29.13)	66.25 (35.61)	6.0	0.75 (0.35)	0.85 (0.39)	−0.10	5.21 (1.36)
4	81 (17.53)	0.71 (0.22)	2.06 (0.28)	90.00 (18.85)	72.08 (30.26)	18.0	0.91 (0.62)	0.56 (0.18)	0.36	4.85 (1.09)
4.5	77 (21.66)	0.77 (0.24)	2.13 (0.25)	87.50 (20.68)	71.25 (28.42)	16.0	0.81 (0.35)	0.74 (0.29)	0.07	4.90 (1.48)
5	88 (10.37)	0.73 (0.25)	1.82 (0.28)	88.75 (18.98)	83.75 (18.63)	5.0	0.77 (0.42)	0.71 (0.26)	0.06	4.95 (1.15)
5.5	89 (16.30)	0.70 (0.20)	1.77 (0.26)	95.00 (13.08)	83.75 (18.63)	11.0	0.76 (0.32)	0.69 (0.35)	0.06	4.55 (0.94)
6	94 (14.26)	0.59 (0.15)	1.90 (0.26)	98.75 (5.59)	91.25 (24.70)	7.5	0.61 (0.17)	0.61 (0.23)	−0.01	4.50 (0.83)
6.5	97 (5.54)	0.61 (0.11)	1.75 (0.15)	97.50 (7.69)	96.25 (9.16)	1.0	0.67 (0.32)	0.58 (0.18)	0.09	4.20 (0.52)
7	98 (3.44)	0.57 (0.19)	1.59 (0.29)	100.00 (0.0)	97.50 (7.69)	3.0	0.61 (0.27)	0.56 (0.23)	0.05	4.10 (0.45)

Note. Mean Response $Latency_2$ was calculated only for half the subjects.
Numbers in parenthesis indicate standard deviations.

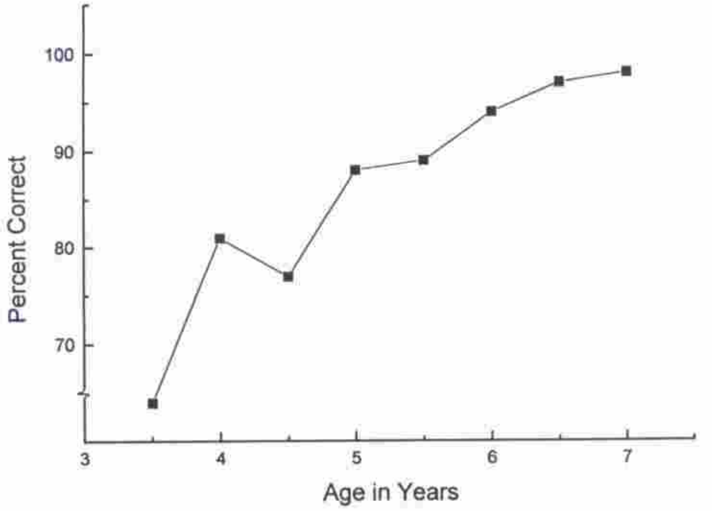

Fig. 1. Percentage of correct responses on the tapping task over age.

2.20, $p < .03$; paired $t_2(77) = 3.57$, $p < .001$; see Figure 4). In general, response latency was more uniform over the 16 trials of testing among older subjects. This age difference in the change in response latency (first four trials vs. last four trials) was statistically significant, repeated measures regression: $F(1,2552) = 33.29$, $p < .0001$.

There was no significant correlation between percentage of correct responses and response latency overall, $r = .02$, n.s. However, there was a marginally significant correlation between these two measures at 4½ and 5½ years of age, $r = .37$, $p = .10$, and $r = .42$, $p = .07$, respectively.

There was a significant decrease over age in the number of subjects who were not usable, regression: $F(1,192) = 14.62$, $p < .0001$, with a particularly large decrease between the ages of 4 and 4½ (10 subjects vs. 3 subjects; binomial < .05). Recall, also, that a number of subjects were found to be unusable at 3 years of age (15 subjects), which is why the 3-year age group was dropped from the study. The number of subjects who were unusable specifically because they failed the pretest decreased significantly with age, linear regression: $F(1,7) = 25.2$, $p < .002$. Older children who passed the pretest required fewer trials to do so than did younger children (regression of number of pretest trials on age: $F(1,159) = 4.26$, $p < .0003$; see Table 4). The largest decrease in the number of training trials occurred between the ages of 5 and 5½ years, orthogonal contrast: $F = 6.71$, $p = .01$.

The most common error on the tapping task was to always tap once, or always tap twice, regardless of what the experimenter did. It may be that subjects who made these errors were able to remember only one of the rules. In any case, 14 subjects between the ages of 3½ and 5½ years erred in this way.

Six children between the ages of 3½ and 4 mirrored what the experimenter did, rather than doing the opposite as instructed. This error was not seen in any subject older than 4 years. Three other children between 3½ and 4 years erred by tapping more than two times on several trials. This was clearly a great temptation for many of the other younger subjects; it was difficult for them to stop tapping after only one or two taps. Subjects who were not able to resist this temptation on any trial were declared "unusable" (see Table 3).

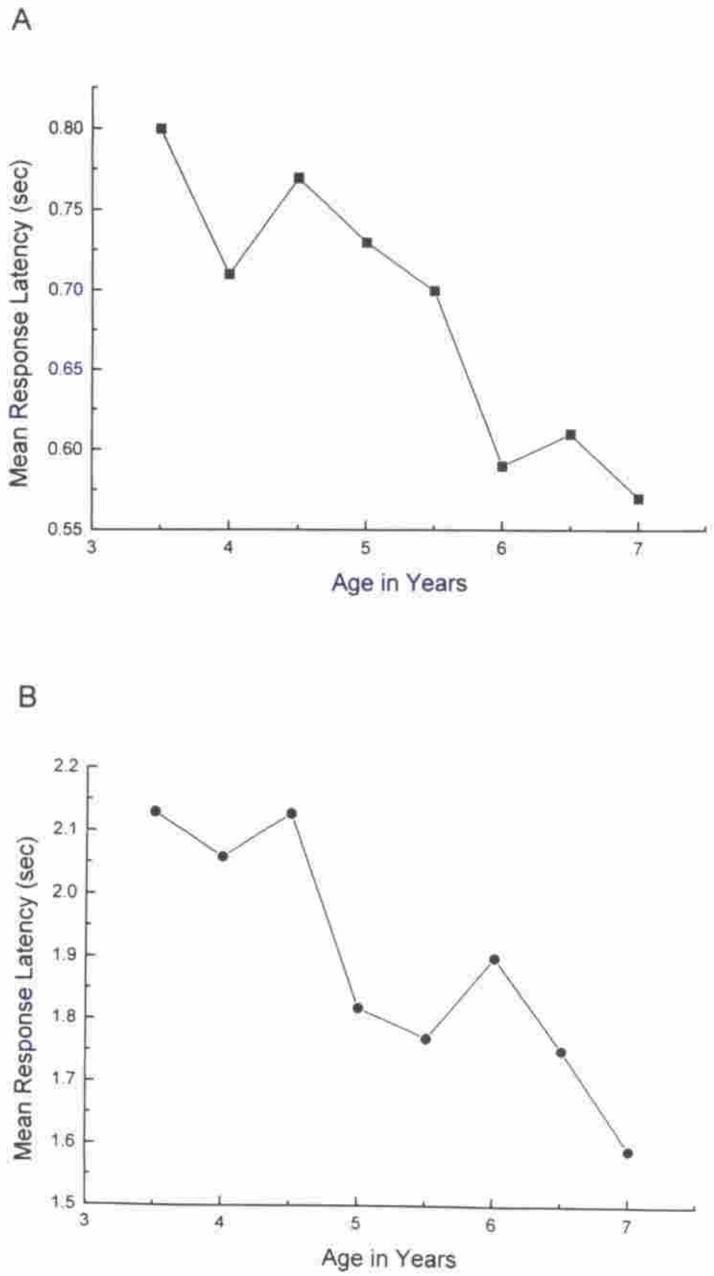

Fig. 2. (a) Mean response latency on the tapping task over age, using latency measure #1 (b) Mean response latency on the tapping task over age, using latency measure #2.

Results on the Tapping Task for the Subset of 93 Subjects Tested on Both the Tapping and Day–Night Stroop-Like Tasks

Performance on the tapping task improved significantly over age among the subsample of children also tested on the day–night Stroop-like test, regression of percentage of correct responses on age: $F(1,92) = 14.70$, $p < .0001$, as we had found for the sample

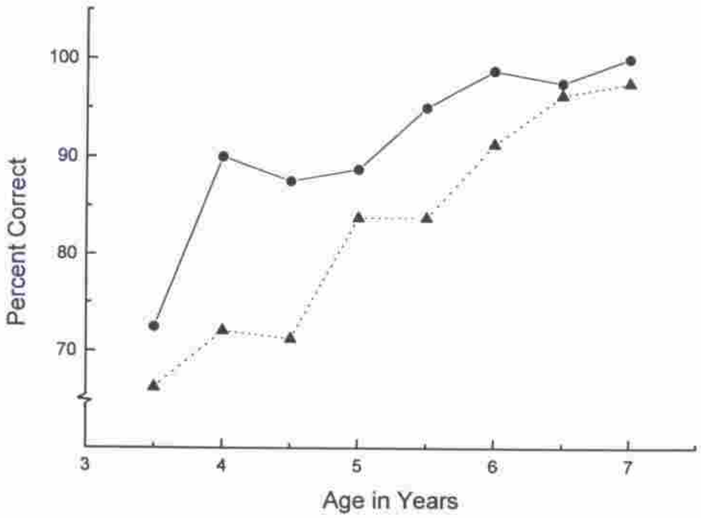

Fig. 3. Percentage of correct responses on the first four and last four trials on the tapping task over age. —●— = Performance on the first four trials. ----▲---- = Performance on the last four trials.

as a whole. However, the change in response latency over age, although in the same direction as for the sample as a whole, did not reach statistical significance within this smaller subsample of children, nor did the difference in performance on the first four trials versus performance on the last four trials. There were no significant differences between the performances of these 93 subjects and the other 67 subjects tested on the tapping task, and no differences in background demographic characteristics. Because the children tested on the day–night test before the tapping test performed comparably to the children tested on the tapping test alone, it is unlikely that prior experience with the day–night test had much effect on tapping test performance.

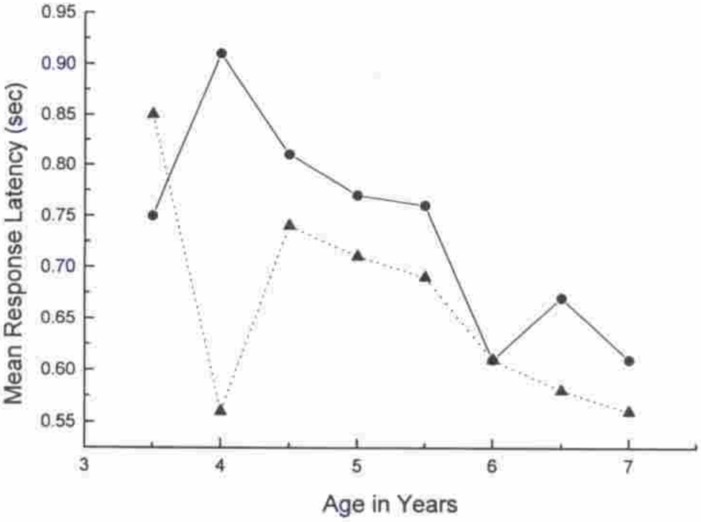

Fig. 4. Mean latency response on the first four and last four trials on the tapping task over age, using latency measure #1. —●— = Performance on the first four trials. ----▲---- = Performance on the last four trials.

Comparison of Performance on the Tapping and Day–Night Stroop-Like Tests for the 93 Subjects Tested on Both

The tapping task proved easier for children than the day–night Stroop-like task (see Table 5). Subjects were correct on significantly more trials on the tapping task than on the day–night task (regression with test as a within subject variable, and sex and age as between subject variables: F for test $(1,92) = 16.57$, $p < .0001$; see Figure 5). Percentage of correct responses was higher on the tapping task than on the day–night task at every age ≥4 years, with the difference being significant only at the individual ages 5½, 6, and 6½ years. The slopes of the Percent Correct × Age functions were similar on the two tasks beginning at age 4½; by 4½ years of age, children were beginning to perform better on the tapping task than on the day–night task and that difference was essentially maintained over the next 2½ years.

The difference between percentage of correct responses on the first four and the last four trials was significantly larger for the day–night Stroop-like test than for the tapping test (7% difference on tapping versus 18% difference on day–night, paired $t(93) = 2.84$, $p < .01$; see Figure 6). This was particularly true at the two youngest ages (3½ and 4 years of age). Children were better able to sustain good performance on the tapping task, while performance on the day–night task deteriorated more over the course of a session. Thus, this difference between performance on the day–night and tapping tasks was due primarily to the significant difference between percentage of correct responses on the last four trials, 85% correct for tapping versus 71% correct for day–night; paired $t(93) = 3.98$, $p < .0001$. Percentage of correct responses on the first four trials was not significantly different on the two tasks, indicating that subjects tended to perform well on both tasks in the beginning. Performance on the tapping task fell off somewhat over the course of a session, but performance fell off much more precipitously over the course of a session on the day–night Stroop-like task.

Response latency on the day–night task was timed beginning from when the child first saw the stimulus card. When latency was timed on the tapping task from the moment the child received the dowel to tap with, latencies tended to be faster on the tapping task than on the day–night task (regression with test (within subject), age and sex (between subject) in the equation: $F(1,91) = 2.99$, $p = .09$), although this difference was only significant among the 7-year-old subjects (paired $F(16) = 5.76$, $p < .05$). The difference in response latency between the first four and last four trials was significantly larger for the day–night test than for the tapping task, paired $t(88) = 2.40$, $p < .04$. This effect was due to the differences between performance on the two tasks in children ≤5 years of age (see Figure 7). The difference in the change in speed of responding on the two tasks decreased over age until it disappeared altogether at the age of 5½ years. Mean latency for the first four trials was significantly shorter on the tapping task than on the day–night Stroop-like task, $F(1,88) = 9.06$, $p < .004$. Mean latency on the last four trials was similar on the day–night and tapping tasks. There were no significant sex differences or interactions.

Before doing the previous analyses, we went back and coded half of the subjects (the girls) on a second latency measure for the tapping task, where timing began from the moment the experimenter stopped tapping, because we thought that might provide a latency measure more comparable to the one used for the day–night task. Using this second latency measure, children tended to be faster on the day–night task than on tapping, but neither the difference across ages nor at any individual age was large enough to reach statistical significance. Children tended to be somewhat faster on the

Table 5

Performance of the Subset of Children Tested on Both the Tapping and Day–Night Stroop-like Tests by Age

Age in Years	Mean Percent Correct	Mean Response $Latency_1$	Mean Percent Correct on First Four Trials	Mean Percent Correct on Last Four Trials	Difference Between Percent Correct on the First and Last Four Trials	Response $Latency_1$ on the First Four Trials	Response $Latency_1$ on the Last Four Trials	Difference Between $Latency_1$ on the First and Last Four Trials
3.5	57	1.36	58.33	62.50	−4.17	1.18	1.06	0.12
	69	*1.65*	*87.50*	*45.83*	*41.67*	*1.86*	*1.45*	*0.41*
4	79	1.37	85.42	66.67	18.75	1.19	1.01	0.18
	69	*1.63*	*89.58*	*54.17*	*35.41*	*1.98*	*1.32*	*0.66*
4.5	72	1.39	89.29	67.86	11.43	1.42	1.35	0.07
	63	*1.17*	*71.43*	*50.00*	*21.43*	*1.73*	*0.84*	*−0.11*
5	87	1.30	87.50	85.71	1.79	1.40	1.26	0.14
	78	*1.18*	*89.29*	*71.43*	*17.86*	*1.44*	*1.18*	*0.26*
5.5	94	1.20	96.15	88.46	7.69	1.29	1.27	0.02
	73	*1.32*	*80.77*	*65.38*	*15.39*	*1.33*	*1.17*	*0.16*
6	97	1.03	100.00	93.75	6.25	1.09	1.08	0.01
	90	*1.36*	*95.83*	*81.25*	*14.58*	*1.48*	*1.40*	*0.08*
6.5	97	1.00	98.08	96.15	1.93	1.10	1.00	0.10
	88	*1.18*	*94.23*	*80.77*	*13.46*	*1.32*	*1.13*	*0.19*
7	98	1.05	100.00	95.31	4.69	0.97	0.86	0.11
	94	*1.34*	*95.31*	*90.63*	*4.68*	*1.25*	*1.43*	*−0.18*

Tapping = values in regular typeface (top line).

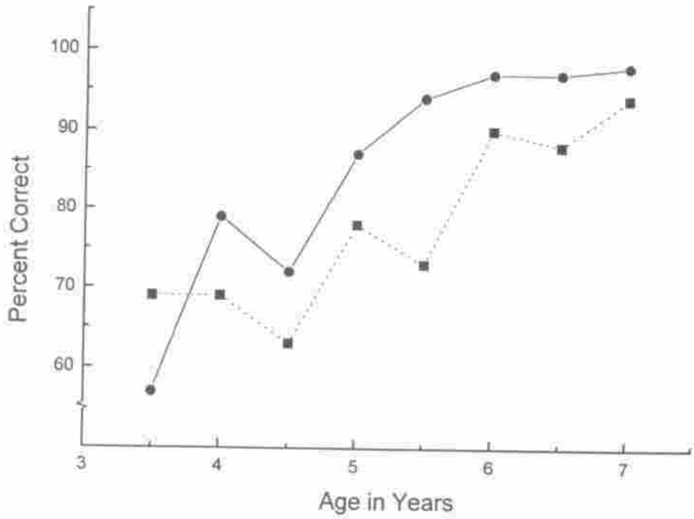

Fig. 5. Percentage of correct responses on the tapping and day–night Stroop-like tasks over age for the 93 subjects tested on both. —●— = Performance on the tapping task. ---■--- = Performance on the day–night Stroop-like task.

day–night task than on the tapping task on the first four trials, paired $t(43) = 2.01$, $p = .06$, but especially on the last four trials, paired $t(43) = 2.98$, $p < .01$. The increase in speed over trials tended to be larger on the day–night task than on the tapping task, paired $t(43) = 1.87$, $p = .08$, but this difference was not statistically significant.

Because only girls were included in the analyses using the second latency measure, we also compared response time on the day–night and tapping tasks using the first latency measure with only girls in the analyses. All results were the same as those found when both males and females had been included in the analyses.

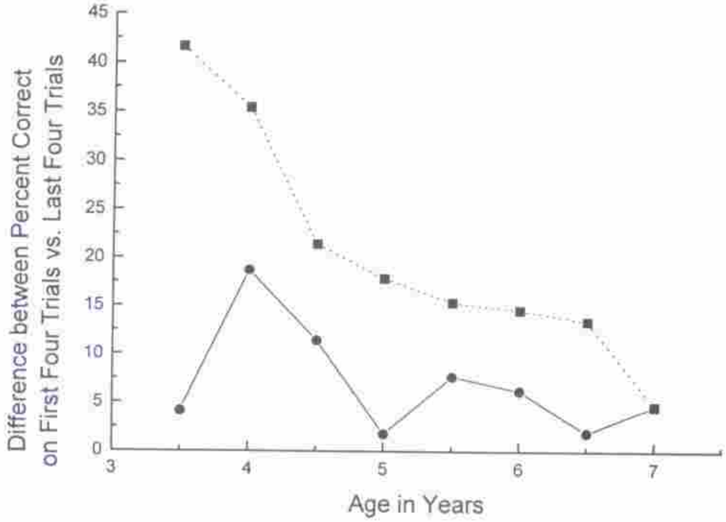

Fig. 6. Percentage of correct responses on the first four minus the last four trials on the tapping and day–night tasks over age for the 93 subjects tested on both. Performance —●— = Change in performance on the tapping task. ---■--- = Change in performance on the day–night Stroop-like task.

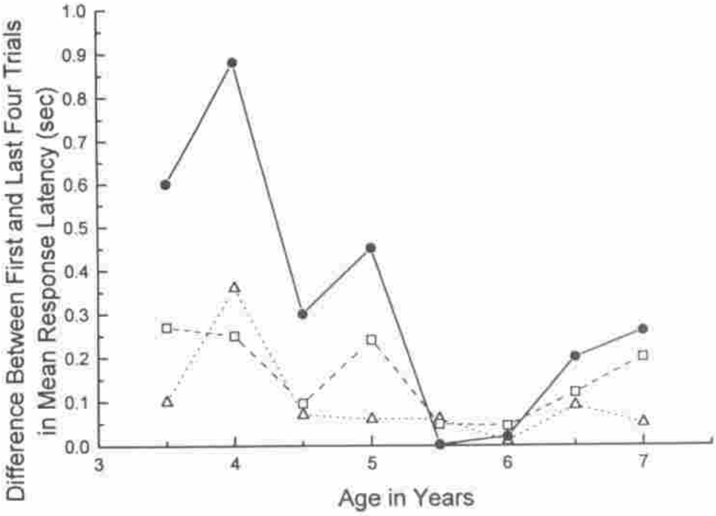

Fig. 7. Difference between mean latency response on the first four and last four trials on the tapping and day–night tasks using both measures 1 and 2 for tapping. ----△---- = Change in performance on the tapping task using latency measure #1. – –□– – = Change in performance on the tapping task using latency measure #2. —●— = Change in performance on the day–night Stroop-like task.

There was no significant difference in the number of unusable subjects on the two tests. However, for the subjects who were usable, significantly more trials were required to pass the pretest on the tapping task than for the day–night Stroop-like task, $t(8) = 19.29$, $p < .0001$. Luria (1961, 1966) reports that adults with frontal lobe damage and young children could often say the correct answer, although they could not demonstrate it in their behavior. Perhaps if the pretest on the tapping task had required subjects to say the correct response instead of demonstrating it, children would not have required more trials on the tapping pretest than they did on the verbal day–night pretest. Luria, Pribram, and Homskaya (1964) also reported that although the frontal patient they studied failed the tapping test, with long practice on each rule she could perform well. Perhaps the relatively long practice that some of our children received on the tapping test may have minimized the age-related improvements on the task and may have helped to make the tapping task appear a bit easier for the children than the day–night task.

Discussion

Over the period of 3½–7 years, children improved in both accuracy and speed on the tapping task, most of the improvement occurring by 6 years. Throughout this age range, children were generally correct on more of the trials earlier in a session than later, with the deterioration in performance from the first four to the last four trials being more evident among the younger children. Across the age range, children responded more slowly on earlier trials, speeding up as the session progressed. With age, fewer subjects failed the pretest and the subjects who passed the pretest required fewer trials to do so. This age-related improvement in performance is consistent with evidence from other studies, summarized in the Introduction, that children improve in their

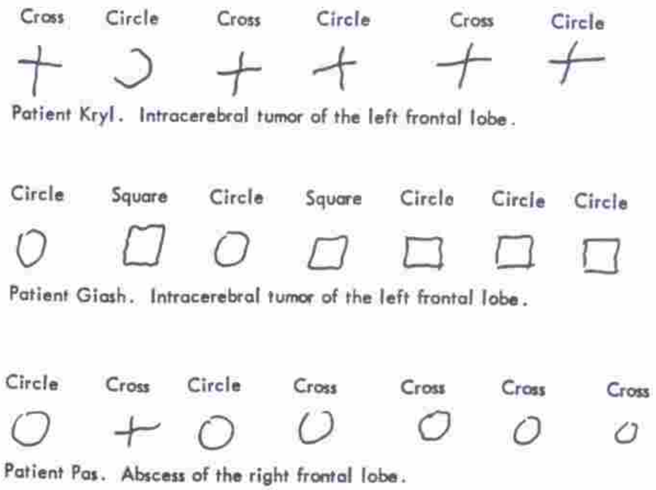

Fig. 8. Performance of subjects with extensive lesions to the frontal lobe when asked to alternately draw two simple shapes (from Luria, 1966). Note how the subjects revert to drawing only one of the instructed shapes. Luria referred to this as "pathological inertia of action."

ability to hold two things in mind and in their ability to inhibit a strong response tendency between 3–6 years of age. These findings are also consistent with the results from other studies of the performance of children 6–12 years of age on the tapping task which found that, at least among Caucasian subjects, most children were already performing at or near ceiling by the age of 6 (Becker, Issac, & Hynd, 1987; Passler, Issac, & Hynd, 1985).

The most common error we found was for a child to always tap once, or always tap twice, regardless of what the experimenter did. This error is reminiscent of a characteristic error Luria (1966) observed in patients with extensive lesions of the frontal lobe. For example, when asked to alternately draw a circle and a cross or a circle and a square, the patients (like most of even our youngest children) started off performing correctly, but like many of our younger children, soon deteriorated into following only one of the rules (e.g., drawing only circles, only crosses, or only squares; see Figure 8).

It is possible that this reflects a memory problem. It may be that children who made these errors were able to remember only one of the two rules. It may have been that these subjects forgot which rule went with which experimenter action or forgot what the other rule was. Or, it may be that subjects lacked the ability to flexibly switch between the two rules although they remembered both. In any case, this kind of error may be conceived of as focusing exclusively on one of the rules, ignoring the other, and ignoring the contextual nature of the instructions (If E taps once, you tap twice; if E taps twice, you tap once.) Older children may perform better on the task because their memories are better, because they are more likely to use the strategy of reducing the two rules to one (do the opposite of what the experimenter does), or because they have more control over their behavior and so can more flexibly switch between two actions. It has been noted by others that between 3–6 years of age, children become more likely to use strategies to aid their performance (e.g., Luria, 1961; Mischel & Mischel, 1983).

Other errors by the children seem more clearly to reflect inadequate inhibitory control. One common error among the younger children was the inability to resist tapping many times instead of just once or twice. Again, this error is reminiscent of behavior Luria noted in patients with excessive damage to the frontal lobe: [When asked] to tap three times or to squeeze the doctor's hand three times . . . although the patient retains the verbal instruction and repeats it correctly, he taps many times or squeezes the doctor's hand five, six, or more times instead of three" (Luria, 1966, p. 252).

Another error was for the child to match what the experimenter did instead of doing the reverse. This may reflect an inhibitory failure, as a subject's first inclination is to mimic the experimenter's action instead of doing the mental calculation that two means one and one means two (i.e., one tap by the experimenter means two taps are called for by me). Becker et al. (1987) reported that African–American children who could not yet perform well on the tapping task performed splendidly when instructed to match the experimenter's actions (Tap once when E taps once, tap twice when E taps twice.) Older children may perform better on the task because they are better able to inhibit response tendencies that would be incorrect in this context.

Luria (1966; Luria et al., 1964) has described extensively such "echopractic" errors in frontal-lobe patients. Indeed, on the tapping task itself, Luria found that although the patients (like our children) could correctly comply with the instructions for a short while, they very soon began to imitate the experimenter's movements, although Luria (like us) had evidence that the subjects had understood the instructions. Indeed, Luria found that the patients could verbalize the rules even as they failed to act in accord with them.

Another task that also requires (a) remembering two rules and (b) inhibiting the response you were inclined to make, making the opposite response instead, is the day–night Stroop-like task. Both tasks are tested over a 16-trial sequence. On both it is possible to reduce the two rules to one: Do, or say, the opposite. The tasks differ, however, in the response modality required—subjects must say the answer on the day–night task; subjects are to enact the answer on the tapping task. The inhibitory requirement is probably less demanding on the tapping task, as subjects' tendency to mimic what the experimenter does is probably not as strong as our tendency to associate the sun with day, and the moon and stars with night. Although the tapping task proved easier for the children (perhaps because of the less-severe inhibitory demand or the longer practice), there were strong similarities in children's performance on the two tasks.

Children were better able to sustain a high level of performance over the course of a session when tested on the tapping task. There was no difference in performance on the two tasks early in a session, but performance on the day–night task fell off more steeply on later trials. The slopes of the developmental progressions on both tasks were comparable: By 4–4½ years of age, children were beginning to perform better on the tapping task; thereafter, improvement on the two tasks proceeded roughly in parallel. Passler et al. (1985), who administered the tapping task and a task similar to our day–night task to children 6–12 years of age, also found that excellent performance on the tapping task appeared at a younger age than on their Stroop-like task.

When we timed response latency on the tapping task from the moment the child received the dowel, latencies tended to be shorter on the tapping task than on the day–night task. When latency was timed from the moment the experimenter stopped tapping, there was a slight tendency for response times to appear faster for the day–night task, although differences between response latencies on the two tasks were even smaller here than when the first latency measure was used. Subjects tended to respond more quickly as a session progressed. This was true on both tasks, but it was especially so on the day–night test for subjects ≤5 years old.

Because subjects made fewer correct responses as a session progressed, the decrease in response time may be more indicative of an inability to sustain a high level of effort over the course of the 16 trials than an indication that the task was becoming easier with practice. According to this line of reasoning, children may have responded more quickly later in a session because they were too cognitively exhausted to figure out the correct answer, so they stopped trying, just did anything, and hence could respond more quickly. Alternatively, the decrease in response time as a session progressed may be indicative of a lack of control over one's behavior manifested as impatience or as an inability to sustain waiting-to-respond until one has figured out the correct answer. In this case, the short response latency would be the cause of the drop in response accuracy (i.e., Children stopped giving themselves enough time to figure out the correct response.) Robinson reports something like this in frontal lobotomy patients administered a task designed by Downey on which subjects are instructed to write "United States of America" as slowly as possible: "[Frontal lobotomy patients] seemed, as always, to be making an effort to do as they were asked, but their hands would move slowly for a few seconds only and then would resume normal speed. Apparently they could not slow themselves down" (Robinson, 1946, p. 431).

Important changes in executive control abilities appear to be occurring between 3–7 years of age. Developmental improvements on the tapping and day–night tests appear to capture those changes well. The tapping test has been empirically linked to frontal cortex and we have hypothesized that the day–night test requires the functions of frontal cortex as well (Gerstadt et al., 1994). Much of the work on the tapping test comes from old studies with patients with massive damage, however. It is not clear from such studies which regions within frontal cortex are critical for the task, or even whether the cortex, rather than the basal ganglia, is the critical site. Indeed, memory and inhibitory abilities have also been linked to the medial temporal lobe. What role, if any, structures such as the hippocampus might play in performance on the tapping task remains to be investigated. We hope that the evidence provided here on the normal developmental progression of performance on the task will help in future studies of the neural basis of performance on the tapping task in young children.

Notes

This work was funded by NIH Grant #R01 MH41842 and BRSG grant #S07 RR07083. We gratefully acknowledge the assistance of Glenda Callender and Weimin Li in running the SAS programs for the data analysis. Glenda also helped graph the results. We thank the children, parents, and teachers who helped make this study possible.

Footnote

[1]Original measure calculated the child's time to respond from the moment he or she was handed the dowel. The second measure, calculated for only half the subjects at each age, was based on the child's time to respond from the moment the experimenter stopped tapping.

References

Barbizet, J. (1970). Prolonged organic amnesias. In J. Barbizet (Ed.), *Human memory and its pathology* (pp. 25–93). San Francisco: W. H. Freeman & Co.

Becker, M. G., Issac, W., & Hynd, G. W. (1987). Neuropsychological development of nonverbal behaviors attributed to "frontal lobe" functioning. *Developmental Neuropsychology, 3*, 275–298.

Bell, J. A., & Livesey, P. J. (1985). Cue significance and response regulation in 3- to 6-year-old children's learning of multiple choice discrimination tasks. *Developmental Psychobiology, 18*, 229–245.

Casey, B. J., Cohen, J. D., Noll, D. C., Forman, S., & Rapoport, J. L. (1993). Activation of the anterior cingulate during the Stroop conflict paradigm using functional MRI. *Society for Neuroscience Abstracts, 19*, 1285.

Damasio, A. R., Damasio, H., & Chui, H. C. (1980). Neglect following damage to frontal lobe or basal ganglia. *Neuropsychologia, 18*, 123–132.

Diamond, A. (1988). The abilities and neural mechanisms underlying A$\overline{B}$ performance. *Child Development, 59*, 523–527.

Diamond, A. (1990). Developmental time course in infants and infant monkeys, and the neural bases of inhibitory control in reaching. In A. Diamond, (Ed.), *The development and neural bases of higher cognitive functions*. New York: *Annals of the New York Academy of Sciences, 608*, 637–676.

Diamond, A. (1991). Neuropsychological insights into the meaning of object concept development. In S. Carey & R. Gelman (Eds.), *The epigenesis of mind: Essays on biology and knowledge* (pp. 67–110). Hillsdale, NJ: Erlbaum.

Drewe, E. A. (1975). An experimental investigation of Luria's theory on the effects of frontal lobe lesions in man. *Neuropsychologia, 13*, 421–429.

Fink, M., & Bender, M. B. (1953). Perception of simultaneous tactile stimuli in normal children. *Neurology, 3*, 27–34.

Flavell, J. H. (1993). Young children's understanding of thinking and consciousness. *Current Directions in Psychological Science, 2*, 40–43.

Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of children 3½–7 years old on a Stroop-like day–night test. *Cognition, 53*, 129–153.

Gladstone, R. (1969). Age, cognitive control, and extinction. *Journal of Experimental Child Psychology, 7*, 31–35.

Goldstein, K. (1936). The modification of behavior consequent to cerebral lesions. *Psychiatric Quarterly, 10*, 586–610.

Goldstein, K. (1944). The mental changes due to frontal lobe damage. *Journal of Psychology, 17*, 187–208.

Guitton, D., Buchtel, H. A., & Douglas, R. M. (1985). Frontal lobe lesions in man cause difficulties in suppressing reflexive glances and in generating goal-directed saccades. *Experimental Brain Research, 58*, 455–472.

Heilman, K. M., & Valenstein, E. (1972). Frontal lobe neglect in man. *Neurology, 22*, 660–664.

L'Hermitte, F. (1983). "Utilization behaviour" and its relation to lesions of the frontal lobes. *Brain, 106*, 237–255.

Livesey, D. J., & Morgan, G. A. (1991). The development of response inhibition in 4- and 5-year-old children. *Australian Journal of Psychology, 43*, 133–137.

Luria, A. R. (1959). The directive function of speech in development and dissolution. *Word, 15*, 341–352.

Luria, A. R. (1961). *The role of speech in the regulation of normal and abnormal behavior*. New York: Liveright Publishing Corporation.

Luria, A. R. (1966). *Higher cortical functions in man*. New York: Basic Books.

Luria, A. R. (1973). *The working brain* (B. Haigh, Trans.) (Original work published 1969). New York: Basic Books.

Luria, A. R., Pribram, K. H., & Homskaya, E. D. (1964). An experimental analysis of behavioral disturbance produced by a left frontal arachnoidal endothelioma (meningioma). *Neuropsychologia, 2*, 257–280.

Mischel, H. N., & Mischel, W. (1983). The development of children's knowledge of self-control strategies. *Child Development, 54*, 603–619.

O'Driscoll, G. A., Alpert, N. M., Matthysse, S. W., Levy, D. L., Rauch, S. L., & Holzman, P. S. (1995). Functional neuroanatomy of antisaccade eye movements investigated with positron emission tomography. *Proceedings of the National Academy of Sciences, 92*, 925–929.

Pardo, J. V., Pardo, P. J., Janer, K. W., & Raichle, M. E. (1990). The anterior cingulate cortex mediates processing selection in the Stroop attentional conflict paradigm. *Proceedings of the National Academy of Sciences, 87*, 256–259.

Passler, P. A., Issac, W., & Hynd, G. W. (1985). Neuropsychological development of behavior attributed to frontal lobe functioning in children. *Developmental Neuropsychology, 4*, 349–370.
Piaget, J., & Inhelder, B. (1969). *The psychology of the child*. New York: Basic Books.
Robinson, M. F. (1946). What price lobotomy? *Journal of Abnormal and Social Psychology, 41*, 421–434.
Russell, J., Mauthner, N., Sharpe, S., & Tidswell, T. (1991). The 'windows task' as a measure of strategic deception in preschoolers and autistic subjects. *British Journal of Developmental Psychology, 9*, 101–119.
Stuss, D. T., & Benson, D. F. (1986). *The frontal lobes*. New York: Raven Press.
Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development, 11*, 37–63.
Zelazo, P. D., Frye, D., Reznick, J. S., Schuster, B. V., & Argitis, G. (1995). *Age related changes in the execution of explicit rules*. Manuscript submitted for publication.