There is convincing evidence that historical changes in the physical and social environment can affect the level of different types of cognitive functioning. The phenomenon that cognitive test scores increase across generations in comparisons of people of the same age who are tested at different times is now so well established that it has been termed the Flynn Effect, after one of the first researchers to document its existence (Flynn, 1984, 1987).

The factors responsible for historical increases in cognitive performance have not yet been identified, and consequently many developmental researchers have relied on year of birth as the basis for determining cohort membership. Furthermore, age-related differences in cognitive functioning are sometimes assumed to be primarily attributable to cohort differences, which implies that differences between birth cohorts should be much larger than differences within birth cohorts. Comparisons of composite scores for five cognitive abilities in different people tested at different ages in different years revealed that within-cohort differences across ages were often as large as between-cohort differences across ages. These results lead to questions about the practice of relying on birth cohort to represent influences on cognitive functioning associated with temporal shifts in characteristics of the environment.

Keywords: cohort differences, cross-sectional, aging, cognitive ability, intelligence
60 years of age, and alternative interpretations of the results have been proposed by other researchers. For example, Horn and Donaldson (1976), who referred to the comparison of independent samples of people of different ages who are tested at different times as quasi-longitudinal, suggested that the age-related trends in the within-cohort data were very similar to those in the cross-sectional analyses. Furthermore, secondary analyses of Schaie’s data, reported in Salthouse (1991), revealed that the age-related variations in independent samples from the same birth cohort were nearly identical to those observed in conventional between-cohort, cross-sectional comparisons.

More recently, Ronnlund and his colleagues (Ronnlund & Nilsson, 2006; Ronnlund, Nyberg, Backman, & Nilsson, 2005) described a project in which the same cognitive tests were administered to independent samples of adults across a wide age range in 1989 and in 1994. Results were reported from a spatial (block-design) test, and for factor scores representing episodic memory (based on five memory tests) and semantic memory (based on a vocabulary test, a general-knowledge test, and three verbal-fluency tests). Comparisons of the tabular data reported in the articles indicate that in the case of the block-design variable, the within-cohort differences were nearly identical to the between-cohort differences at every age. The patterns of results for the episodic memory and semantic memory factors were less consistent, as the within-cohort contrasts were more positive than the between-cohort contrasts for participants in their 40s and 50s in the case of episodic memory, and for participants in their 40s, 50s, and 60s in the case of semantic memory. As in Schaie’s data (Schaie et al., 1973), the patterns of age-related differences were nearly identical in the within-cohort and between-cohort comparisons for adults older than about 60 years of age.

In summary, prior research with independent-samples within-cohort comparisons of age-related differences in cognitive functioning has been inconsistent, in that the same results have been interpreted differently, and different patterns have been reported for different variables within the same project, and for the same variable at different periods of adulthood. Several factors might account for the inconsistencies in these comparisons. One issue is that the samples in different years may not have been comparable, in which case the within-cohort contrasts could have been confounded with sample selectivity. Because even random sampling can result in fluctuations across test years, it is desirable for sample comparability to be directly assessed rather than merely assumed.

A second issue regarding prior research is that the boundaries between birth cohorts are arbitrary, and may not correspond to salient physical or social changes. In the studies by Schaie (Schaie et al., 1973; Schaie & Strother, 1968) and Ronnlund (Ronnlund & Nilsson, 2006; Ronnlund et al., 2005), individuals were grouped into 7-year and 5-year birth cohorts, respectively, presumably because those ranges corresponded to the intervals between test occasions. However, because the relations of birth year to relevant environmental changes are currently unknown, valuable information may be lost by imposing discrete boundaries on a continuous variable.

The current study incorporated these considerations in an examination of between-cohort and within-cohort age-related trends in measures of cognitive functioning in participants from the Virginia Cognitive Aging Project (Salthouse, 2004, 2010a, 2011). The same tests have been administered to new (and returning) samples of adults across a wide age range between 2001 and 2011, and therefore cognitive functioning could be examined as a function of both birth year (within cohort) and test year (between cohort).

The relevant contrasts are illustrated in Figure 1, which portrays a subset of the sample in the current study. Notice that people from different birth years (horizontal axis) were tested in different years (vertical axis). The highlighted row, corresponding to a traditional cross-sectional comparison, consists of people who were born in different years and tested in the same year (i.e., 2006). Because the testing extended over multiple years, it was also possible to compare the performance of people who were born in the same year but tested in different years, as represented by the highlighted column in Figure 1 (birth year = 1946). It is important to note that unlike in a longitudinal comparison, which would be represented as a positive diagonal in Figure 1, in these comparisons the people in each cell are different.

Because birth year and test year were both continuous variables, they were used as simultaneous predictors in regression analyses, instead of as discrete factors in analyses of variance as in Ronnlund’s (Ronnlund & Nilsson, 2006; Ronnlund et al., 2005) and Schaie’s (Schaie et al., 1973; Schaie & Strother, 1968) studies. The relevant information from the regression analyses consists of estimates of the relations between measures of cognitive functioning and both birth year (between-cohort comparisons) and test year (within-cohort comparisons). Variations in both birth year and test year corresponded to differences in age, and therefore the key question was whether there were age-related differences in cognitive functioning in within-cohort comparisons, and if so, how their magnitude compared with the magnitude of the age-related differences in between-cohort comparisons.

Representativeness of samples in different years can be evaluated by comparing the scores of the sample participants with scores from a nationally representative normative sample. Age-adjusted scaled scores from the Wechsler Adult Intelligence Scale—Third Edition (WAIS-III; Wechsler, 1997a) and Wechsler Memory Scale—Third Edition (WMS-III; Wechsler, 1997b) were used for this purpose because they are based on a sample selected to match the proportions in the population with respect to major demographic characteristics.

**Method**

**Participants**

Characteristics of the participants in each test year are reported in Table 1. Across the entire sample, the birth years ranged from...
Within-Cohort Age Differences

from 1907 to 1989, chronological age ranged from 18 to 97, and the test year ranged from 2001 to 2011. There was a wide range of ages among the participants in each test year, and the correlation between birth year and test year was only .06.

Eighty-one percent of the participants classified themselves as White, 10% classified themselves as Black, and the remainder identified themselves as belonging to other ethnicities or to more than one ethnic group. The gender distribution of the sample was 65% female and 35% male, and participants had, on average, 15.7 years of formal education.

Between 205 and 913 adults were recruited as new participants each year except 2010, when no new participants were recruited. The average scaled scores for tests from the WAIS-III battery (Vocabulary and Digit Symbol) and from the WMS-III battery (Logical Memory and Word Recall) in each test year are reported in Table 1. In the normative sample, the

Table 1. Descriptive Characteristics of the Samples in This Study

<table>
<thead>
<tr>
<th>Test year</th>
<th>N</th>
<th>Age (years)</th>
<th>Birth year</th>
<th>Females (proportion)</th>
<th>Health</th>
<th>Education (years)</th>
<th>Vocabulary</th>
<th>Digit Symbol</th>
<th>Logical Memory</th>
<th>Word Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>205</td>
<td>50.7 (17.3)</td>
<td>1950 (17.3)</td>
<td>.65 (.5)</td>
<td>2.1 (0.9)</td>
<td>16.0 (2.5)</td>
<td>13.3 (3.2)</td>
<td>12.2 (2.8)</td>
<td>12.0 (3.1)</td>
<td>13.0 (2.9)</td>
</tr>
<tr>
<td>2002</td>
<td>269</td>
<td>48.3 (17.8)</td>
<td>1954 (17.8)</td>
<td>.63 (.5)</td>
<td>2.0 (0.8)</td>
<td>15.9 (2.8)</td>
<td>12.2 (3.1)</td>
<td>11.9 (3.1)</td>
<td>11.5 (2.7)</td>
<td>12.5 (3.0)</td>
</tr>
<tr>
<td>2003</td>
<td>330</td>
<td>47.4 (17.5)</td>
<td>1956 (17.5)</td>
<td>.66 (.5)</td>
<td>2.0 (0.8)</td>
<td>15.7 (2.6)</td>
<td>12.6 (3.0)</td>
<td>11.9 (2.9)</td>
<td>12.3 (2.8)</td>
<td>13.3 (3.4)</td>
</tr>
<tr>
<td>2004</td>
<td>458</td>
<td>51.3 (18.0)</td>
<td>1953 (18.0)</td>
<td>.63 (.5)</td>
<td>2.0 (0.9)</td>
<td>15.6 (2.6)</td>
<td>12.9 (2.9)</td>
<td>11.3 (2.8)</td>
<td>12.1 (2.6)</td>
<td>12.4 (3.4)</td>
</tr>
<tr>
<td>2005</td>
<td>670</td>
<td>49.8 (19.8)</td>
<td>1955 (19.8)</td>
<td>.66 (.5)</td>
<td>2.0 (0.8)</td>
<td>15.4 (2.8)</td>
<td>12.9 (3.3)</td>
<td>11.0 (2.9)</td>
<td>11.7 (3.0)</td>
<td>12.1 (3.2)</td>
</tr>
<tr>
<td>2006</td>
<td>913</td>
<td>51.1 (19.2)</td>
<td>1955 (19.2)</td>
<td>.66 (.5)</td>
<td>2.0 (0.9)</td>
<td>15.7 (2.5)</td>
<td>12.9 (2.7)</td>
<td>11.4 (2.7)</td>
<td>11.9 (2.8)</td>
<td>12.4 (3.2)</td>
</tr>
<tr>
<td>2007</td>
<td>444</td>
<td>52.3 (19.4)</td>
<td>1955 (19.4)</td>
<td>.62 (.5)</td>
<td>2.1 (0.9)</td>
<td>15.6 (2.8)</td>
<td>11.9 (2.8)</td>
<td>10.9 (2.9)</td>
<td>11.5 (3.0)</td>
<td>11.5 (3.5)</td>
</tr>
<tr>
<td>2008</td>
<td>255</td>
<td>54.6 (16.8)</td>
<td>1953 (16.8)</td>
<td>.64 (.5)</td>
<td>2.2 (0.9)</td>
<td>15.5 (2.9)</td>
<td>11.2 (3.2)</td>
<td>10.7 (3.0)</td>
<td>11.1 (3.0)</td>
<td>10.7 (3.4)</td>
</tr>
<tr>
<td>2009</td>
<td>236</td>
<td>57.1 (15.7)</td>
<td>1952 (15.7)</td>
<td>.69 (.5)</td>
<td>2.1 (0.8)</td>
<td>15.7 (2.6)</td>
<td>11.8 (3.1)</td>
<td>11.1 (2.7)</td>
<td>11.4 (3.0)</td>
<td>11.3 (3.4)</td>
</tr>
<tr>
<td>2011</td>
<td>440</td>
<td>52.5 (14.5)</td>
<td>1958 (14.6)</td>
<td>.68 (.5)</td>
<td>2.2 (0.8)</td>
<td>15.9 (3.0)</td>
<td>11.7 (3.6)</td>
<td>10.8 (2.8)</td>
<td>11.4 (3.4)</td>
<td>11.3 (2.9)</td>
</tr>
</tbody>
</table>

Note: The table presents mean values, with standard deviations in parentheses. Health was self-reported on a scale from 1, excellent, to 5, poor. The Vocabulary and Digit Symbol scaled scores are age-adjusted scores derived from the Wechsler Adult Intelligence Scale—Third Edition (Wechsler, 1997a), and the Logical Memory and Word Recall scaled scores are age-adjusted scores derived from the Wechsler Memory Scale—Third Edition (Wechsler, 1997b).
mean age-adjusted scaled score for each test is fixed at 10, with a standard deviation of 3. Most of the averages in Table 1 are between 11 and 13, and therefore one can infer that the participants in this project were functioning at about the 75th percentile of the nationally representative normative sample. The scaled scores varied across test years, as the correlations with test year were −.13 for Vocabulary, −.12 for Digit Symbol, −.09 for Logical Memory, and −.16 for Word Recall (all ps < .01). The samples may therefore have differed in degree of representativeness relative to the normative sample, being somewhat less select in more recent years.

**Measures**

Sixteen different tests were used to assess five cognitive abilities. Reasoning was assessed with the Raven’s Progressive Matrices test (Raven, 1962), the Shipley Abstraction test (Zachary, 1986), and the Letter Sets Test (Ekstrom, French, Harman, & Dermen, 1976). Spatial visualization was assessed with the Spatial Relations test (Bennett, Seashore, & Wesman, 1997), the Paper Folding Test (Ekstrom et al., 1976), and the Form Boards Test (Ekstrom et al., 1976). Vocabulary was assessed with the Vocabulary test from the WAIS-III (Wechsler, 1997a), the Picture Vocabulary test from the Woodcock-Johnson Cognitive Battery (Woodcock & Johnson, 1990), and multiple-choice synonym and antonym tests (Salthouse, 1993). Verbal memory was assessed with the immediate Logical Memory and Word Recall tests from the WMS-III (Wechsler, 1997b) and a paired-associates test (Salthouse, Fristoe, & Rhee, 1996). Perceptual speed was assessed with the Digit Symbol test from the WAIS-III (Wechsler, 1997a) and with the Pattern Comparison and Letter Comparison tests (Salthouse & Babcock, 1991). Previous research established that these tests were all reliable, and valid in the sense that they had moderate to high loadings on their respective ability factors (e.g., Salthouse, 2004; Salthouse & Ferrer-Caja, 2003; Salthouse, Pink, & Tucker-Drob, 2008).

**Results**

Because composite scores are more reliable and less influenced by test-specific factors than scores from individual tests are, all analyses were conducted on composite scores formed by averaging z scores for the tests representing each ability. Figure 2 portrays the average scores across all test years as a function of birth year (and age). As is typically found, reasoning, spatial visualization, verbal memory, and perceptual speed had a negative relation with age (i.e., a positive birth-cohort relation), and vocabulary had a positive relation with age (i.e., a negative birth-cohort relation) until about age 60. Because later birth years (corresponding to younger ages) were associated with higher levels of performance in four of the abilities, the slopes of cognitive performance as a function of birth year were positive for all composite scores except vocabulary, and

![Fig. 2. Mean composite scores in the five cognitive abilities as a function of birth year (and age). Error bars indicate standard errors.](image-url)
corresponded to differences of about 0.02 standard deviations per year.

Age-related differences in cognitive functioning were investigated with multiple regression analyses in which birth year and test year were simultaneous predictors of the cognitive-ability composite scores. The unstandardized regression coefficients from these analyses are presented in the left half of Table 2. The coefficients for birth year (between-cohort contrasts) represent the difference per year in the composite score when the variation in test year was controlled at the average test year (i.e., 2005.8).

The coefficients for test year indicate the difference per year in the composite score when the variation in birth year was controlled at the average birth year (i.e., 1954.6). Because increases in test year were associated with increases in age, these coefficients provide an estimate of the age-related differences within the same birth cohort. Note that all of these values were considerably larger than their standard errors, and each was more than twice as large as the estimate of the corresponding between-cohort (birth-year) age-related difference.

A second set of analyses included the age-adjusted scaled scores for the Vocabulary, Digit Symbol, Logical Memory, and Word Recall tests as covariates to adjust for differences in sample selectivity across test years. The effect of including these variables as covariates was to carry out all comparisons at the average values of the scaled scores (i.e., 12.5 for Vocabulary, 11.3 for Digit Symbol, 11.7 for Logical Memory, and 12.1 for Word Recall). Results of these analyses are presented in the right half of Table 2. The estimates of the birth-year effects were very similar to the values from the first set of analyses, in which there was no adjustment for sample selectivity. However, controlling the variation in scaled scores reduced the estimates of the test-year effects for all five cognitive abilities. The estimates were very small for the vocabulary and reasoning composite scores, but the within-cohort estimates of the age-related differences were approximately the same magnitude as the between-cohort (birth-year) estimates of the age-related differences for the spatial-visualization, verbal-memory, and perceptual-speed composite scores.

An additional set of analyses was conducted to determine if the relations of birth year and test year to cognitive ability varied according to birth cohort (and age). For this purpose, the analyses were repeated in separate birth-cohort groups corresponding to the birth-year boundaries specified in the horizontal axis of Figure 2. Because the standard deviations of the birth-year and test-year variables were very similar in each of these groups, these comparisons allowed the effects of birth year and test year to be evaluated when the range of variation was similar for the two predictor variables. Sample sizes in these groups ranged from 210 to 946, with a median of 597.

The regression coefficients from these analyses, in which the scaled scores were included as additional predictors to control for sample selectivity, are presented in the five panels of Figure 3. The coefficients are estimates of the age-related differences per year in the between-cohort (birth-year) and within-cohort (test-year) contrasts within the specified range of birth years. The regression coefficients for birth year were multiplied by −1 to facilitate comparison of the between-cohort and within-cohort effects. After this transformation, the two types of comparisons were expressed in the same direction, with negative coefficients corresponding to lower levels of cognitive ability with increased age.

There are two major points to note in Figure 3. The first is that the patterns across birth cohorts (and age groups) varied somewhat for different cognitive abilities. The functions were relatively flat for spatial-visualization ability, which indicates that the relation between performance and age was nearly constant across successive age (cohort) groups. In contrast, the functions were negative for the other abilities, which indicates that the relation between age and ability were more negative at older ages (and in earlier cohorts). The second, and more interesting, point regarding the results in Figure 3 is that in each panel, the magnitudes of the between-cohort (birth-year) and within-cohort (test-year) estimates of age-related differences are about the same. These results imply that age-related differences in cognitive ability at the average test year in between-cohort comparisons were very similar to age-related differences in cognitive ability at the average birth year in within-cohort comparisons.

<table>
<thead>
<tr>
<th>Cognitive ability</th>
<th>Original analyses</th>
<th>Analyses controlling for scaled scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Birth year</td>
<td>Test year</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>−0.012 (0.001)</td>
<td>−0.037 (0.005)</td>
</tr>
<tr>
<td>Reasoning</td>
<td>0.020 (0.001)</td>
<td>−0.051 (0.007)</td>
</tr>
<tr>
<td>Spatial visualization</td>
<td>0.022 (0.001)</td>
<td>−0.050 (0.005)</td>
</tr>
<tr>
<td>Verbal memory</td>
<td>0.019 (0.001)</td>
<td>−0.067 (0.005)</td>
</tr>
<tr>
<td>Perceptual speed</td>
<td>0.029 (0.001)</td>
<td>−0.071 (0.004)</td>
</tr>
</tbody>
</table>

Note: Standard errors are in parentheses.
Fig. 3. Estimates of the differences in composite cognitive-functioning scores per year in between-cohort (birth-year) and within-cohort (test-year) comparisons, as a function of birth year (and age). The coefficients presented are from analyses in which scaled scores were used as covariates. Note that the estimates for birth year have been multiplied by −1 so that for both sets of coefficients, negative values indicate lower scores with increased age.
Discussion

The major finding of this study is the discovery that within-cohort age-related differences in five cognitive abilities were nearly the same magnitude as the age-related differences in conventional cross-sectional (between-cohort) comparisons. The initial estimates for the within-cohort age-related differences were considerably more negative than those for the between-cohort age-related differences. However, because the samples in different test years may have differed in selectivity, the estimates after controlling for the scaled scores may provide the best indication of the within-cohort relations between age and cognitive ability. It is therefore noteworthy that these adjusted within-cohort estimates were nearly the same magnitude as the between-cohort estimates for the spatial-visualization, verbal-memory, and perceptual-speed composite scores.

The variation in scaled scores across test years was assumed to be attributable to differential selectivity of the samples, and thus the scaled scores were used as covariates to equate the samples on these variables. Although it is possible that differences across test years reflect systematic time-lag effects, this seems unlikely because the direction of the effects is opposite that reported for the Flynn Effect. That is, the trends in Table 1 indicate a slight decrease in the average scaled scores from early to later test years, instead of higher cognitive scores in more recent test periods, as is typically found. However, the trend is consistent with a broader, and less selective, recruitment of participants in later test years.

The finding that age-related differences in cognitive functioning are as large in contrasts within the same birth cohort as in contrasts involving people from different birth cohorts implies that birth year is a poor proxy for the changing aspects of the physical and social environment that may influence cognitive performance. Previously, I have discussed several limitations of operationalizing environmental changes in terms of birth cohort (Salthouse, 1991). Among these limitations are the following:

It is unlikely that individuals will be meaningfully grouped with respect to common experiences when they are classified according to arbitrary temporal boundaries...not all experiences are presumed to be relevant, and there is no assurance that individuals classified together on the basis of year of birth all share the critical experiences...[and]...because the impact of experiences may not be uniform among individuals born within specified temporal intervals, relying upon birth year to define cohorts also leads to the problem of possible differential effects of critical experiences among individuals treated as equivalent. (p. 117)

These concerns, in combination with the current findings that within-cohort age-related differences are nearly as large as between-cohort age-related differences, indicate that it may not be meaningful to refer to age-related differences in cognitive functioning as cohort differences when cohort is defined exclusively in terms of birth year.

It is important to emphasize that the current results are not inconsistent with the documented time-lag improvements in average scores in cognitive tests (i.e., the Flynn Effect). That is, the Flynn Effect can exist without necessarily being relevant to the interpretation of age-related differences or changes in cognitive functioning if the environmental changes affect people of all ages, and not merely people during certain developmental periods. The historical shifts in relevant aspects of the environment that influence cognitive-test scores could be analogous to the effects of inflation on income (Salthouse, 2010b). That is, when inflation is operating, average income for a given age will be higher in more recent years, but these time-related increases will not necessarily compromise the meaningfulness of cross-sectional age comparisons in any given year. It is not yet clear whether people of all ages benefit equally from time-lag improvements in cognitive functioning, but to the extent that they do, the distortion in maturational trends may be more pronounced in longitudinal comparisons than in cross-sectional comparisons (the same pattern found in the case of the effect of inflation on income).

Disentangling potential determinants of age-related cognitive change is challenging because there is still limited understanding of what is responsible for the relations between age and cognitive functioning. It is widely recognized that age is not a causal variable, but instead is best conceptualized as a continuum along which causal factors operate. The goal of many developmental researchers has therefore been to identify causal factors that could be used to replace the age variable in explanations. Because the current results suggest that age-related differences in several cognitive abilities are as large in contrasts within the same birth cohort as in conventional cross-sectional contrasts involving different birth cohorts, a high priority for future research should be the replacement of birth year as a proxy for time-related environmental influences. Rather than attributing age-related differences to unknown correlates of birth year, researchers may find it more productive to specify and measure the relevant characteristics, and include them in analyses to evaluate the role of these factors in the age-related differences and changes in cognitive functioning. In fact, one could argue that a decompositional strategy such as this is desirable with every presumed cause of human behavior.

The importance of investigating the aspects of the environment responsible for variations in human behavior has also been emphasized by other researchers. For example, Plomin and Daniels (1987) proposed a three-step approach to investigate the nature of environmental influences in behavioral genetics: document differential experiences, document associations between differential experiences and differential outcomes, and investigate the extent to which the associations between differential experiences and differential outcomes are causal. For researchers interested in interventions, an additional step might be to specify and investigate the mechanisms...
responsible for the environment-behavior associations. Implementing these steps will be difficult, and interpretations can be complicated if different aspects of the environment are relevant for different dimensions of behavior, and at different periods in development. However, environmental influences are too important to be inferred on the basis of what is not accounted for by other influences, or by using crude proxies that fail to capture the relevant aspects responsible for the behavioral variations of primary interest.

Declaration of Conflicting Interests
The author declared that he had no conflicts of interest with respect to his authorship or the publication of this article.

Funding
This research was supported by Award No. R37AG024270 from the National Institute on Aging. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institute on Aging or the National Institutes of Health.

References