Dear all,

This paper comes from my early work on my dissertation-in-progress (expected 2025). The dissertation follows the relationship between social science research on the elderly, corporate retirement practices, and age-based antidiscrimination law in the US since 1945. My main argument is still evolving, but I believe that this strain of research had effects beyond its immediate application in banning mandatory retirement ages; social science reshaped how Americans think about the mind in general and how it changes over time. This section comes from a middle chapter of the dissertation. In earlier chapters the reader has been introduced to work on the psychology and sociology of old age in the 1940s and 1950s, mainly focused on the University of Chicago's department of sociology. I've tried to make this paper as self-contained as possible, though there is a short section here on geropsychology that would make more sense with the context of previous chapters.

Thank you for reading my paper!

Jamie

Placing the Individual in Time: Intellectual Decline, Social Science Research, and Antidiscrimination Law

In the United States, unlike in most other wealthy countries, it is illegal in almost all cases to retire an employee once they reach a set age. Mandatory retirement, commonly practiced throughout the mid-twentieth century, was banned piecemeal, state by state and for public or private employees, culminating in a 1986 amendment to the federal Age Discrimination in Employment Act (ADEA).[1] Mandatory retirement was presented by its opponents as a civil rights issue, a discriminatory practice based on ageist stereotypes that portrayed the elderly as physically and mentally incapable of work.[2] Framed this way, voluntary retirement took on an air of progressive inevitability, simply the latest step in the long civil rights revolution that saw disparate marginalized groups receive state protection.

But the end of mandatory retirement was not inevitable, as illustrated by its continued use in other countries. Different explanations for its end in the US assign causation to different actors: perhaps corporations needed to make more efficient use of human resources in a globalizing, competitive economy; or a growing sense of crisis around the viability of Social Security made later retirement seem attractive to legislators; or the elderly organized into an effective lobbying force that could protect its own interests.[3] Each of these arguments has something to offer, but they all miss an important detail: the case for voluntary retirement, like

[1] Age Discrimination in Employment Amendments of 1986, Pub. L. No. 99-592, 100 Stat. 3342 (1986).
[2] Eliminating Mandatory Retirement: A Report by the Chairman of the Subcommittee on Health and Long-Term Care of the Select Committee on Aging, Comm. Pub. No. 99-561 (1986).
[3] William Graebner, *A History of Retirement: The Meaning and Function of an American Institution, 1885-1978* (New Haven: Yale University Press, 1980); Carole Haber and Brian Gratton, *Old Age and the Search for Security: An American Social History* (Bloomington: Indiana University Press, 1994); David Van Tassel and Peter N. Stearns, eds., *Old Age in a Bureaucratic Society: The Elderly, the Experts, and the State in American History* (New York: Greenwood Press, 1986); William Graebner, "Age and Retirement: Major Issues in the American Experience," in *Age in America: The Colonial Era to the Present*, ed. Corinne T. Field and Nicholas L. Syrett (New York: New York University Press, 2015); Julie C. Suk, "From Antidiscrimination to Equality: Stereotypes and the Life Cycle in the United States and Europe," *American Journal of Comparative Law* 60, no. 1 (2012): 75-98.

other antidiscrimination practices, was made using scientific evidence. The most famous example of this in US history is the *Brown v. Board of Education* Supreme Court case of 1954, in which Kenneth and Mamie Clark's work on black and white children's play with dolls was presented as evidence of the psychological impact of racism. To acknowledge the importance of the Clarks' scientific work is not to downplay the long political, social, and economic history of the civil rights movement, which extends far beyond a handful of psychological studies. But this scientific work was important, shaping Americans' understanding of racial discrimination for decades.[4] Likewise, the 1986 ADEA amendment was preceded by Congressional hearings in which expert witnesses presented evidence that a uniform retirement age did not reflect the variety of capabilities of older people and could harm employees by forcing them out of work before they were ready.

One of the loudest scientific voices in the fight against ageism was K. Warner Schaie, a practitioner in the field of life-span developmental psychology (LSDP). Little-known today, LSDP emerged in the 1950s and 1960s out of the attempt to apply the methodology of child developmentalism to subjects of any age. Schaie and his colleagues gathered a body of evidence that suggested that, contrary to widespread belief, adults' performance on intelligence tests did not decline from youth to old age. A 70-year-old was likely just as mentally capable as they had been at 50, knocking out a major argument in favor of mandatory retirement. LSDP researchers were most confident in these claims in the 1970s and early 1980s—precisely the moment when Congress moved to ban mandatory retirement.

[4] James T. Patterson, *Brown v. Board of Education: A Civil Rights Milestone and Its Troubled Legacy* (Oxford: Oxford University Press, 2001); Jonathan Zimmerman, "*Brown*-ing the American Textbook: History, Psychology, and the Origins of Modern Multiculturalism," *History of Education Quarterly* 44 (2004): 47-69.

This chapter follows Schaie's research from the 1950s, when he began the Seattle Longitudinal Study (SLS) on intelligence and aging, to 1980, when he appeared twice before Congress to present evidence for elderly people's capacity to continue working past 65.[5] Three findings bolstered his argument for voluntary retirement. First, evidence from the SLS suggested that different psychological faculties decline at different rates. An elderly woman who struggles with quick calculations might maintain a robust vocabulary into her 70s. This insight was possible because Schaie adopted for his study the Thurstone test of Primary Mental Abilities, an intelligence test that divided the mind into several orthogonal factors. Second, Schaie theorized that generational differences explained much of the apparent decline associated with old age. Old people only appear to perform poorly on tests when they are compared with generations born later, who benefit from more education, better nutrition, or any number of other environmental factors in their favor. Third, LSDP researchers grew confident that cognitive interventions could boost test performance and maintain intellectual functioning into old age. Even when decline occurred, it was not inevitable: good habits and regular practice could keep brains functioning late into life.

This line of research required that scientists think about time in new ways and adjust their tools for use in new contexts. Essentially, the project of LSDP was to apply the methodologies of child developmentalism and psychometrics to the research subjects of gerontology, bringing larger and larger spans of time into their analysis. Historians of science have thought through similar transformations using frameworks of scale and plasticity. Thinking about small and large scales simultaneously requires tools that can translate between the global, regional, and local,

[5] "Work after 65: Options for the 80's," Hearing before the Special Committee on Aging, United States Senate (April 24, 1980); "How Old Is 'Old'? The Effects of Aging on Learning and Working," Hearing before the Special Committee on Aging, United States Senate (April 30, 1980).

while manipulating time in new ways destabilizes the meanings of familiar scientific terms.[6] For LSDP researchers, studying a wider range of ages and developing the tools to compare the young and the old forced a reconsideration of what is meant by aging and decline, and whether such a thing as "normal" aging exists.

The two most important tools here, the longitudinal study and the intelligence test, moved between temporal scales in different ways. When the longitudinal study was applied across long spans of time, tracking subjects from young adulthood to old age, its methodology had to change to account for both age differences and generational differences. These methodological innovations were applied in the SLS and helped to define LSDP as a field separate from traditional developmentalism. Intelligence tests were different: in order to make meaningful comparisons between the young and old, it was essential that tests *not* change when they were applied to new subjects. Researchers suspected that elderly people might benefit from extra accommodations on tests, such as no time limits, but changing test conditions would make comparisons across age impossible. This represented a different kind of temporal manipulation: while the history of psychometrics in the U.S. is typically confined to the early-twentieth century, intelligence test data, carried forward by a long-term longitudinal study, continued to shape public policy into the 1980s.[7]

Factor Analysis and Adult Intelligence

[6] Deborah R. Coen, *Climate in Motion: Science, Empire, and the Problem of Scale* (Chicago: University of Chicago Press, 2018); Hannah Landecker, *Culturing Life: How Cells Became Technologies* (Cambridge, MA: Harvard University Press, 2007).
[7] John Carson, *The Measure of Merit: Talents, Intelligence, and Inequality in the French and American Republics, 1750-1940* (Princeton: Princeton University Press, 2007).

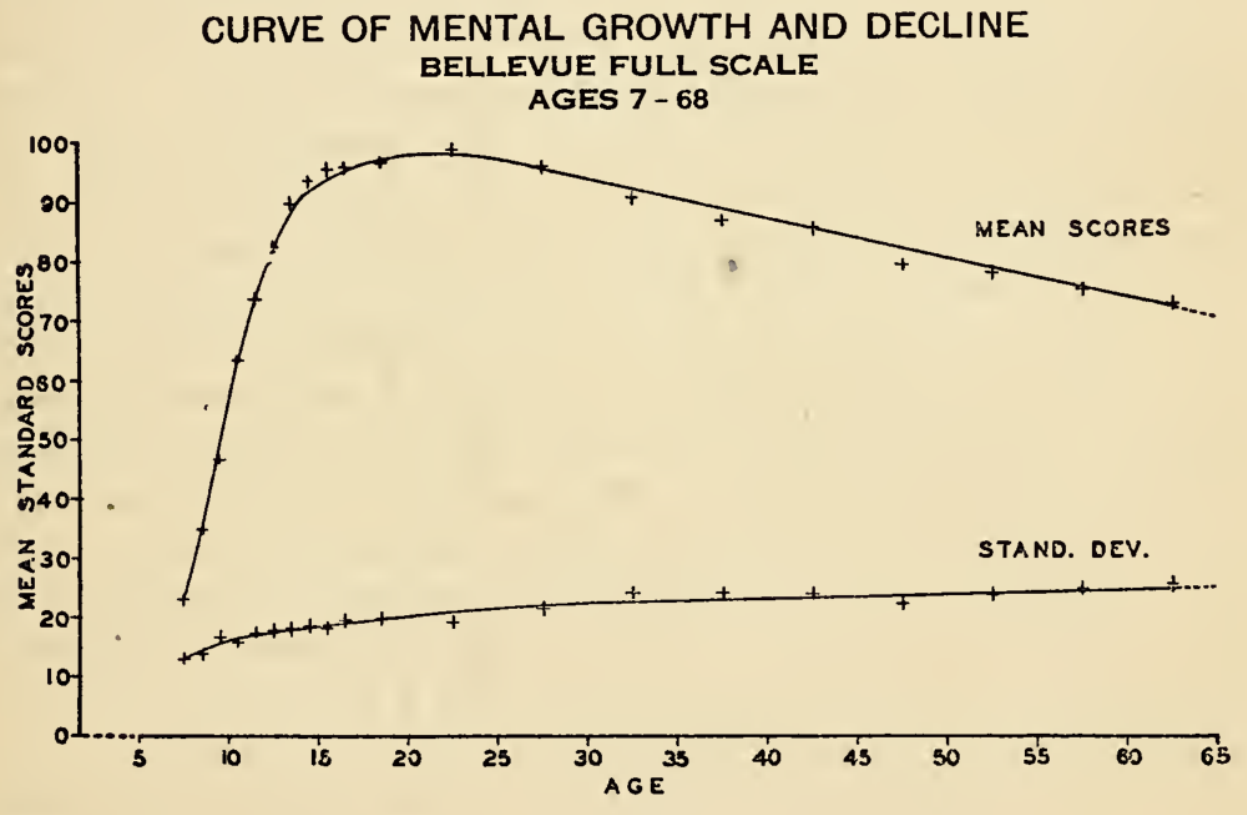
The Seattle Longitudinal Study (SLS) ran from 1956 to 2012, collecting data on its participants every seven years. It aimed to track changes in cognition over time, so the study's primary measurement was a battery of psychometric tests. The number of tests grew over time, but the core measurement, used throughout the SLS, was the Thurstone test of Primary Mental Abilities (PMAs), a battery of tests that would measure five forms of intelligence: verbal meaning, space, reasoning, number, and word fluency. This test emerged out of well-known psychometric debates of the 1920s and 1930s on the nature of intelligence. But Schaie, as director of the SLS, chose to use the PMA test to answer specific questions on the nature of age-related intellectual decline, not because of any deep commitment to a particular theory of the nature of intelligence. Once the PMA test was adopted, it had to be retained for the entire SLS in order for data from 1956 to be directly comparable to data from later decades. The nature of this longitudinal study extended the influence of classic psychometrics through the end of the twentieth century.

Klaus Warner Schaie was born an only child to a Jewish family in Stettin, Pomerania, Germany (today Szczecin, Poland) in 1928. In 1938, on Kristallnacht, Nazi supporters destroyed the Schaie family's store, and their father began seeking passage out of Germany. In June 1939, the family traveled via Trieste to Shanghai. Young Klaus, now a teenager, remained in Shanghai through World War II, learning English first in a refugee school and then in a newspaper job. In 1947, the family's father died of a stroke and the Chinese People's Liberation Army advanced closer to Shanghai. Schaie and his mother left for the United States, arriving in San Francisco in December. Soon afterward, he enrolled as an undergraduate at the University of California-Berkeley, where his exposure to intelligence testing and developmental psychology shaped his future career. Schaie's major advisor was the psychologist Read Tuddenham, who introduced

him to contemporary research in psychometrics. Schaie recalls being particularly interested in the research of Louis Leon and Thelma Gwinn Thurstone, and took as his first research project the application of their work on children to adults.[8]

In the 1950s, the leading researcher on adult intelligence was David Wechsler. Earlier psychometricians, such as Lewis Terman, had focused on testing childhood and adolescent intelligence, largely for the purpose of sorting and evaluating students. Early versions of Terman's Stanford-Binet test reported subjects' performance in terms of "mental age," under the assumption that a normal child's test performance would improve in some predictable way over time. But this scoring system broke down for adults, because, Wechsler argued, intelligence only increased from birth into the teenage years, followed by a slow decline through the rest of life. Wechsler was responsible for shifting psychometrics away from mental age and toward the intelligence quotient (IQ), a metric that can be applied to subjects of any age. The result of this research, the Wechsler Adult Intelligence Scale (WAIS), overtook the Stanford-Binet test in popularity, and remains the most widely-used intelligence test in the U.S.[9]

[8] K. Warner Schaie, "Living with Gerontology," in "A Tribute to K. Warner Schaie," ed. Phyllis Miller, special issue, *Mensa Research Journal* 32, no. 3 (2001): 6-25.
[9] David Wechsler, *The Measurement of Adult Intelligence*, 2nd ed. (Baltimore: Williams and Wilkins, 1941), especially 19-35.



According to David Wechsler's data, intelligence peaked between ages 15 and 20 and then declined throughout adulthood. Throughout the mid-twentieth century this narrative of decline was widely accepted. Wechsler, *Measurement of Adult Intelligence*, 29.

Schaie's first research project was to measure adult intelligence using alternatives to the WAIS. Wechsler's test was based on the theory of general intelligence, developed by the English psychologist Charles Spearman. According to Spearman, all measurements of mental ability could be separated into two factors, which he named *g* and *s*: *g* was a description of a person's mental ability in general, while *s* was specific to particular tasks. In other words, Spearman acknowledged that people might be better or worse at particular mental tasks, but maintained that someone who performed well in one area was likely to perform well in other areas. Intelligence

tests were tools for measuring *g*, the exact ontological meaning of which was unclear; Spearman frequently characterized it as all-purpose "mental energy."[10]

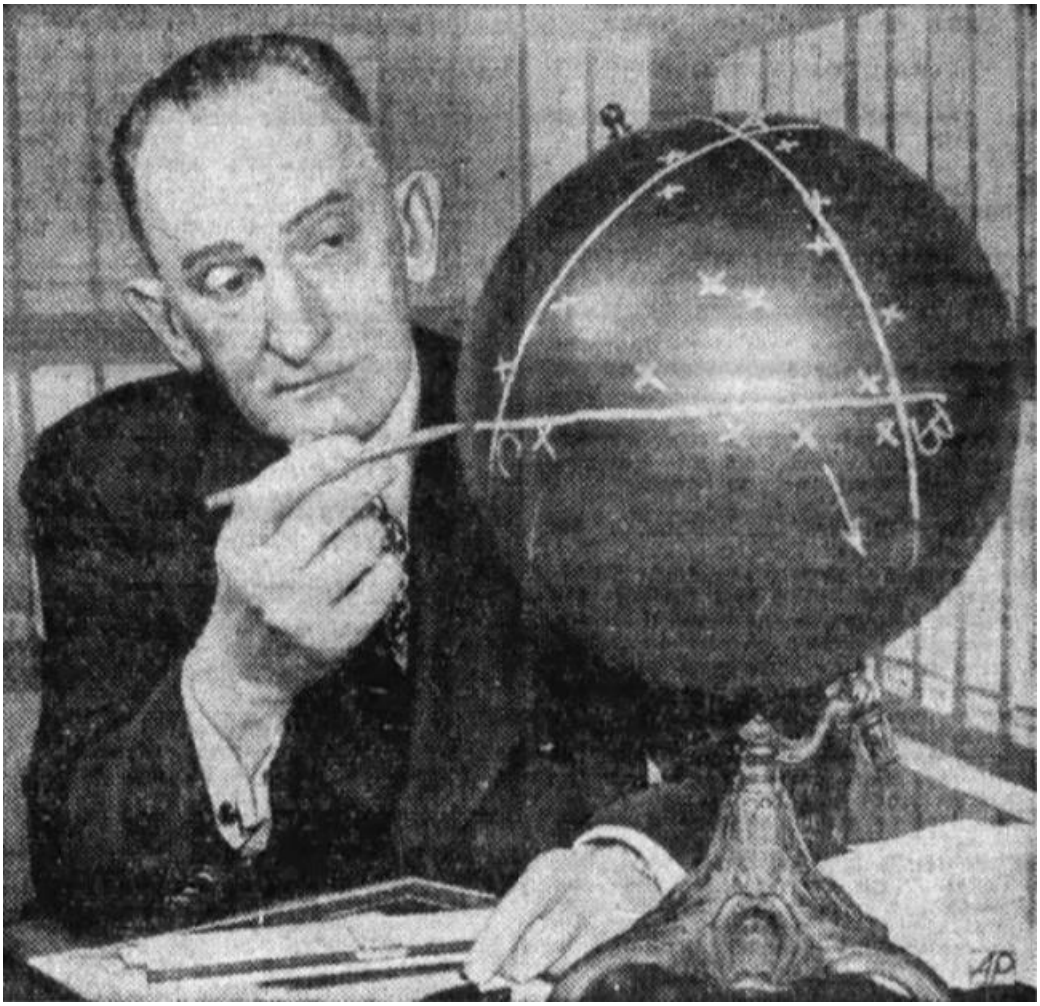
The most popular alternative to general intelligence, and the theory that appealed most to Schaie, was the system of primary mental abilities (PMAs), developed over the 1930s by husband-and-wife team Louis Leon Thurstone and Thelma Gwinn Thurstone.[11] According to this theory, intelligence was not comprised of a single all-purpose "energy," but seven distinct abilities, which the Thurstones named verbal meaning, space, reasoning, number, word fluency, memory, and perceptual speed.[12] Traditional tests, the Thurstones argued, were valid measures of intelligence, but could not distinguish between these factors of intelligence: a subject might receive an average intelligence score because they had average performance in all seven PMAs, or because they performed above average in some and below average in others. The PMA test battery, developed by the Thurstones and sold by the American Council on Education, promised to give educators and researchers a more complete picture of young people's minds than competing general intelligence tests.

This project of breaking down mental functioning into distinct abilities was not new. The Thurstones acknowledged that the PMA system bore a passing resemblance to nineteenth-century faculty psychology—the belief that specific mental and emotional capabilities were associated with discrete physical locations on the brain. Famously, this theory was the basis of

[10] H. Wildon Carr, A. Wolf, and C. Spearman, "Symposium: The Nature of Intelligence, "*Proceedings of the Aristotelian Society, Supplementary Volumes* 5 (1925): 1-27, quoted on 27.
[11] Thelma's first published work on PMAs appeared in 1941; before then, all their work was credited solely to Louis. By Louis's own admission, the two collaborated on this project throughout their career, though the exact scope of Thelma's contributions in the 1930s are unclear. See Carolyn T. Bashaw and W. L. Bashaw, "Thelma Gwinn Thurstone," in *Women in Psychology: A Bio-Bibliographic Sourcebook*, ed. Agnes N. O'Connell and Nancy Felipe Russo (New York: Greenwood Press, 1990).
[12] L. L. Thurstone, *Primary Mental Abilities* (Chicago: University of Chicago Press, 1938).

phrenologists' practice of skull measurement.[13] But the Thurstones insisted that the PMAs were "objective" measures of mental abilities. While faculty psychology simply projected common-sense assumptions about human behavior onto the brain, the PMAs were based on factor analysis, a mathematical technique that the Thurstones helped develop.[14] This technique promised to simplify scientific work by reducing the complexity of a problem: in the best-case scenario, factor analysis can explain a large number of variables with a small number of common factors. The Thurstones arrived at the PMA system by gathering a mass of test data, administering 56 tests of a variety of skills to a group of volunteers from the University of Chicago. Analyzing the correlations between average scores on these tests, the Thurstones found that these 56 scores could be explained with 12 common factors. Seven factors loaded highly with clusters of tests recognizable as everyday abilities; these became the seven PMAs. Three factors seemed to correspond to random clusters of tests—the Thurstones discarded these as meaningless. The remaining two factors were borderline cases, which the Thurstones named "restriction" and "deduction." While the Thurstones initially treated these two factors as areas for future research, they eventually dropped them from their system, leaving a canonical list of seven PMAs.[15]

[13] Courtney E. Thompson, "Phrenology," *Encyclopedia of the History of Science* (November 2021), accessed March 27, 2023. https://doi.org/10.34758/ymce-b249.
[14] L. L. Thurstone, *The Vectors of Mind: Multiple-Factor Analysis for the Isolation of Primary Traits* (Chicago: University of Chicago Press, 1935).
[15] Thurstone, *Primary Mental Abilities*, 88.



L. L. Thurstone demonstrates factor analysis in three dimensions: test scores, represented by X's, are vectors that are described with a coordinate system of primary abilities labeled A, B, and C. Like the phrenological bust, this sphere was meant to represent intelligence visually for a popular audience; unlike the phrenological bust, it never caught on. *Enid Daily Eagle* (Enid, OK), May 2, 1945, p. 9.

This background in factor analysis points to a tension in the PMA theory. On the one hand, the Thurstones wanted the seven abilities to be objectively real, claiming that they had discovered the "building blocks" of the human mind. On the other hand, portraying themselves as responsible empiricists, the Thurstones acknowledged that the PMAs were simply mathematical abstractions—different data or different decisions in its analysis could have yielded different factors, or a different number of factors. Their work was provisional, and they expected that future research in neurology or genetics would confirm whether the PMAs were

truly fundamental to the mind.[16] Much has already been written about this problem of reification in psychometrics.[17] The important point here is that the PMAs were relatively stable entities, suitable as the basis for a long-running longitudinal study, yet individual PMAs could be added or dropped without damaging the theory as a whole. While the Thurstones typically listed seven PMAs in their theoretical writings or when addressing the public, the number of PMAs fluctuated between different publications of their test battery. The 1941 battery for adolescents only tested six of the seven abilities, dropping perceptual speed.[18] For the 1949 revision of this battery, the Thurstones constructed a shortened, simplified test that could be completed in a few hours rather than several days; they dropped the memory test, which was too long to administer in this context.[19]

Schaie adopted this 1949 test battery for the research that would eventually become the SLS, unwittingly committing himself to a lifetime of studying five factors of intelligence: verbal meaning, space, reasoning, number, and word fluency. His first project was a 1951 pilot study of PMA performance in adults aged 53 to 78, which would test whether Wechsler's prediction of intelligence decline throughout adulthood held for individual abilities. But this raised questions about the test's applicability to different ages. The Thurstone battery was designed for teenagers, and administering it to a 70-year-old required rethinking the test's structure and difficulty. Schaie's solution was to administer the tests with a time limit, as directed in the test manual, but then to allow participants to complete the test with no time limit, yielding separate timed and

[16] Thurstone, *Primary Mental Abilities*, 2.
[17] Stephen Jay Gould, *The Mismeasure of Man* (New York: W. W. Norton, 1981), especially the discussion of the Thurstones on 296-316; Kurt Danziger, *Constructing the Subject: Historical Origins of Psychological Research* (Cambridge: Cambridge University Press, 1990).
[18] L. L. Thurstone and Thelma Gwinn Thurstone, "Instructions for Examiner: The Chicago Tests of Primary Mental Abilities for Ages 11 to 17" (Washington, D. C.: American Council on Education, 1941).
[19] L. L. Thurstone and Thelma Gwinn Thurstone, "Examiner Manual for the SRA Primary Mental Abilities Test, Intermediate—Ages 11 to 17" (Chicago: Science Research Associates, 1949), Counseling Center Records, box 3, Carnegie Mellon University Archives.

untimed scores. The rationale—based simply on common-sense assumptions about the differences between the young and old—was that elderly people would likely need additional time to complete test questions.[20]

However, untimed testing proved to be impractical. Schaie's goal was to make comparisons across the lifespan, which required consistency in measuring people of different ages. In his research from the 50s, Schaie compared adult test performance to the norms provided in the PMA test battery manual: rather than providing raw test scores, which lacked context, he expressed his subjects' performance as better or worse than the average 17-year-old's performance, making the extend of old-age decline clear. These comparisons would only make sense if adolescents and adults were given the same test—if both groups were subject to the same time limits. Schaie reported the results of his untimed tests in his 1951 pilot test, but reverted to timed PMA tests for all future research.

Schaie's dissertation, completed at the University of Washington in 1956, was both an expansion of this pilot study and the first measurement round of the future SLS. Subjects were randomly drawn from the membership rolls of the Group Health Cooperative of Puget Sound, one of the first health maintenance organizations (HMOs) in the country. In the following decades, the growth of the American health insurance industry would provide the SLS with a steady supply of research subjects. Again, Schaie conducted a cross-sectional analysis of PMA test scores, this time on adults from age 20 to 70. He also administered a test he had designed himself, the Test of Behavioral Rigidity (TBR), designed to measure subjects' ability to quickly adapt to new situations.[21] The results of this study were consistent with Schaie's earlier pilot

[20] K. Warner Schaie, Fred Rosenthal, and Robert M. Perlman, "Differential Deterioration of Factorially 'Pure' Functions in Later Maturity," *Journal of Gerontology* 8 (1953): 191-96.
[21] K. Warner Schaie, "Test of Behavioral Rigidity: Preliminary Manual" (Palo Alto: Consulting Psychologists Press, 1960), Counseling Center Records, box 3, Carnegie Mellon University Archives.

study and with Wechsler's results. Different PMA and TBR scores might decline at different rates, but decline was inevitable. He characterized decline as "a maturational phenomenon which occurs at a relatively constant rate and whose degree is presumed to be a function of original endowment."[22] Schaie would eventually interpret variability in PMA decline rates as an argument against the "myth" of monolithic decline, but this position would take time to develop.

Research Designs across the Lifespan

It is impossible to pinpoint an exact moment when this 1956 dissertation ceased to be a one-time study and became the start of an ongoing longitudinal study, but after two follow-up studies in 1963 and 1970 it was clear that the SLS was applying new methodological tools to stake out a new position on the question of old-age decline. Schaie's unique theoretical and methodological commitments, the new professional identity of life-span developmental psychology, and the SLS itself all emerged together over the 1960s. Before then, Schaie's early studies on adult intelligence were categorized as "geropsychology," or the subfield of gerontology dealing with the psychology of aging. Gerontology grew rapidly after World War II, and its growth only accelerated as the U.S. population aged.[23] Schaie gave his first conference presentation on age and intelligence at the Second International Gerontological Conference, held in St. Louis in September 1951. Writing retrospectively decades later, he described meeting and speaking with major geropsychologists such as Robert Havighurst, James Birren, and Robert

[22] The dissertation results were published as K. Warner Schaie, "Rigidity-Flexibility and Intelligence: A Cross-Sectional Study of the Adult Life Span from 20 to 70 Years," *Psychological Monographs: General and Applied* 72, no. 9 (1958): 1-26, quoted on 23.
[23] W. Andrew Achenbaum, *Crossing Frontiers: Gerontology Emerges as a Science* (New York: Cambridge University Press, 1995).

Kleemeier at this conference—recollections that demonstrate the impact of geropsychology on Schaie's early career.[24]

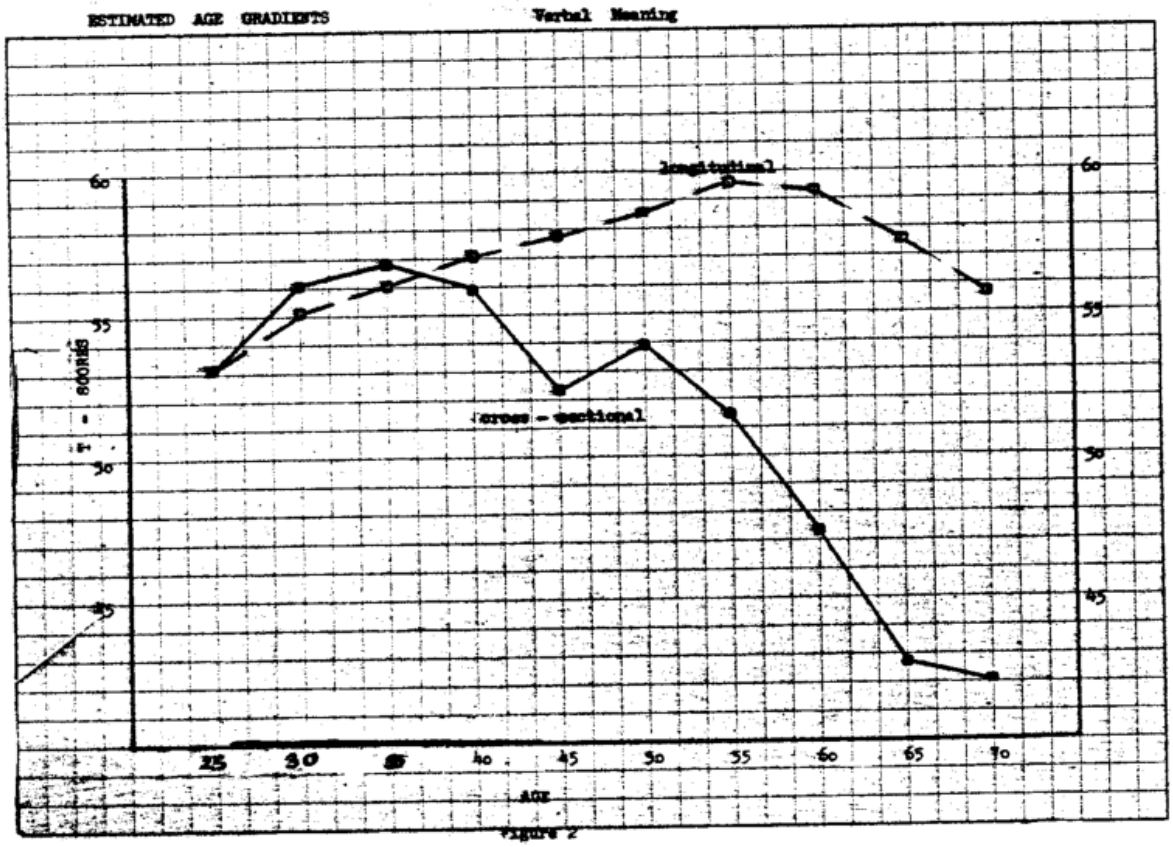
By the end of the 1960s, Schaie and his colleagues had established the new field of life-span developmental psychology (LSDP) as a discipline apart from geropsychology. Schaie took early steps in this direction by turning from empirical studies of aging to deeper methodological questions. Researchers in many fields had recognized contradictions between the results of cross-sectional and longitudinal studies. A cross-sectional study takes place all at once, comparing different subgroups of a population (in this case, people of different ages) and taking a snapshot of their differences. A longitudinal study takes place over time, following a defined group and tracking changes as they age. In theory, both kinds of studies should give insight into the differences between old and young people and allow researchers to express various health metrics as a function of time, plotting changes on graphs stretching from birth to middle age to death. Yet the two study designs yielded different results. Cross-sectional studies often found that subjects' performance peaked in their 20s and 30s, followed by steady decline through middle and old age. Longitudinal studies on the same metric typically found peaks in performance later in life followed by only moderate decline, with any major performance decrease occurring within a few years of death. These findings reappeared across many health metrics, from grip strength to lung capacity to intelligence test performance (Wechsler's conclusion that intelligence peaks early was based on cross-sectional studies).[25] This pattern

[24] Schaie, "Living with Gerontology," 11-12.
[25] Raymond G. Kuhlen, "Age and Intelligence: The Significance of Cultural Change in Longitudinal vs. Cross-Sectional Findings," *Vita Humana* 6, no. 3 (1963): 113-24; A. Damon, "Discrepancies Between Findings of Longitudinal and Cross-Sectional Studies in Adult Life: Physique and Physiology," *Human Development* 8 (1965): 16-22.

across domains suggested that the problem was some methodological issue with cross-sectional or longitudinal studies, or both.

In 1963, Schaie conducted a follow-up to his dissertation study that addressed the discrepancy between research designs directly. The 1956 study had been a cross-sectional comparison of 500 adults ranging from 20 to 70 years old. Seven years later, with funding from the National Institute of Child Health and Human Development (NICHD), Schaie contacted and retested 302 of these original subjects on the PMA and TBR batteries, along with an additional random sample of new participants, yielding a second cross-sectional comparison of adults between 20 and 77 years of age. But because many of these subjects had been tested twice, the data could also be interpreted as a collection of longitudinal studies—though unlike most longitudinal studies, this tracked adults of many different birth years. Schaie showed that the discrepancy between the two designs appeared within the same dataset: when interpreted as a cross-sectional study, the data from 1963 suggested that test scores peak in a person's 20s or 30s and then decline; when interpreted as a longitudinal study, the changes between the 1956 and 1963 datasets suggested that scores peak after age 50, followed by moderate decline. The only areas where early peaks persisted in longitudinal studies were in tests that relied on quick reaction time. Schaie concluded that neither cross-sectional nor longitudinal studies simply measured the effects of aging: other variables interacted and caused the two designs to produce different results.[26]

[26] K. Warner Schaie and Charles R. Strother, "A Cross-Sequential Study of Age Changes in Cognitive Behavior," paper presented at the meeting of the Midwest Psychological Association at St. Louis, MO, April 1964. https://sls.psychiatry.uw.edu/wp-content/uploads/2020/02/CrssSeqStudyAgeCog.pdf.



Data from Schaie's 1963 study of PMA test performance over the lifespan. According to cross-sectional data, scores in verbal meaning peak at 35; according to longitudinal data, scores peak at 55. (Schaie and Strother, "Cross-Sequential Study.")

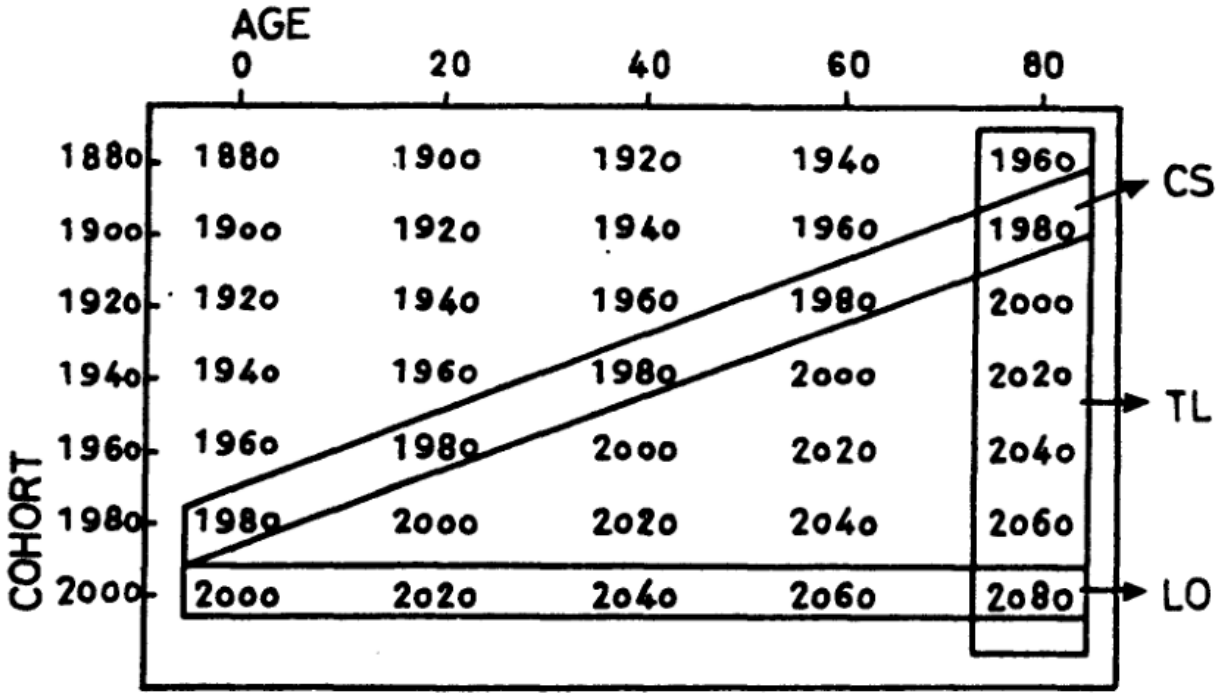
Confounding variables were the subject of Schaie's 1965 paper "A General Model for the Study of Developmental Problems," probably the most often-cited paper of his career.[27] Schaie's "general developmental model" would resolve the discrepancies between cross-sectional and longitudinal studies and propose new methods to more accurately track developmental changes. In classic developmental science, any human characteristic, from height to intelligence to social adjustment, could be expressed mathematically as a function of aging.[28] The goal of

[27] K. Warner Schaie, "A General Model for the Study of Developmental Problems," *Psychological Bulletin* 64, no. 2 (Aug 1965): 92-107.
[28] William Kessen, "Research Design in the Study of Developmental Problems," in *Handbook of Research Methods in Child Development,* ed. Paul H. Mussen (New York: John Wiley & Sons, 1960), 36-70.

developmental science, then, is to produce average growth curves showing how some attribute changes as individuals age. But Schaie argued that aging was not the only change over time that affected development. Individuals in different cohorts (that is, who were born in different years) grow up in separate environments, which can introduce any number of confounders that make comparison difficult—maybe one generation has access to better nutrition or education than another, or parenting methods have changed over time. And even in comparisons within the same cohort, changes in the environment could confound age-related changes: if a study finds behavioral differences between ages five and ten for children born in 1960, this might be caused by differences between five- and ten-year-olds or by differences between the environments of 1965 and 1970. The core of Schaie's general developmental model was to account for these ambiguities by expressing development as a function of not one but three variables. In addition to aging, development was also a function of cohort and time of measurement.

Traditional study designs can only control for one of these variables, leaving the other two as confounders. Schaie's paper recontextualized cross-sectional and longitudinal designs in terms of his new three-variable developmental model, showing how they were unable to accurately track changes in human behavior over time. Cross-sectional studies keep the time of measurement constant, but compare subjects of different ages and different cohorts. If a cross-sectional study finds that the young and old behave differently, it is impossible to tell whether this is the result of aging or generational differences. Longitudinal studies track a single cohort, but make measurements at different times as the subjects continue to age. Schaie listed a third study design, used occasionally in social science research, which he named "time-lag studies." In this design, subjects of a constant age are tested at multiple points in time—perhaps a group of 12-year-olds is tested in 1950, then a separate group of 12-year-olds is tested in 1960, then a

third group in 1970, and so on. This controls for the subjects' age, but allows both cohort and time of measurement to vary. None of these designs could provide researchers with unambiguous interpretations.

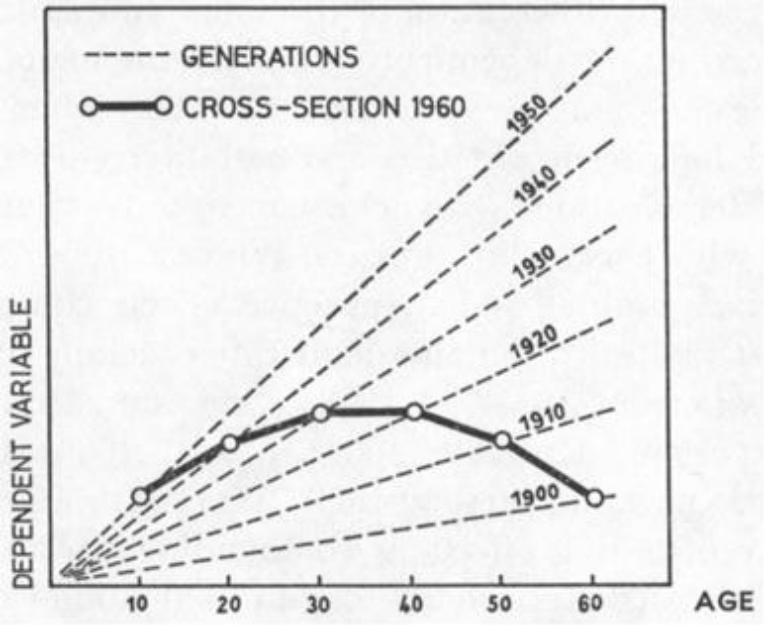


This schematic represents the life-span developmental psychologists' critique of traditional research methods. Each cell in the table represents a possible time of measurement for a given cohort (birth year) at a given age. Cross-sectional, longitudinal, and time-lag studies correspond to a single diagonal, row, or column of the table. In all three cases, the remaining two variables are not held constant, which creates confounders. The SLS attempted to solve this problem by occupying as many cells in the table as possible, tracking many cohorts over many ages. (P. B. Baltes, "Longitudinal and Cross-Sectional Sequences in the Study of Age and Generation Effects," *Human Development* 11, no. 3 (1968): 158)

Schaie's solution was to combine multiple of these traditional designs in more complicated studies. He initially used the term "sequential strategies" to refer to these combinations of cross-sectional, longitudinal, and time-lag studies, although this term never caught on. The most efficient design for developmental research, according to Schaie, was to begin by measuring a random sample of subjects of many ages, then measure these subjects again at a later time, and accompany this with a new random sample of subjects of all ages. Using a subset of this data yielded the traditional cross-sectional, longitudinal, and time-lag

designs. These designs still had to deal with the problem of confounding, but comparing data between different designs could help researchers isolate whether change occurred due to differences in age, differences in cohort, or differences in time of measurement. Sequential strategies provided a theoretical justification for Schaie's 1956 and 1963 studies after the fact. The SLS would take these two studies as its basis and continue to apply this most efficient design over many iterations. Theory and experimental practice informed one another.

In developing these methods, Schaie staked out what made the new field of life-span developmental psychology (LSDP) different from both gerontology and child developmentalism. Like the Thurstones' PMAs, Schaie's general developmental model was an attempt to break the object of scientific study down into its component parts. But this discussion of age, cohort, and time of measurement effects was not an application of factor analysis. Instead, it was an elaboration of "change over time," the basis of developmental research, and an attempt to disambiguate what "time" meant for developmentalists. The problem of age and generational differences confounding one another is common to all areas of developmental science, but it only becomes a problem when comparing across large age gaps and long spans of time—when studying minors, a cross-sectional study can only involve a cohort difference of at most 18 years, not enough to introduce serious generational differences. The new field of LSDP was set apart through a complex interaction of methodology and choice of subject. Studying elderly people made LSDP different from child developmentalism; its developmental framework made LSDP different from gerontology; the methodological innovations of the SLS and general developmental model made LSDP more than a simple combination of child developmentalism and gerontology.



An example of how cross-sectional studies can confound age and cohort effects. In this highly-artificial example, all subjects' performance on a test increases linearly with time, but subjects born in later generations improve more steeply. If young and old subjects are tested at the same time, older subjects are placed at a disadvantage because of these generational differences. The study results suggest that performance peaks in middle age and then declines; in reality, performance increases across the lifespan. (P. B. Baltes, "Longitudinal and Cross-Sectional Sequences in the Study of Age and Generation Effects," *Human Development* 11, no. 3 (1968): 152.)

The years immediately after the general model paper's publication saw LSDP emerge as a distinct field, largely through the personal efforts of Schaie, now chair of the psychology department at West Virginia University (WVU). In the context of President Johnson's Great Society, unprecedented federal funds flowed into Appalachia to encourage its development. Between 1968 and 1972, the WVU psychology department received over $250,000 in grants from the Public Health Service and National Institute of Child Health and Human Development (NICHD).[29] During this period, Schaie invited Paul Baltes and John Nesselroade (who would become longtime LSDP colleagues) to faculty positions at WVU and started the first doctoral

[29] "2 WVU Grants Announced," *Dominion News* (Morgantown, WV), June 29, 1968, p. 14; "Health Service Gives WVU Funds," *Weirton Daily Times*, January 21, 1970, p. 23; "WVU Receives Grant for 'Aging' Research," *Sunday Gazette-Mail* (Charleston, WV), March 26, 1972, p. 18; "WVU Gets 2 Grants," *Beckley Post-Herald*, August 14, 1972, p. 14; "Behavior Grant," *Beckley Post-Herald*, October 15, 1972, p. 3.

program in life-span developmental psychology.[30] Two conferences hosted at the university track the development of LSDP as an autonomous field. The first, hosted by Schaie in 1967, was framed as a conference on "geropsychology," suggesting that Schaie's research was a specialty within the established field of gerontology.[31] The second, more significant conference, hosted by Baltes and Larry Goulet in 1969, was explicitly devoted to life-span developmental psychology and helped to define the field, its goals, and its methods.[32] In addition to a paper by Schaie on the SLS and his developmental model, the conference featured research on complex forms of factor analysis, visual perception, and learning in children and adults.

Popularization and Controversy

Shortly after these conferences, newly-confident LSDP researchers began publicizing their results. Continuing the pattern of gathering new data every seven years, the SLS conducted its third study in 1970. Schaie and his colleagues again found significant differences between generations on test scores but little decline within specific generations' scores from 1956 to 1963 to 1970.[33] In 1974, Schaie and Baltes shared these results in an article for the popular magazine *Psychology Today*.[34] Their language reflected the confidence of LSDP researchers in the mid-1970s: intellectual decline was a "myth," the apparent decline in test scores being almost entirely the result of cohort differences. The myth persisted because young people stereotyped the old,

[30] Schaie, "Living with Gerontology," 17-19.
[31] K. Warner Schaie, ed., *Theories and Methods for Research on Aging* (Morgantown: West Virginia University Press, 1968).
[32] L. R. Goulet and Paul B. Baltes, eds., *Life-Span Developmental Psychology: Research and Theory* (New York: Academic Press, 1970).
[33] K. Warner Schaie, Gisela V. Labouvie, and Barbara U. Buech, "Generational and Cohort-Specific Differences in Adult Cognitive Functioning: A Fourteen-Year Study of Independent Samples," *Developmental Psychology* 9, no. 2 (1973): 151-66.
[34] Paul B. Baltes and K. Warner Schaie, "The Myth of the Twilight Years," *Psychology Today* 8 (March, 1974): 35-40.

punishing them for acting independently and creating a self-fulfilling prophecy of elderly incompetence. Schaie had noted from the beginning of the SLS that different PMA scores change at different rates; now this was presented as evidence that intelligence was maintained into old age, as the authors noted that decline was only evidence on tests that relied on quick reaction time. Schaie and Baltes recommended compensatory education for the elderly, pointing to pilot studies by Baltes and colleagues showing that proper training could speed up subjects' responses to questions.[35] This term, "compensatory education," originally referred to early-childhood education programs such as Head Start; in both methodology and policy proposals, LSDP borrowed from child psychology. In light of the SLS results and the potential of compensatory education, Schaie and Baltes argued that older adults should not be unilaterally excluded from work, and should be given opportunities for second-career training and voluntary (rather than mandatory) retirement.

This *Psychology Today* article appeared in the midst of a growing movement for elderly rights. The term "ageism" was coined by Robert Butler in 1968, providing new language to conceptualize mistreatment of the elderly as a civil rights issue. Organizations devoted to elderly rights ranged from the moderate National Council of Senior Citizens, which lobbied for improvements to Social Security and Medicare, to the radical Gray Panthers, which allied itself with the movements for nuclear disarmament and withdrawal from the war in Vietnam.[36] This popular activity was accompanied by increased funding for research on aging, a trend encapsulated by the founding of the National Institute on Aging (NIA) in 1974. Soon afterwards, the major SLS funding source switched from the NICHD to the NIA. In this context, Schaie and

[35] W. J. Hoyer, Gisela V. Labouvie, and P. B. Baltes, "Modification of Response Speed Deficits and Intellectual Performance in the Elderly," *Human Development* 16 (1973): 233-42.
[36] Jesse F. Ballenger, *Self, Senility, and Alzheimer's Disease in Modern America: A History* (Baltimore: Johns Hopkins University Press, 2006), 75-78.

his colleagues found an audience eager for good news on the prospect of maintaining intelligence into old age. Earlier generations of gerontologists had acted as public advocates for the elderly; in this new context, LSDP researchers could portray themselves almost as civil rights activists.

In bringing their research to a popular audience, Schaie and Baltes staked out a position as optimists on the question of aging; they also invited accusations of wishful thinking. Their *Psychology Today* article made the by-now familiar critique of cross-sectional studies, pointing specifically to recent cross-sectional research by Raymond Cattell and his disciple John L. Horn, which found that intelligence peaked and declined early. Horn and his colleague at the University of Denver, Gary Donaldson, responded with an attack on the SLS and its findings, particularly what they saw as irresponsible use of the term "myth." This became the most acrimonious debate in the history of LSDP. Between 1976 and 1977, the two sides—Horn and Donaldson opposite Schaie and Baltes—published responses and counter-responses in *American Psychologist*, the flagship journal of the American Psychological Association.[37] Schaie and Baltes had made confident claims to a public audience; they now had to defend their research in a professional setting.

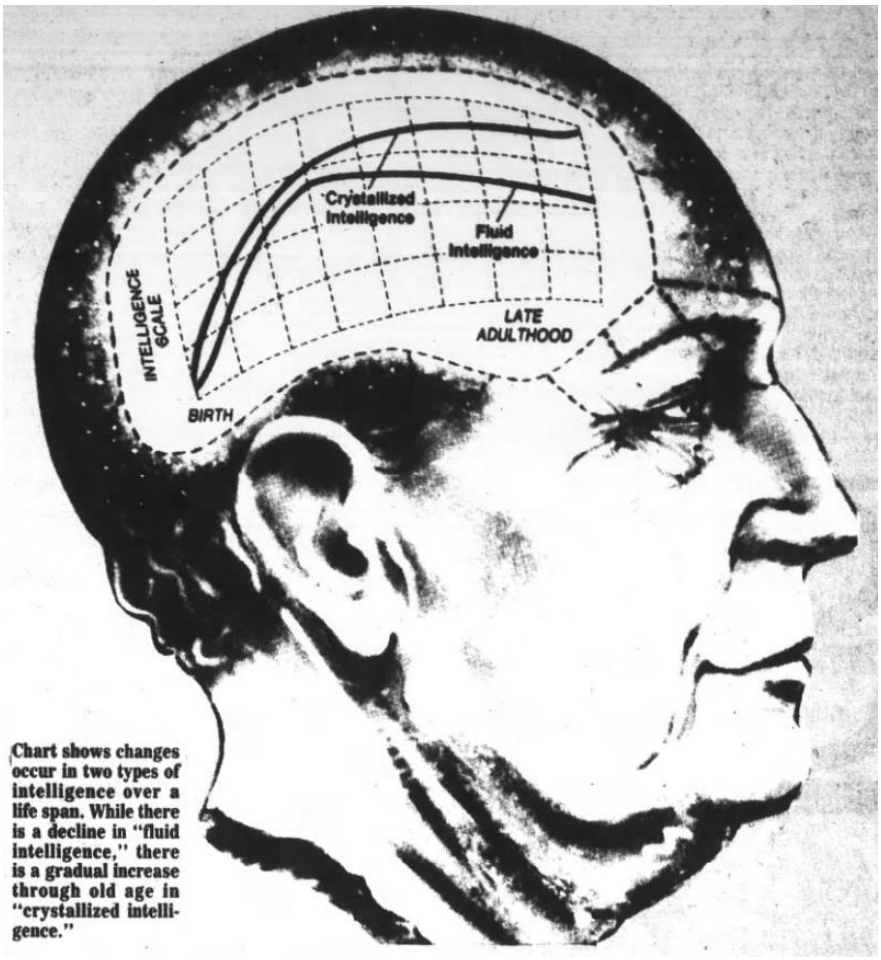
The heart of Horn and Donaldson's critique was that the SLS, as an observational study, could only describe differences between the young and old; it could not explain whether these differences are caused by age or generational differences. Ideally, scientists would test Schaie's hypothesis in a randomized controlled experiment, where different treatments are assigned to different participants. But it is impossible to assign an age or a cohort to a research subject.

[37] John L. Horn and Gary Donaldson, "On the Myth of Intellectual Decline in Adulthood," *American Psychologist* 31, no. 10 (Oct. 1976): 701-19; Paul B. Baltes and K. Warner Schaie, "On the Plasticity of Intelligence in Adulthood and Old Age: Where Horn and Donaldson Fail," *American Psychologist* 31, no. 10 (October 1976): 720-25; John L. Horn and Gary Donaldson, "Faith Is Not Enough: A Response to the Baltes-Schaie Claim That Intelligence Does Not Wane," *American Psychologist* 32, no. 5 (May 1977): 369-73; K. Warner Schaie and Paul B. Baltes, "Some Faith Helps to See the Forest: A Final Comment on the Horn and Donaldson Myth of the Baltes-Schaie Position on Adult Intelligence," *American Psychologist* 32, no. 12 (December 1977): 1118-20.

These criticisms could apply to any longitudinal study, and in fact some of them had been anticipated in Schaie's earlier writings. For example, Schaie understood that the retesting effect might improve subjects' scores on later rounds of a longitudinal study, but claimed that the seven-year gap between SLS rounds was long enough to mitigate this effect. All four participants agreed that the longitudinal study, even when modified with Schaie's sequential strategies, was an imperfect tool for measuring developmental change; they disagreed on whether it was good enough to justify making definitive claims about the myth of cognitive decline.

Horn and Donaldson then went a step further, claiming that an alternative developmental model—the theory of fluid and crystallized intelligence—better explained the data gathered by the SLS. This theory had been introduced by Cattell and refined by Horn.[38] Like the Thurstones, Cattell used factor analysis to divide intelligence into its component parts. But he was less interested in use-dependent faculties than in how the components of intelligence changed over time. Thus he introduced two factors: fluid intelligence and crystallized intelligence. Fluid intelligence represents a person's ability to form new logical connections and make sense of new situations, and is largely determined genetically. It peaks early in life (Cattell first claimed around 15 years old, but later revised this to around 30), followed by inevitable decline. Crystallized intelligence represents the abilities acquired through education and experience, such as breadth of vocabulary or use of advanced mathematical operations. Depending on a person's education, crystallized intelligence might peak or plateau at different points in life.

[38] Raymond B. Cattell, "Theory of Fluid and Crystallized Intelligence: A Critical Experiment," *Journal of Educational Psychology* 54, no. 1 (1963): 1-22; John L. Horn and Raymond B. Cattell, "Age Differences in Fluid and Crystallized Intelligence," *Acta Psychologica* 26 (1967): 107-29.



According to Horn, fluid intelligence peaks in early adulthood and then declines, while crystallized intelligence can continue to grow throughout life. Note the use of phrenological imagery. (Daniel Goleman, "Intelligence doesn't necessarily decline in old age," *Index Journal* (Greenwood, SC), March 18, 1984, p. 6C.)

The vitriol of this debate is surprising, given the close professional connections and shared methodological commitments between the four participants. Horn had been a participant at the 1969 WVU conference, presenting research on fluid and crystallized intelligence; clearly the two camps were familiar with each other's research.[39] Like Horn, Schaie and Baltes were influenced by Cattell's earlier research on intelligence and factor analysis. One of their colleagues at WVU, John Nesselroade, was another of Cattell's student. The PMAs and fluid and

[39] John L. Horn, "Organization of Data on Life-Span Development of Human Abilities," in Goulet and Baltes, *Life-Span Developmental Psychology*, 424-66.

crystallized intelligence are, in fact, mathematically consistent with one another: factor analysis allows intelligence to be split into two or seven or any number of factors, which can then be recombined in different ways. Unlike the famous psychometric debates of the early-twentieth century, this was not a dispute between fundamentally different conceptions of the mind, but rather a disagreement over emphasis and what constituted a responsible inference from imperfect data.

Part of this difference in emphasis can be explained by the context in which Cattell developed fluid and crystallized intelligence. Although Raymond Cattell was one of the most prolific and influential psychologists of the twentieth century, he remains controversial for his lifelong commitment to eugenics.[40] In the 1930s, Cattell used psychometrics to test the dysgenic prediction that average intelligence will decline over time, as people with less desirable genes have more children. Testing the intelligence of ten-year-olds in Leicester and Devonshire, England, he found a significant negative correlation between a child's IQ and number of siblings.[41] But his prediction of declining intelligence did not hold. After World War II, Cattell conducted a follow-up study on these two populations and found that the average IQ had increased by at least one point. Unwilling to abandon his beliefs on family patterns and heredity, Cattell theorized that his tests confounded inherent intelligence with the effects of schooling. Dysgenic decline was taking place, but this was counteracted by increases in schooling over time.[42] In the language that Cattell would develop over time, dysgenics affected fluid intelligence, but not crystallized intelligence.

---

[40] William H. Tucker, *The Cattell Controversy: Race, Science, and Ideology* (Champaign: University of Illinois Press, 2009).
[41] Raymond B. Cattell, "Is National Intelligence Declining?" *Eugenics Review* 28, no. 3 (1936): 181-203; *The Fight for Our National Intelligence* (London: P.S. King and Son, 1937).
[42] Raymond B. Cattell, "The Fate of National Intelligence: Test of a Thirteen-Year Prediction," *Eugenics Review* 42, no. 3 (1950): 136-48.

Despite the mathematical and methodological near-identity of Schaie's and Cattell's scientific projects, their research led to completely different politics. Factor analysis could be used to divide the mind up into different applications of intelligence, as in the PMA system, or to divide it into different sources of intelligence, as in fluid and crystallized intelligence. These were not incompatible models of the mind, but rather choices about what qualities of the mind were important to measure. Their choice of factors made it easier for Schaie to think in terms of capability to work and for Cattell to think in terms of genetic changes. Cattell's research design, comparing random samples of ten-year-olds in the 1930s and again in the 1940s, was an example of what Schaie would later call a time-lag study. As Schaie explained in his general developmental model paper, this design confounded environmental changes over time (such as increases in the number of years spent in school) with genetic changes between generations. Different researchers interpreted such ambiguities in line with their politics. Cattell emphasized genetic shifts; Schaie and Baltes emphasized environmental differences between generations; Horn and Donaldson made a point of avoiding interpretation as far as possible, effectively critiquing the LSDP researchers for "politicizing" their science.

In sum, neither factor analysis nor particular research designs led scientists inevitably to eugenicist or antidiscriminatory politics. Yet the historical and formal overlap between Schaie's and Cattell's research is important to understanding their policy recommendations. Their political beliefs were far apart, but they expressed their beliefs in a common language: human behavior can be measured quantitatively using psychological tests, this behavior changes in measurable ways over the lifespan and between generations, and behavior can be factored into components that change at different rates. These shared assumptions pushed Schaie and Cattell to express their political beliefs as narratives of change over time. Cattell did not simply believe

that some genetic stock was more desirable than others, but that the state ought to intervene and prevent dysgenic decline between generations. Schaie did not simply believe that stereotypes about the elderly were harmful, but that these stereotypes could be explained by shifts between generations that made comparisons between young and old unfair. This was the argument Schaie presented to Congress in 1980 and which was put into law with an amendment to the ADEA.

Conclusion

The 1976-77 debate marked the peak of LSDP confidence, with the field's two most prominent researchers making strong claims about intellectual maintenance in popular and professional venues. This confidence would not last. Although Schaie and Baltes remained friendly colleagues, their views of aging would drift apart. Schaie remained committed to the central claims of the SLS, characterizing intellectual decline as a myth throughout his career. But Baltes shifted away from this optimism beginning in the 1980s, accepting some of Horn and Donaldson's arguments for the inevitability of decline.[43] Instead, he began to investigate how the elderly could adapt to decline by making use of their accumulated life experiences—relying on crystallized rather than fluid intelligence, essentially.[44] Like advocates for other marginalized groups, the LSDP researchers split over whether to portray elderly people as the same as the young or as different, but still deserving of respect. The SLS's background in psychometrics forced Schaie to make direct quantitative comparisons between the young and old and helps explain his commitment to the equality position.

---

[43] Paul B. Baltes, "Autobiographical Reflections: From Developmental Methodology and Lifespan Psychology to Gerontology," in *A History of Geropsychology in Autobiography*, ed. James E. Birren and Johannes J. F. Schroots (Washington, D.C.: American Psychological Association, 2000), 21-22.

[44] Paul B. Baltes and Margret M. Baltes, "Psychological Perspectives on Successful Aging: The Model of Selective Optimization with Compensation," in *Successful Aging: Perspectives from the Behavioral Sciences*, ed. Paul B. Baltes and Margret M. Baltes (Cambridge: Cambridge University Press, 1990), 1-34.

The other legacy of the debate with Horn and Donaldson was to highlight the importance of intervention in producing authoritative knowledge. Much of their critique of the SLS rested on its status as an observational study rather than an experiment. After the debate, Baltes committed himself further to research on compensatory education for the elderly, running the ADEPT (Adult Development and Enrichment Project) study with Sherry Willis in the late 1970s.[45] While Baltes turned away from the intellectual maintenance thesis, Willis did not; she married Schaie in 1981 and brought compensatory education into the SLS. Beginning with the 1984 study, the SLS combined its by-now traditional psychometric battery with training programs designed to improve subjects' test scores over successive seven-year gaps. While earlier research ended up in Congressional debates over retirement, this strain of LSDP research has been cited by the contemporary brain training industry. Lumosity, the largest and one of the earliest brain training programs, relied heavily on LSDP intervention studies in its initial pilot study.[46]

Life-span developmental psychology did not transform the social sciences. Although one can find social scientists today who follow a "lifespan approach," this does not constitute a cohesive field as it did in the 1970s. Methodologically, Schaie's influence on social science is faint but identifiable: his 1965 general developmental model paper is still cited today. Schaie introduced the term "time-lag" design to refer to studies such as Cattell's on dysgenic decline, and that coinage has stuck. The relationship between aging and intelligence is still up for debate: since the 2000s, the psychologist Timothy Salthouse has argued that cognitive decline begins before age 30 (returning to Wechsler's position from the 1940s), criticizing the SLS specifically

[45] Rosemary Blieszner, Sherry L. Willis, and Paul B. Baltes, "Training Research in Aging on the Fluid Ability of Inductive Reasoning," *Journal of Applied Developmental Psychology* 2, no. 3 (1981): 247-65.
[46] Michael Scanlon, David Drescher, and Kunal Sarkar, "Improvement of Visual Attention and Working Memory through a Web-based Cognitive Training Program," Lumos Labs, 2007.

for failing to account for the retesting effect in longitudinal studies.[47] In the 1970s, popular magazines reported that Schaie's longitudinal studies had overturned the received wisdom of early cognitive decline; in the 2000s, they reported that Salthouse's cross-sectional studies had overturned the received wisdom of late cognitive decline.[48]

But the SLS is important because it reveals how social scientists are forced to reckon with time at different scales. Moving from the very young to the very old required creative and improvisational use of familiar tools, which played out in the piecemeal construction of the SLS over twenty years. Some tools had to change; others had to stay the same. And this thinking through time is evident in the contemporary legacies of LSDP. The brain training industry, which emerged out of Baltes's notion of "compensatory education" for the elderly, suggests a habitual, disciplined conception of time—users are encouraged to use apps or complete puzzles daily, encapsulated by the mantra "use it or lose it," repeated endlessly as advice for maintaining cognitive health.[49] The contemporary focus of cognitive aging research is Alzheimer's disease prevention, which, researchers believe, can begin early in life. By building up "cognitive reserve" in youth, a person might stave off the worst effects of cognitive decline.[50] Of course, testing this theory would revive all the methodological questions about comparing the young and old that Schaie introduced over fifty years ago. On the macro and micro scales, temporality

[47] Timothy A. Salthouse, "When Does Age-Related Cognitive Decline Begin?" *Neurobiology of Aging* 30, no. 4 (2009): 507-14; K. Warner Schaie, "Salthouse Again Reifies the 'Cross-Sectional Fallacy,'" *Neurobiology of Aging* 30, no. 4 (2009): 528-29.
[48] "'Brain decline' begins at age 27," *BBC News*, March 16, 2009. http://news.bbc.co.uk/2/hi/health/7945569.stm. Accessed May 10, 2023.
[49] Henry M. Cowles, "Habit and the Limits of History," *Reviews in American History* 48 (2020): 191-96; Melissa Healy, "Your aging brain: Is it 'use it or lose it'?" *Los Angeles Times*, December 11, 2018, https://www.latimes.com/science/sciencenow/la-sci-sn-aging-brain-mental-exercise-20181210-story.html. Accessed May 11, 2023.
[50] William S. Kremen et al, "Influence of young adult cognitive ability and additional education on later-life cognition," *Proceedings of the National Academy of Sciences of the United States of America* 116, no. 6 (2019): 2021-26.

continues to structure how we understand mental fitness and our personal responsibility to maintain it.