Jensen on "Jensenism"

ARTHUR R. JENSEN University of California, Berkeley

Though "Jensenism" is a term listed in several dictionaries. Arthur Jensen has produced a more extensive body of work than suggested by the dictionary entry. To the public, he is mainly known for his work on the genetics of intelligence. This article discusses the work that is publicly less well known. Work discussed includes studies in learning, memory, the cumulative deficit hypothesis, Spearman's hypothesis, and speed of information processing, to name a few. The publicly better known work is also discussed. A bibliography of Jensen's publications is included in an appendix. (Abstract written by D. Detterman)

To discover that one's name has entered the dictionary as an "ism" is both flattering and embarrassing, and is cause for reflection. I know because it happened to me. Recent editions of a number of dictionaries contain the word "Jensenism." The Random House and Webster's Unabridged Dictionaries, for example, contain the following entry:

Jen-sen-ism (jen'se niz'em), *n*, the theory that an individual's IQ is largely due to heredity, including racial heritage, [1965-1970]; after Arthur R. Jensen (born 1923), U.S. educational psychologist, who proposed such a theory; see -ism]—Jen'sen-ist, Jen'sen-ite', *n., adj.*

For those who understand the meaning of heritability in quantitative genetics, the wording is rather inept and the "theory" attributed to me has been around at least since the time of Francis Galton (1822-1911), whose *Hereditary Genius* (1869) predated the very article that led the popular press to label me a "hereditarian" by exactly one century. The dictionary definition can't be overly derided, however, as it is quite true that, in 1969, I did present a fairly comprehensive review of the evidence that IQ is substantially heritable and had stated that it is a reasonable hypothesis that genetic as well as environmental factors are involved in the well documented Black-White average difference in IQ. Also, I like to think that I was partly responsible for getting Galtonian thinking back on track in differential psychology after it had been derailed in the behavioral sciences for at least a generation following World War II (the period dominated by what Sandra Scarr once referred to as "naive environmentalism").

INTELLIGENCE 26(3): 181-208 ISSN: 0160-2896

Direct all correspondence to: Arthur R. Jensen, School of Education, University of California, Berkeley, Berkeley, CA 94720-1670.

Copyright © 1998 by Ablex Publishing Corporation All rights of reproduction in any form reserved.

However, the more serious disadvantage of having one's name turned into an "ism" is that, from that moment on, one is liable to be identified only as the "ism" in the dictionary. The rest of one's research activity can be unfairly eclipsed, and findings and formulations that are unique and perhaps even fundamentally more important are forgotten. One of my aims here is to forestall this threatened eclipse of other aspects of my research and shine some light on how that which got me labelled as an "ism" fits into the larger orbit of my lifetime's work.

Essentially, I have always been a differential psychologist. Human idiosyncracies and individual differences in behavior interested me before I had ever heard of psychology. The first book I read on the subject, more or less by accident while in high school, was J.B. Watson's Psychology From the Standpoint of a Behaviorist (1929). It was probably the main reason I chose to major in psychology in college, after reluctantly but realistically deciding not to pursue a career in music. Though I became acquainted with some wellknown psychologists, such as Edward Tolman and Egon Brunswik, as an undergraduate psychology major at Berkeley, the one psychologist whose work most captured my attention (but whom I never saw in person) was the then Sterling Professor of Psychology at Yale, Clark L. Hull, a latter-day Watsonian and Pavlovian behaviorist. One could say that I became a Hullian, and I recall writing a long term paper for one of my courses extolling Hull's theory of learning-excessively so, according to the comments of the TA (one of Edward Tolman's graduate students) who graded my paper. Primed, I suppose, by Watson, I was especially attracted to Hull's purely mechanistic system for explaining behavior, as spelled out in his Principles of Behavior (1942). B.F. Skinner's Behavior of Organisms (1938) was also appealing but lacked the systematic theoretical system that made Hull's approach seem more promising to me.

I was totally unaware at the time that these now classic works in psychology, and indeed my whole undergraduate education in psychology, neglected individual differences and the influence of genetic factors on behavior. These topics were scarcely admitted as part of the field of psychology, at least as it was presented at Berkeley in the 1940s. Experimental psychology dominated the department at that time, and the implicit assumption of experimental psychology was that individual differences in the behavioral realm originated entirely outside the organism, through its exposure to different environmental contingencies, and they could be explained, if one were at all interested in doing so, in the purely stimulus-response-reinforcement terms of conditioning and learning. In its focus on discovering general laws or principles of behavior, experimental psychology traditionally treated individual differences as a nuisance variable, or as merely error variance in the statistical analyses of its data.

This limited perspective of my undergraduate courses in psychology was extremely implicit and so completely taken for granted that it did not enter my consciousness until some years later. I occasionally meet psychologists even today who think of individual differences as error variance or as purely a product of environmental diversity. I was still largely operating on this assumption in 1964 when I wrote a major paper that attempted to explain social class differences in scholastic learning entirely in terms of the then current S-R theories and principles of verbal learning (67). Ironically, the publication of that paper was so long-delayed that it appeared *after* my position on the major basis of individual and group differences had changed in a hereditarian direction. Large differences in the publication lag of one's articles and book chapters during certain periods may even create a false

impression of contradictory vascillations in one's theoretical stance. The publication dates of one's articles are not always perfectly correlated with the actual chronology of one's changing position on theoretical issues.

Thanks to the beautiful "recreational reading" room (the Morrison Library) on the Berkeley campus, where I spent most of my evenings, I believe I got as much or more of my undergraduate education from entirely self-selected extracurricular reading as I got from my courses and textbooks. The most lasting influence I recall are works by M.K. Gandhi, Bertrand Russell, G.B. Shaw, Havelock Ellis, H.G. Wells, Aldous Huxley, Alfred Korzybski, and biographies of famous musicians, scientists, and philosophers. They instilled a certain critical sense as well as a humanistic idealism that, in the long run, made a greater impression on me than did most of the relatively uninspiring textbooks I was required to read in my courses. To make more time for the reading I most liked, it was my policy never to read anything in my college textbooks more than once.

It was my extracurricular reading, probably more than anything else, that led me to look for the ways psychological science might have practical applications that could benefit individuals and society. Some years later when I decided to enter graduate school to work toward a Ph.D., I examined various university catalogues to see what they offered in applied psychology. I recognized more of the names of psychologists whose works in applied areas, such as clinical and educational psychology, that I had already come across in my reading on faculty of Teachers College, Columbia University than in any other university catalogue. Egon Brunswik's course on the history of psychology had also left me with a distinct impression of Columbia as having one of the pioneer departments of psychology, shaped by such luminaries as James McKeen Cattell, E.L. Thorndike, and Robert S. Woodworth. (When I arrived at Columbia, Woodworth was still lecturing at age 87, and I audited his two courses.) The fact that Columbia University is located in New York City, home to Carnegie Hall, Toscanini and the NBC Symphony, the New York Philharmonic, and the Metropolitan Opera, provided a powerful added attraction. The musical capital of America, New York is visited each year by many of the world's greatest orchestras, conductors, and virtuosos. And my interest in music has never been second to my interest in psychology, though I have necessarily devoted more time to the latter, of course, since it has been my livelihood. When I wasn't on the Columbia campus, chances are I was hanging out in Carnegie Hall, either at a concert or a rehearsal.

At Columbia's TC I studied educational, clinical, and personality psychology. My Ph.D. dissertation (under Professor Percival Symonds) was on the Thematic Apperception Test as a measure of aggression (2,3).¹ I found my three years as Symonds' research assistant much to my liking. However, my interest in the subject of his research at that time, based on the psychoanalytic or "dynamic" interpretation of various projective techniques (8, 20), proved short-lived. Though Symonds was a man of noble character and in many ways a fine mentor, my three years at TC were probably influenced more by the lingering shadow of the psychologist who had been Symonds' mentor but who had died three years before I arrived at TC—Edward L. Thorndike, probably America's greatest psychologist. Thorndike's influence and his conception of psychology still pervaded the intellectual atmosphere at TC during my tenure and was repeatedly reinforced by an imposing portrait of the great man that hung on the wall above the card-catalogues in the TC library. I felt compelled to read some of Thorndike's books and I liked them a lot, especially for their

clear thinking and their objective and empirically anchored approach to the remarkably broad range of subject matter in psychology with which he dealt.

It is amazing how much of what today is viewed as established fact in psychology was either discovered or presaged by E.L. Thorndike. As he was one of the leading pioneers of psychology as a natural science, he became the first of my "heroes" in psychology; the other two (Galton and Spearman) I discovered a few years later. These are the three psychologists whose key works I return to and re-read for their wealth of hypotheses, original and insightful ideas, and inspiration, always to be rewarded. If there have been any authentic geniuses in the history of scientific psychology, in my estimation they include at least these three. (I have written about Galton [238, 352, 383] and Spearman [239, 353, 383].)

During my clinical internship at the University of Maryland Psychiatric Institute in Baltimore, I examined a great variety of psychiatric patients, using all of the prevailing techniques of clinical psychology, and typically wrote a clinical report on each patient. During my training experience in psychotherapy. I quickly came to realize that I was less satisfied and less effective working with people directly than in working with data. I did not enjoy many of the routine aspects of clinical work, probably because I am quite low in extraversion. Hence I welcomed collaboration with one of my clinical supervisors in some research we did on the Rorschach as an index of pathological thinking that completely eschewed the traditional systems of scoring Rorschach protocols and was solely based on characteristics of the subject's verbal expression (5; see also 39, 46).

It was also at this time that I began seriously reading the books and articles of Hans J. Eysenck, who was then a maverick personality researcher and the professor of psychology in the University of London's Institute of Psychiatry. Most of my evenings that year were spent reading every book and article by Eysenck that I could find in the university library, as well as many of the references he made to the influential work of others, particularly Galton, Spearman, and Thurstone. This provided a much needed antidote to the predominantly Freudian or psychoanalytic concepts that informed my clinical work. It was not so much the specific aspects of Eysenck's own theories or his research, but rather his general approach to psychology as a natural science that provided my first real sense of finally having discovered my true vocation. I felt I was no longer groping for the path that I should take to make my life seem worthwhile. I believed that research and teaching in the field of differential psychology, broadly conceived, was exactly the path for me. So I wrote to Eysenck for his permission and applied to NIMH for a postdoctoral fellowship to spend a year in Eysenck's department in London. Luckily, both were granted and, with a year's extension of my fellowship, I had the good fortune to spent two full years with Eysenck. (Six years later I returned to his department as a Guggenheim Fellow for another full year during my first sabbatical leave from Berkeley [detailed in 149, 378]).

I emphasize my postdoctoral work with Eysenck, because I believe it planted the seeds of virtually everything I have done since then. It put me on the path that I have followed, in one way or another, for all of my later research. Although each of the many subsequent byways could not have been anticipated, they all led more or less consistently in one general direction—what came to be known as the London School of differential psychology, originated by Galton and with Spearman, Burt, and Eysenck successively as its leading exponents (283, 376, 377). (I knew personally only Eysenck [378] and Burt [126, 225, 326, 367].) The London School is not really a school or even a doctrine or a theory. Rather, it is a general view of psychology as a *natural science* and as essentially a branch of *biology*.

Its central concern is variability in human behavior. It is Darwinian in that it views both interspecies variation and an important part of intraspecies variation (both individual and group differences) in certain classes of behavior as products of the evolutionary process. It is behavior-genetic in that the evolutionary process depends upon genetic variation and selection, and the neural basis of behavioral capacities is subject to these evolutionary mechanisms the same as other physical characteristics. It is *quantitative* in that it emphasizes the objective measurement and taxonomy of behavior and the operational definition of latent traits or hypothetical constructs. It is *analytical* in that it subjects quantitative data to mathematical formulation and statistical inference. It is experimental in that it typically obtains measurements, both behavioral and physiological, under specifically defined and controlled conditions. It is *reductionist* in that it aims theoretically to explain complex phenomena in terms of simpler, more elemental processes. It is *monistic* (as opposed to dualistic) in that it neither posits nor seeks any explanatory principle that does not consist of strictly physical processes; it views complex psychological phenomena as emerging solely from interactions among more elemental neurophysiological processes and their past and present interactions with environmental conditions.

Besides the extensive reading, studying (courses in multivariate statistics with Patrick Slater and factor analysis with A.E. Maxwell), and writing (4, 6, 7, 9, -14) that occupied my time as a postdoc, I undertook laboratory research on individual differences in the effects of massed and distributed practice in selective stimulus-response learning and I devised a special apparatus for the directly measuring individual differences in reactive inhibition independent of any form of learning per se (51). The specific hypotheses I tested derived from Eysenck's theory of the basic neural processes responsible for individual differences in extraversion-introversion (or E, as it was called), which had been established as a continuous unitary trait by the factor analysis of personality inventories, behavioral ratings, and objective behavior measurements derived from certain laboratory techniques. Eysenck's theory of E at that time brought me back to Clark Hull's theory of learning, which had first fascinated me as an undergraduate. I became a born-again Hullian, this time around becoming more thoroughly versed in every facet of Hull's theory and most of the theoretical and empirical literature related to it, including Pavlov's classic work on conditioning.

Eysenck's theory held that the basis of E is the rate of build-up of a hypothetical neural process called *reactive inhibition*, or $I_{\rm R}$ (as defined in Hull's system). The theory contends that trait extraversion reflects a more rapid build-up and a slower spontaneous dissipation of I_R under the conditions in which I_R is hypothesized to be manifested in behavior, such as the experimental extinction and spontaneous recovery of conditioned responses, the effect of massed trials versus spaced trials in serial rote learning, and the reminiscence effect in motor learning (as on the pursuit-rotor). It was this aspect of Eysenck's research program that led me into theories of learning and the experimental psychology of human learning, which soon completely overshadowed my interest in personality research. I saw the study of individual differences in learning in its own right as a more fundamental and scientifically researchable subject than the study of personality. The last postulate in Hull's behavior theory in its final form (in 1952) states that individual differences in learning, or excitatory potential $({}_{S}E_{R})$, are a net product of individual differences in each of the hypothetical constructs in his system, such as habit strength $(_{S}H_{R})$, drive (D), reactive inhibition (I_R) , conditioned inhibition $({}_{S}I_R)$, sensory limen or threshold of stimulus activation $({}_{S}L_R)$, and spontaneous oscillation of reaction potential (${}_{S}O_{R}$). I thought this approach provided

the basis for a rigorous, quantitative and experimental approach to the study of individual differences in human learning. I later elaborated on this idea in a paper that, in retrospect, strikes me as an overly ambitious and practically unfeasible program for research on individual differences in human learning (59). Since the largest part of the individual differences variance in the forms of learning that are important for education and the acquisition of many other real-life skills is factorially indistinguishable from Spearman's g, or general intelligence, 1 now believe a program of research on the nature of g to be probably more fruitful than focusing on learning per se (189, 301).

But before going on with my story. I should point out what may not be well known to younger readers, that Hull's system, which dominated the learning field from about 1940 to 1960 (he died in 1952), waned rapidly in the early sixties and became virtually extinct by 1970. Since then, Hull's citation index has hovered close to zero. This is a remarkable fate for one who, for over a period of at least twenty years, many considered the leading theorist in scientific psychology. In marked contrast, E.L. Thorndike, 48 years after his death, remains among the 100 most frequently cited psychologists in recent decades, and the number of citations of Charles Spearman has increased in each decade since his death (in 1945) and risen most rapidly since 1970.

Thorndike's and Spearman's intuitions, hypotheses, theories, and the phenomena on which their interests were focused, mainly learning and cognitive ability, were evidently more important, more original, and scientifically more productive than Hull's precisely formulated theory of learning, however impressive his achievement seemed in its day. The problem, I think, was not Hull's in use of the hypothetico-deductive method, which I believe was exemplary, but that the many interrelated parts of his whole grand theoretical edifice, its postulates (as Hull called them), were erected on too slim a foundation of empirical studies. Hence the subsequently growing number of experiments inspired by the theory and devised to test it increasingly failed to confirm its predictions. Though modifications and additional ad hoc principles were proposed to meet the explanatory demands imposed by new empirical evidence, Hull's system gradually collapsed beyond repair and was eventually discarded, much like the geocentric theory in astronomy and the phlogiston theory in chemistry. In the history of science, of course, this is a perfectly respectable demise for a theory. The really fatal shortcoming of Hullian theory, however, was its nonbiological behaviorism, a position that was bound ultimately to leave it theoretically barren.

Rather early in my career, while still a Hullian, I tried to modify Hull's theory to make it accommodate some of the contradictory experimental evidence by proposing a fundamental mathematical reformulation of the treatment of reactive and conditioned inhibition within the Hullian framework (18). But this kind of ad hoc doctoring could not save Hull's system any more than postulating retrograde motion of the planetary epicycles could preserve Ptolemy's geocentric theory. Though I gradually lost interest in Hull's theory, my interest in human learning, particularly in its individual differences aspect, was undiminished.

Now that the stage is set, with a backdrop of the values and attitudes against which all my later activity can be more understandable, I will provide a brief account of the specific studies that I believe mark the key points in my research activity, and how and why I moved from each point to the next. Studies never arise from thin air, of course, but also one does not have to go looking for things to research. Each new project, it seems, is absolutely compelled by the preceding ones, or by one's purposeful and critical reading of the literature, or by one's self-criticism and others' criticisms of one's previous work. The completion of each study always leave some loose ends. Problems abound and one continually searches for what seems the most fruitful path toward each problem's solution. Given the pages allotted me, this account is necessarily quite telegraphic, referencing only my main publications on each topic. A perusal of my bibliography (see Appendix) indicates that my publications fall into six main categories: (*i*) clinical and personality, (*ii*) human learning, (*iii*) behavior genetics, (*iv*) racial- cultural differences, (*v*) test bias, and (*vi*) mental chronometry and g theory. (I will ignore the first category, with some dozen or so articles, which in retrospect I consider of much less importance or interest.)



Figure 1. A typical serial position curve for a 9-time list, showing the percent of total errors made in learning the items at each position up to a criterion of mastery (i.e., the first errorless trial).

Human Learning

Serial Rote Learning. As I was hired by the University of California at Berkeley specifically to teach courses in human learning and mental testing, it was appropriate that I also continue the research in this area that I had begun while in London. I was given a laboratory near my office and funds which I used to devise an electrical apparatus that I thought had essential advantages over the conventional memory drum used in those days for research on human learning. My device was programmable via punched tape so it could automatically present stimuli (back-projected on a disdplay screen) at any rate, in any desired order, from trial-to-trial (29). I used it in a great many experiments, only abandoning it years later when dependable microcomputers became purchasable at a reasonable price.

One prediction from Eysenck's Hullian-based theory of extraversion, which made use of Hull's concept of reactive inhibition (I_R), was that, since extraverts build up I_R faster than do introverts, extraverts should show a more steeply peaked serial position effect in serial rote learning. The serial position effect produces the distinctive serial-position curve (shown in Figure 1) that plots the number of errors made as a function of the position of each of the to-be-learned items during the courses of the learning trials that precede the first error-free trial. Hull had theorized that the serial position curve resulted from the differing amounts of I_R built up at each of the item positions, being greatest near the middle position in the serial order and decreasing toward the beginning and end of the series.

An experimental test of Eysenck's prediction completely failed to bear out his hypothesis. Introverts and extraverts had virtually identical serial position curves. But this finding could not invalidate Eysenck's theory unless it could also be shown that two other conditions were true, namely, (a) that there are reliable individual differences in the shape of the serial position curve (SPC), for whatever reason, and (b) that Hull's theory is correct that the distinctive shape of the SPC is the result of a differential build-up of I_R on the items at each of the the different positions in the list during the course of learning them. In a series of experiments, I found that, although there were reliable individual differences in the number of trials subjects needed to learn a serial list to a criterion of mastery (i.e., one errorless trial) and in their total errors prior to mastery, there were no significant individual differences in the peakedness of the SPC per se (24). That is, when the number of errors occurring at each position is converted to the proportion of total errors at each position, the height of the SPC's peak becomes virtually the same for all subjects. The only feature of the curve that shows reliable individual differences is its degree of skewness; that is, the departure of the SPC from perfect symmetry (see Figure 1), as determined by which position accumulates the largest proportion of errors (36). In a 9-item list, for example, the peak for some subjects falls at position 5; for others, at position 6; and for a relatively few subjects, at position 7. These individual differences in skewness had high test-retest reliability across different serial learning tasks.

So my interest shifted from individual differences in the peakedness (technically, the kurtosis) of the SPC to individual differences in the skewness of the SPC. In a series of further experiments, I tested three new hypotheses:

(*i*) The items (of a given type) in a serial list are learned in a single trial, the number of trials required being a function of the number of items in the series, the item's position in the series, and the subject's overall rate of learning.

JENSEN ON "JENSENISM"

Indeed, my data showed that the backward learning curves (i.e., the course of learning plotted trial-by-trial from the criterion trial of mastery backwards to the first learning trial) for each item, regardless of position, are virtually flat and at the chance level up to the trial on which the item is learned, thereafter maintaining a high probability of correct response throughout all subsequent trials (22).

(*ii*) The serial position effect simply reflects the rank *order* in which the items in the series are learned (which is different from their serial order of presentation). The most typical order of learning the items in, say, a 9-item list, is as follows:

Position in series: 1 2 3 4 5 6 7 8 9

Order of learning: 1 2 9 3 8 4 7 5 6

The proportion of the total errors accrued at each position is directly related to the order of learning the item at that position (22). The number in **boldface** indicates the position at the peak of the SPC.

(*iii*) Individual differences in the degree of skewness of the curve are a positive function of the subject's forward memory span, that is, subjects with a larger memory span tended to learn the first two (or three) items in the series on a single trial. Thus the order of learning for persons with different memory spans is as follows:

Serial Position: 1 2 3 4 5 6 7 8 9

Order of learning for

Short memory span: 1 9 2 8 3 7 4 6 5 Average memory span: 1 2 9 3 8 4 7 5 6 Long memory span: 1 2 3 9 4 8 5 6 7

Note that the peak of the SPC moves further beyond the middle position (#5) as memory span increases. In fact, the only SPCs that were symmetrical (i.e., those with their peak consistently in the middle position) were those produced by mildly retarded persons, with IQs below 75, a group for whom a short memory span is typical. We found that even among bright college students, individual differences in the skewness of the SPC are significantly related to individual differences in memory span (measured by forward digit span $\hat{a} \, la$ the Wechsler Digit Span test). All these studies are detailed in a series of articles (22, 24, 33, 36).

The next research taskwas to discover what determined the order in which items were learned. Order was a very stable phenomenon. The examples above show the distinctive pattern in the order of learning. The first one (or two) items are learned, then the last item, than the item adjacent to the first (or second) than the item adjacent to the last, and so on. This systematic alternation "rule" held for every length of serial list, provided, of course, that the number of items in the list exceeded the subject's immediate, one-trial memory span. In my reductionist effort to explain the SPC, I hypothesized that "new" (i.e., not-yet-learned) items had a tendency to "stick" to "old" items (i.e., those previously learned). I termed this hypothetical phenomenon the *adjacency* effect. To have any reductionistic explanatory value for the SPC, of course, the "adjacency effect" would have to be shown to be a more general phenomenon beyond the SPC per se and be manifested in other forms

of learning besides serial learning. It had to appear even in learning tasks that ruled out any possibility for serial learning.

To test the generality of the adjacency effect. I set up an experiment using free recall, in which 40-item lists of common nouns were presented (at a 2-second rate) in a completely different order on each learning trial. After each of the first three presentations, subjects were asked to recall freely as many of the nouns as they could remember. Then, on the fourth presentation, a random half of the forty nouns in the list were deleted and replaced by twenty "new" nouns, and subjects then had to recall freely as many of all the nouns

Trial	Positions Learned
1	1
2	1 -2
3	7-1-2
4	7-1-2-3
5	6-7-1-2-3
6	6-7-1-2-3-4
7	5-6-7-1-2-3-4

Figure 2. A temporal adjacency effect determining the order of learning a 7-item serial list in seven trials. The serial poisition of the last item learned on each trial is shown in **boldface**.

(both "old" and "new") presented on the fourth trial as they could. These recall data indeed showed a significant adjacency effect (41); that is, if a "new" (N) word was adjacent to an "old" (O) word, it had a higher probability of recall than if it were adjacent to another new word. The probability of recall of a "new" target word (N) decreased from the most to the least optimal condition for recall as follows: ONO, ONN, NON, and NNN.

The above experiment strongly supports the existence of the hypothesized *adjacency effect* in free recall, and that the effect is consistent with the hypothesized explanation for the SPC. In serial learning, each successive item that is learned on a given trial is the one that is *temporally* most adjacent to the last item learned. It is as if each "new" item waiting to be learned "sticks" to the item that was most recently learned on a previous trial. Figure 2 shows the order of learning items at a given position in a 7-item list (assuming a learning rate of one item per trial) as each "new" item "sticks" to the last item learned on the previous trial. Thus in a 7-item list, the peak of the SPC typically falls at position 5. The same algorithm probably applies to supra-span lists of any length, but has been tested only on serial lists of up to 13 items.

But why should a "new" item "stick" to an "old" item that is temporally adjacent to it? I hypothesized the cause was the time needed for processing and consolidating the information. When an "old" item follows a "new" item, it does not impinge on the full time needed to process the "new" item for retention; and when a "new" item precedes an "old" item its processing time can extend into the time period occupied by the "old" item, which has already been processed, and therefore does not distract or interfere as much with the processing of the "new" item. Processing an item for retention requires a certain amount of time, and individual's differ in the amount of time required, hence individual differences in overall learning rate for any given task.

At the time I published my study of the adjacency effect (41), I expressed certain reservations about its adequacy as an explanation of the SPC. In light of subsequent studies, however, I now think it is probably the best explanation of the serial position effect that I have yet seen. In retrospect I wish I had continued investigating it until it was more definitively proved (or disproved). But I got sidetracked by questions concerning serial learning phenomena that called for better understanding besides the serial position effect per se.

The many other experiments on serial rote learning were performed in my lab. They showed that serial learning does not consist of the stimulus-response chaining of items in the manner of paired-associates learning (25, 49). (Prior learned paired associates placed into a serial list do not afford any advantage or affect the shape of the SPC; and adjacent pairs of items taken from a previously learned serial list had to be re-learned when presented as a typical paired-associate learning task.) The mystery, then, was what exactly is learned in serial rote learning (49)? The typical SPC is manifested even when all of the items of a serial list are presented simultaneously for study (23, 31). Individual differences in personality factors, particularly neuroticism, or anxiety, interact with rate of presentation in determining the number of trial to attain a criterion of mastery (27). And, surprisingly, the probability of the location of the incorrect letters in misspelled words is predicted, on average, by the serial position effect (26). Also, we found that elementary school children, when tested at various stages of learning the alphabet, show that they learn it just as one learns a serial list; the learning proceeds from both ends of the alphabet, working toward the middle, which is the last part of the alphabet to be mastered.

One might ask why anyone should be interested in something so esoteric as the serial position effect, paired-associates learning, and related phenomena observed in the learning laboratory. My belief was (and still is) that psychology, to develop as a natural science, has to begin by trying to explain the simplest, most universal, and most reproducible behavioral phenomena. Scientific explanation is essentially reductionistic, showing the causal links by which the phenomenon to be explained is related to certain simpler, more elemental, and more general principles. An explanatory hypothesis, invoking simpler mechanisms and general principles, originates from hunches and inductive reasoning based on observations of a phenomenon: certain consequences of the initial hypothesis then are logically deduced from it and are empirically tested for their validity by systematically making new observations and performing controlled experiments. This procedure, if followed properly, gradually builds up a nomothetic network, or general theory, within which an increasing number of related phenomena can be explained. The problems is in discovering the basic and essential elements and their interrelations in a given behavioral domain. Unless one begins investigation with phenomenon that is reliable, reproducible, and objectively observable or measurable, there is little hope of developing a scientific explanation of it.

Most of the explanations of human behavior offered in popular psychology, such as one often sees on TV, when someone introduced as a psychologist provides a "psycholoigical explanation" of, say, why some celebrity committed adultery, got divorced, or did whatever made the tabloids are simply nonsense from a scientific viewpoint. They only reinforce skeptics who argue (alas, all too often correctly) that psychologists are the twentieth century's shamans and witch doctors. I'm always amazed to see a psychologist offering a glib explanation of some immensely complex behavioral individual incident when psychological science has not even provided explanations for comparatively simple general phenomena such as individual differences in optical illusions (e.g., rate of oscillation in the Necker cube), or the serial position effect, or why 12-year- olds have a longer digit span or a faster reaction time than 10- year olds, or why two individuals of the same age have reliably different digit spans. All these phenomena are universal, reliably observable and measurable, and amenable to exacting analysis in the laboratory. Indeed, many, if not most, of the simplest, most basic psychological phenomena mentioned in instroductory psychology textbooks have not yet been understood in the true sense of a scientific explanation. It may hardly matter with which dependable behavioral phenomenon one begins investigating. If reductionist thinking and empirical hypothetico-deductive procedures, along with considerable investiggative ingenuity, are applied, advances of a scientific nature will emerge.

I was occasionally asked by my faculty colleagues and my dean why I investigated such esoteric psychological phenomena that apparently had so little direct relevance to education. After all, my faculty appointment was officially as an educational psychologist in a graduate school of education. If my fields of teaching and research were human learning and individual differences, they wondered if my research shouldn't reflect some concern with the more obvious learning problems seen in the public schools. A theory of serial rote learning hardly qualified, and some might even say I was fiddling while Rome burns.

I took these hints from colleagues to heart when one of my graduate students, a school psychologist in a nearby district with a large Mexican-American population, came to me with questions about her experience in testing children who technically qualified for placement in special classes for the educable mentally retarded (EMR), with IQ below 75. From

193

her observations of these children's behavior and social interactions on the playground, she got the impression that the Mexican-American children did not appear to be as retarded as their EMR White "Anglo" classmates who had the same IQ and the same low level of scholastic performance as the Mexican-American children. If the Mexican-American and Anglo pupils, despite having the same IQ, really differed in ways that are relevant to their learning ability, should they receive the same kind of instruction in the same EMR classes? Here, I thought, was a chance to put my research on individual differences in learning ability to some (at least seemingly) practical use. Having some acquaintance with the literature claiming cultural bias in the standard IO tests. I supposed that these were most probably biased against many Mexican-American children, for both linguistic and cultural reasons, and that we should test these children (and, for comparison, also their Anglo counterparts) on greatly simplified forms of the serial and paired associate tasks used in my laboratory, as performance on such tests did not have to call call for any items of knowledge acquired outside the testing room. Until that time, the only subjects in my laboratory studies of learning were Berkeley undergraduates - not the most ideal population in which to study individual differences, particularly in cognitive abilities, as they have much less variance in IO than exists in the general population; extremely few score below the 75th percentile (IQ 110). Therefore, the prospect of getting research data from pupils in the public schools was enticing.

The Level I-Level II "Theory". My studies of Mexican-American and Anglo pupils used serial learning, paired-associates learning, and free-recall learning paradigms. The learning materials were actual highly familiar objects (e.g., spoon, comb, toy car, etc.), which all pupils could readily identify (in whatever language they preferred). Among pupils in EMR classes, the Mexican-Americans performed significantly better than the Anglos of the same age and IQ; many of the Mexican-Americans performed on a par with the Anglos of average IQ in regular classes (19). Yet the two groups performed about the same on completely nonverbal IQ tests that evinced no evidence of cultural bias. We seemed to have here two rather distinct classes of tests; one consisting of direct rote-learning tasks for which the Mexican-Americans in EMR classes outperformed their EMR Anglo classmates, and the other consisting of nonverbal reasoning tests in which the two groups performed about the same.

I tried the same learning tests in another elementary school with Black and White pupils, both in EMR and in regular classes, and found the same phenomenon I had found in the initial study with Mexican-Americans. The results were even more pronounced. It is not unusual to find Black children, even with IQ below 75, whose performance on rote-learning tasks is on a par with that of White children of average IQ. Yet these same children achieve poorly in school and perform at a low level when given relatively simple tests that involve some form of reasoning or problem solving, such as the Block Designs subtest of the Wechsler Scale and Raven's Colored Progressive Matrices.

In trying to explain these observations, I formulated the Level I-Level II "theory" (65, 66, 222). I put "theory" in quotes, because it was not really a theory but rather an empirical generalization. Level I tasks were those that required little or no mental manipulation of the input information to achieve correct response output. Level I tests showed individual differences in the the capacity for registering information and recalling it in much the same form in which it was presented. A clear example (and test) of Level I ability is forward digit

span (FDS), in which the subject recalls a series of digits, in order, after a single presentation. Level II, however, requires some transformation or mental manipulation of the input in order to arrive at the appropriate response. A clear example is backward digit span (BDS), in which the subject must take in a series of digits and then recall them in reverse order. Most of the items on typical IQ tests, especially items that call for what R.B. Cattell has termed "fluid" intelligence, involve Level II ability.

I also hypothesized a hierarchical relationship between Level I and Level II; that is, in a given person, high Level II ability was seldom, if ever, accompanied by low Level II ability, while low Level II ability was not uncommon in persons of high Level I ability. The latter condition, we found, is more prevalent among children of low socioeconomic status, especially Black children (157). On this basis I argued that there were two types of familial mental retardation, both showing an IQ below 70-primary, which consists of having both Level I and Level II abilities more than two standard deviations below the population mean, and secondary, in which only Level II ability is -20 below the mean, but Level I ability is in the normal range (88, 89, 113). This distinction had useful implications even in the low- average range of ability. Following a talk I gave about Level I-Level II to U.S. Navy personnel psychologists, they tried using a Level I test (auditory forward digit span) with Category IV recruits (i.e., those with AFQT [Level II] scores below the 30th percentile). They subsequently found that assessments of Level I ability were better at predicting success in job training for Category IV recruits than the AFQT scores. Level I provided virtually no incremental validity, however, for recruits with average or above-average AFOT scores. Their findings are consistent with the hypothesized hierarchical relationship between Level I and Level II.

The hypothesis that Blacks differ from Whites much less, if at all, in Level I ability than in Level II ability, on average, would seem to account for our major findings. The simplest, most clear-cut test of this hypothesis was performed by comparing large random samples of Black and White children in California schools on the forward digit span (FDS) and backward digit span (BDS) subtests of the Wechsler Intelligence Scale for Children-Revised (WISC-R). The hypothesis was strongly borne out: the standardized White-Black (W-B) difference on BDS was almost twice as great on FDS, and BDS correlated almost twice as much with Full Scale IQ as did FDS (170). The Level I-II hypthesis was also confirmed using other tests of learning and memory (98, 146, 157) and by the measurement of the degree of clustering (i.e., mental manipulation) of items in free recall of serially- presented common nouns (148).

The persistent question of just how and why slow and faster learners differed led me to include mildly mentally retarded adolescents and young adults in my learning experiments, and to compare them against young children who had comparable raw scores (i.e., mental age) but were of average or superior IQ (73). The children were faster learners than the retarded subjects, despite having the same raw scores as the retarded adults on ordinary IQ tests.

One noticeable difference between the groups was that, in paired-associates learning, for example, the average-IQ children actively and spontaneously used some form of verbal mediation to acquire the connections between the paired items (which consisted of colored pictures of familiar objects, e.g., cat-apple), such as imagining or verbalizing to themselves something like "The CAT eats the APPLE" (37, 38, 48). Even if the normal children did not initially think of using this "trick," as soon as it was suggested by the experimenter,

they used it on every paired-associate list thereafter, and it facilitated their performance greatly. The retarded persons, however, could benefit from verbal mediation only when a mediating association for each paired-associate was provided by the experimenter, in which case it also improved their speed of learning. But most of the retarded subjects would not spontaneously generate their own mediators on subsequent trials. In another study, performed with junior high school students, I found that in a nominally nonverbal trial-and-error S-R selective learning task (with informative feedback) children in classes for "slow learners" show a marked gain in learning speed when they were instructed simply to name the stimuli aloud as they performed the task. In contrast, children of average and above-average IQ showed virtually no performance gain after instructions to name the stimuli aloud, presumably because they were already naming them spontaneously, but subvocally, prior to the instruction (34; see also 52, 55, 56, 67, 78, 93).

In searching further for sources of individual differences in learning, I conducted a series of studies (45) aimed at determining if there were reliable individual differences and common factors in various types of interference. Interference is measured in the classic laboratory paradigms for proactive and retroactive interference in a variety of learning and short-term memory tasks, both visual and auditory (100), and by the interference effect measured by the Stroop color-word test (44, 58). There are reliable individual differences in all of these types of interference effects and some (least of all the Stroop effect) are related to SAT scores and college grade-point- average. Individual differences in learning were conceptualized largely in terms of the degree to which the various learning tasks interacted with individual differences in susceptibility to proactive and retroactive inhibition on every to-be-learned item on every trial up to the point of mastery of the task.

Today these data would be interpreted in terms of individual differences in the capacity of Working Memory, the hypothetical construct of modern cognitive psychology that seems most relevant to understanding the g factor, which during the period of my interest in human learning was almost invisibly far in the background of my thinking. But I was more an experimental psychologist than a psychometrician in those days, and at that time experimental psychologists typically looked on the g construct, and even the study of individual differences as being of interest in their own right, only with disdain.

The Behavior-Genetics of Intelligence

When, in 1966, I was invited to spend a year at the Center for Advanced Study in the Behavioral Sciences, I had enough research material on the learning characteristics of culturally disadvantaged children, as they were called at that time, that I thought I could best spend my year at the Center writing a book about my findings. I took all of my research material with me and began work at the Center, a wonderfully undisturbed and heavenly atmosphere for study and writing, with a most helpful staff and the intellectual companion-ship of the many other fellows at the Center.

A reasonably comprehensive book about the educationally disadvantaged children, I thought, should contain one short chapter addressing the issue of the inheritance of intelligence, if only to show that this line of explanation for individual and group differences in scholastic performance could be dismissed as outmoded and scientifically discredited. I had never given this topic much thought and knew shamefully little about it at that time. It had never been touched upon in my entire education to that point and the subject was generally either unmentioned or scorned by virtually everyone I knew working in the field of learning and the educationally disadvantaged. My first exposure to it was in 1957 during my postdoc in London, when I attended a lecture on "the inheritance of mental ability" by Sir Cyril Burt. Though I was highly impressed by Burt's lecture as a brilliant tour de force, its subject didn't really capture my interest at that time. Burt's lecture was later published in the *American Psychologist* (1958, *13*, 1–150), and it seemed a good place to start what became my program of reading virtually the entire world literature on the genetics of mental ability at that time. I hadn't expected to go that far into the subject, but the more I read, the more I realized it couldn't be dismissed and had to be taken seriously.

In order to be able to evaluate much of what I read, I had to tackle the technical aspects of quantitative genetics. Luckily, a professor of ethology and genetics was also a fellow at the Center that year and was a most helpful tutor and guide to the literature on quantitative genetics. I felt most resentful of the fact that I had reached that stage of my education and of my career and had not been exposed to the existing scientific knowledge on the genetics of mental ability. I was even more dismayed to realize that my case was all too typical of those working in most branches of psychology, particularly experimental, educational, and clinical. All human variation in abilities was attributed to the learning opportunities afforded by different environmental and cultural circumstances to which individuals were exposed. Though at that time the literature on behavior genetics was but a fraction of its present volume, what there was seemed sufficient to call in question the prevailing 100 percent environmentalism of the 1950s and '60s. My task was cut out for me: to help dispel the ignorance that generally prevailed in educational psychology concerning the role of genetic factors. In reading E.L. Thorndike, the father of American educational psychology, I found that he was on the right track in his intuition about the importance of genetic factors in individual differences, but his line of thought on this subject rapidly went out of fashion shortly after World War II, for no good scientific reason.

Therefore, during my year at the Center, I wrote several articles that stemmed from my new-found interest in the genetics of mental ability and its implications for education (61, 62, 63, 64, 68, 70). The most frequently cited of these articles is based on my examination of the famous Holzinger formula for estimating heritability from the difference between the correlations between MZ twins and between DZ twins. I showed that Holzinger's formula, which was virtually the only one ever used in studies of the heritability of intelligence up to that time, did not estimate heritability as it is defined in quantitative genetics, nor did it take account of the effect of assortative mating on the estimation of heritability from twin data(61; see also 178). I provided a new formula that not only accorded with the meaning of broad heritability as defined in genetics but also took account of assortative mating. (The formula could also be generalized for estimating heritability with other kinships besides twins, such as full siblings and half siblings.) I used this formula to recalculate heritability coefficients for IQ on every published study of MZ and DZ twins.

Although the articles I wrote that year emphasized the evidence for the substantial heritability of *individual differences* in IQ, I thought (and wrote) that it was unnecessary to invoke genetic causes for the observed racial differences in IQ, which I thought could be explained in terms of cultural bias in the tests and poor environmental opportunities for acquiring the particular knowledge and skills called for by conventional tests. One of my articles written at the Center (63), which originated as an invited address at a convention was titled "How Much Can We Boost IQ and Scholastic Achievement?" It came to the attention of the editors of the *Harvard Educational Review*, and in 1968 (when I was back at Berkeley) they asked me to expand it into a more comprehensive article for the Spring, 1969 issue of the *Review*. They even provided an outline of the topics they wanted me to deal with in the article, including my view on the heritability of race differences (which I had not previously mentioned). I gladly accepted the editor's invitation, as an opportunity to consolidate what I had been studying and thinking about during my year at the Center. The result was a 200-page typescript which became a 125-page article in the *Harvard Educational Review* (*HER*) titled "How Much Can We Boost IQ and Scholastic Achievement?" (76). Though unexpected at the time, I suppose it was the article that forever changed my life, for better or worse.

My Year of Turmoil

Based on a review of the empirical literature, my HER article made four main claims: (*i*) experimental attempts to raise the IQ of children at risk for low IQ and poor scholastic performance by various psychological and educational manipulations had yielded little, if any, lasting gains in IQ or scholastic achievement; (*ii*) individual differences in IQ have a high heritability (.70–.80, corrected for attenuation), but environment also plays an important part; (*iii*) most of the exclusively cultural- environment explanations for racial differences in IQ and scholastic achievement were inconsistent and inadequate, so genetic as well as environmental factors should be considered; and (*iv*) certain abilities, particularly rote-learning and memory (i.e., Level I ability) have only a weak relationship to IQ, which suggests that these Level I abilities might be used to compensate to some extent for low IQ (i.e., Level II ability) and thereby make school instruction more beneficial for many children, regardless of their racial or social class background, who are below average in Level II but are average or above in Level I. (Pupils with this pattern of abilities constitute the majority of those who are most at risk for failure under traditional classroom instruction.)

Viewed as a whole, it seemed quite reasonable. But it was the few pages on race differences in IQ and achievement (about 5% of the article) that aroused so much sound and fury, most of it focused on the one sentence that violated what I later came to realize is the greatest taboo in the latter half of the twentieth century. Here is what I wrote concerning the Black-White difference in IQ: "The preponderance of the evidence is, in my opinion, less consistent with a strictly environmental hypothesis than with a genetic hypothesis, which, of course, does not exclude the influence of environment or its interaction with genetic factors" (76, p. 82).

That one aspect of the article was blown up by the mass media, with feature articles in *TIME*, *Newsweek*, *LIFE*, *U.S. News & World Report*, the *New York Times Magazine*, and many other newspapers and magazines, as well as radio and TV programs. The Berkeley campus was in an uproar for weeks (and sporadically for months and even years thereafter) with bands of demonstrators disrupting my classes, slashing all the tires on my car, amd painting swastikas on my office door. The student paper, *The Daily Cal*, carried many denunciations and only a few defenses of my position, and there were demands from dissident groups that I be fired. The campus police assigned plainclothes bodyguards to accompany me whenever I left my office, and for several months the campus bomb squad handled the screening and opening all of my mail, even some of the unidentified mail received at my home. There were telephoned and mailed threats on my life and on my fam-

ily; phone calls were routed (and recorded) through the local police station. A number of the calls that came in over one period of several days so worried the police that they urged me and my family to spend a week away from our home at some unknown location, as the police could not provide 24-hour protection. (We stayed with friends in a neighboring suburb; an inconvenience, but as they had a lovely swimming pool, it was a pleasant diversion.) Worst of all, from my standpoint, was that my on-going research in the Berkeley schools was immediately terminated and permanently proscribed by the Berkeley school officials (128). When I asked one official for an explanation, he remarked, "The Berkeley schools are a political unit, not a research institute."

Many other harrowing incidents followed, some taking place when I was lecturing on other college campuses, both in the United States and abroad, even when my lectures didn't touch on the subjects of genetics or race. The largest, most tumultuous demonstration I ever experienced was, surprisingly, at the University of Melbourne, in 1977, where about fifty policemen had to rescue me from a madding mob. The unprovocative topic of my [undelivered] lecture: The relationship between intelligence and learning [189; see also 301]. The very next day the same thing happened in the same setting to Hans Eysenck. His topic: the relationship between personality and learning. (I have written at greater length in the Preface to my *Genetics and Education* about the bizarre events following the publication of my HER article [112; see also 149].)

The really important consequence of the HER article for my subsequent activity was that it raised a number of questions and issues concerning subjects that called for fuller explication or further research. In many articles (from #77 on) and three books (143, 144, 206), I consolidated my position on these subjects as best as empirical research permitted at that time and launched new research on the remaining unanswered questions and speculative hypotheses. Some people advised me to get out of this controversial area altogether. One eminent psychologist friend warned that if I scorned the Zeitgeist, it would in turn scorn me. However, rather than duck for cover, which I peculiarly felt would be disgracefully un-Gandhian, Iresolved not to be whipsawed by the prevailing orthodoxy in the social and behavioral sciences, but to do whatever I could to reform the social sciences. And I believe that at least the scientific community, if not the media and the political establishment, has indeed changed its mind if not its voice over the past 30 years, with an almost total collapse of naive environmentalism and an increasing recognition, at least in the pages of academic journals, of the importance of genetic factors and of environmental factors with biological effects on the development of human mental ability. The well-known survey by Synderman and Rothman (1988) of over 600 psychologists in the relevant fields showed that their modal response on every question that involved the very issues considered heretical my HER article agreed with the position I then stated. To what extent my own work may have helped usher in the new perspective would be impossible to estimate, but I believe I have played a role. Many other influences, of course, have brought about the demise of doctrinaire environmentalism and advanced the biological orientation of mainstream behavioral science.

Bias in Mental Testing

At about the time of my HER article, the question of culture bias in mental tests was frequently brought up. I was familiar with the early resarch on social class bias in standard

tests, pioneered by Kenneth Eells (who had been one of my professors), but I found rather little more than speculation in the literature regarding test bias with respect to racial or ethnic groups. Yet that question was crucial. I reviewed what little evidence existed on the subject in the mid-1960s (69, 99), but found it inadequate and largely unconvincing. Besides the educational, social, and economic unfairness of using tests that are differentially and systematically inaccurate for different racial, ethnic, and social-class groups in our population. I considered also the scientific importance of test bias for the field of psychology in its own right. Psychometrics—the science and technology of mental measurement—is of course basic to many fields in psychology, as indeed reliable and valid measurement is essential for the development of any science or technology. To the degree that the standard psychometric instruments then in use were biased, either by culture, social status, or gender, basic research in differential and educational psychology as well as the practical applications of testing in educational placement, in college admission, in personnel selection, and in assigning recruits to various training schools in the armed services were all compromised by having to operate with deficient tools.

I thought it imperative to devise methods for detecting the presence of various kinds of psychometric bias. This became the main focus of my research effort for the next few years (109, 153, 176, 179, 181, 182, 289). It culminated, in 1980, with the publication of my Bias in Mental Testing (199; see also 202, 203, 217), an 800-page tome which was then (and may still be) the most comprehensive work on the subject. Research on criteria of bias based on a test's so-called *external* validity, that is, its practical predictive validity (i.e., both the regression [and correlation] of criterion measures on test scores) in different subpopulations had already been quite well investigated by psychometricians during the period between 1970 and 1980. Though I fully explicated this work in Bias, my own research contributions emphasized internal indicators of bias, such as whether different groups, (e.g., Black-White, male-female) differ significantly in various psychometric features such as the test's reliability, the test items' rank order of difficulty, the test scores' correlation (and regression) with chronological age, the relative frequency of choosing various distractors (i.e., error responses) in multiple-choice tests, the groups' similarity in the factor structure, and the groups' similarity in kinship correlations and heritability values for the test in question.

A methodological innovation I introduced was the use of what I termed "pseudo-race age groups." For example, when I found significant differences between Black and White school children in their specific choices of error distractors (in the Raven Progressive Matrices test), I created two "pseudo-race" groups composed entirely of White children, the groups differing in age such that the younger group and the older group had the very same mean difference in total test score as the mean difference between age-matched Blacks and Whites. I discovered that the same-age Black-White differences in the frequencies of selecting a particular distractor (i.e., a wrong answer) among the several distractors for each item on the Raven test were virtually identical to the differences between the two groups of White children that differed in age by almost two years (approximately ages 8 years and 10 years). Applying this method to a variety of tests, including Gesell's Figure Copying test (a good predictor of scholastic performance in the primary grades), free drawing, and several Piagetian tests, we found that in every feature of test performance, agematched groups of Black and White children differed in exactly the same way as did "pseudo-race" groups of different-age White children when the age of the younger group is about 80% that of the older group.

Our findings with pseudo-race groups suggested that the observed racial differences in performance were not attributable to test bias, but reflected a developmental difference in rate of mental maturation, with Whites (and more so, Asians), on average, having a steeper trajectory and a higher asymptote. From my own studies and my review of the total empirical literature on test bias, I concluded in *Bias*: "...the currently most widely used standardized tests of mental ability—IQ, scholastic aptitude, and achievement tests—are, by and large, *not* biased against any of the the native-born English-speaking minority groups on which the amount of research evidence is sufficient for an objective determination of bias, if the tests were in fact biased" (p. ix). Essentially the same conclusion was announced independently two years later in a joint report by the National Research Council and the National Academy of Sciences (Wigdor & Garner, 1982), which had chosen a panel of nineteen leading experts in psychometrics to review the evidence.

The Cumulative Deficit Hypothesis

One hypothesis proposed in the 1960s to explain the Black-White difference in the trajectory of raw scores on mental tests and on scholastic achievement across grades 1 to 12, held that the increasing racial disparity in test performance with increasing age is the result of a cumulative deficit in learning, such that failure to learn particular knowledge or skills thoroughly at one grade level hinders the ability to learn more advanced material in later grades. Because Black children, on average, begin school having learned less of the prerequisites for learning in the primary grades, they fall further and further below national norms in scholastic achievement with each additional year. Progressive learning deficit is thought to act cumulatively, like mounting credit card debt. This hypothesis was popular in the 1960s and provided much of the rationale for Head Start and other compensatory education programs (54; see also 158, 162, 304, 314, 380).

My investigation of the phenomenon, however, found the evidence for it ambiguous, at best. The divergence between Black and White test scores with increasing age or grade level in school was fully apparent when looking at *raw* scores, but there was little, if any, evidence for a divergence of Black and White mean scores when the scores are expressed as *standardized* scores. That is, when measured in age-standardized scores, the mean Black-White difference of about one standard deviation remains constant from kindergarten to 12th grade, because the standard deviation *within* each group also increases proportionally with age.

I reasoned that if there were a true cumulative deficit effect for IQ, and if it was the cause of Blacks' lower average IQ, and if IQ declined the longer a child stayed in a culturally disadvantaged environment, then Black children at any given age should have, on average, a lower IQ than that of their younger siblings. The cumulative deficit theory predicts that the positive difference between the ages of the older (O) and the younger (Y) sibling is positively correlated with the Y-O difference in IQ. There should be a substantial such correlation among Blacks (i.e., the older sib should have a lower IQ than the younger sib), but this effect should be negligible or nonexistent for middle-class and upper-class Whites. A significant correlation for Blacks would support the favored environmental explanation of the cumulative deficit, because there is nothing in genetic theory which would predict a systematic difference. on average, between the genotypes of full siblings for any given characteristic, as each sib receives a *random* set of half of each of its parent's genes.

I tested this prediction with several hundred pairs of Black siblings and White siblings, all of school age (156). Despite the large samples, the hypothesis was not born out, although the theory-predicted correlation was significant (p < .05, one-tail test) only for verbal IQ in the all-male sib pairs, and then only for those in the primary grades. No other subdivision of the data revealed the slightest indication of the predicted correlation, for either Blacks or Whites.

This null outcome made me wonder if the Black population in Berkeley, California, despite its typically lower IQ compared to Whites and Asians in the same schools, wase somehow atypical of the general Black population, perhaps being less environmentally disadvantaged. The cumulative deficit might occur only in children whose environmental disadvantage falls below a critical a threshold necessary for normal phenotypic development of the individual's genotypic potential. I realized, therefore, that another study using exactly the same methods would have to be done in an area where there could be no question that the vast majority of the Black school children lived in a conspicuously impoverished environment.

I found the necessary data for this study in a school district in one of the poorest counties in the rural South. The IQ of the Black pupils was 71, averaged over kindergarten to 12th grade; the average IQ of the White pupils was 101. All of the full siblings, White and Black, enrolled in all of the schools of this rural county were included in the study. The findings were startling. The White school population showed no evidence of an age-related decline in IQ, in this respect being like the White sample in my Berkeley study. The Blacks, however, showed a marked age-decrement in IQ, as indicated by the younger minus older sibling IQ difference – a decrement of about one IQ point for each year of the Y-O sibling age difference. In other words, with family background controlled (by the sibling design) these Black children declined, on average, about one IQ point per year throughout their time in school (180). One might have argued that this was not necessarily an environmental effect but a racially genetic difference in the trajectory of the mental growth curves for Blacks and Whites. The California data, however, seemed to rule out this interpretation, as they evinced no such effect for Blacks. If the effect observed in Southern rural Blacks were a genetic racial characteristic rather than an environmental effect, it should have shown up, at least to some degree, in the California Blacks as well. It therefore seems most likely that some substantial part of the IQ deficit for Blacks in the poorest environments is a result of environment, most probably environmental factors that have biological consequences, such as unfavorable prenatal conditions, poor nutrition, and childhood illnesses, which can limit mental development.

Spearman's Hypothesis

While re-reading Spearman's major work, *The Abilities of Man* (1927), I came across a brief passage (p. 379) that had not previously captured my attention. On second reading, however, it made a major impact. On the basis of one slight study (by American psychologists) of Black-White differences on a variety of cognitive tests, Spearman conjectured that variation in the magnitude of the B-W difference across various tests is directly related to

the size of each test's g loading. (Because the article presented only the tests' means but not their intercorrelations, Spearman could not extract the g factor with which to test his conjecture.) I henceforth referred to this conjecture as "Spearman's hypothesis." It struck me as of quintessential importance, because, if true, it is a much more general hypothesis than my Level I-Level II formulation, which appears to be just a special case of Spearman's hypothesis. Spearman's hypothesis also seemed to explain why the size of the B-W difference (in standardized units) varied so widely across different cognitive tests. This question had always been a stumbling block to the prevailing environmental theories, which were a plethora of piecemeal, ad hoc, inconsistent, and unconvincing explanations. Spearman's hypothesis, if true, would mean that the B-W difference was essentially a difference in g. Therefore, if we are to understand the phenotypic B-W difference in measurements of cognitive ability, it would be necessary to understand the nature of g itself. First, Spearman's hypothesis had to be put to a rigorous empirical test. This called for representative samples of Blacks and Whites measured on as wide a variety of mental tests as could be found. I tested Spearman's hypothesis on a large scale (224, 256, 266, 267,

268, 288, 290, 296, 324, 325, 339, 375). By publishing my analysis of much of the evidence as a target article in *The Behavioral and Brain Sciences* (266), its was subjected to commentaries by over thirty experts in psychometrics and cognitive psychology. In brief, the total evidence strongly bears out Spearman's hypothesis. It is no longer a hypothesis, but must now be regarded as an empirical fact, as much so as Galton's Law of Filial Regression or Thorndike's Law of Effect. (The most recent comprehensive summary of the methodology and evidence on Spearman's hypothesis is provided in Chapter 11 of my recent book, The g Factor [383].)

Speed of Information Processing and g

From the standpoint of developing a practically useful test of mental ability, Binet chose right path. From the standpoint of a scientific analysis of mental ability, it was, however, Galton who showed the way. The complexity of information processing required by conventional g-loaded psychometric tests, such as IQ, is much too great to afford a finegrained analysis of the processes that underlie g. The Galtonian approach requires the study of elementary cognitive tasks (ECTs). A given ECT is so simple that all subjects can perform it easily and correctly, and the only reliable source of individual differences is response latency or reaction time (RT).

I devised a computerized chronometric apparatus (194; see also 218, 219, 264, 286) based on the Hick paradigm, on which the subject depresses a "home" button, hears a preparatory stimulus ("beep"), then, after 1 to 3 sec., a light/button (6 inches from the "home" button) goes "on" which the subject quickly turns off by pressing the light/button. The target light/button can be presented in an array consisting of either 1, 2, 4, or 8 light/buttons (corresponding to 0, 1, 2, or 3 bits of information). Called the RT-MT apparatus (some have called it "the Jensen button-box"), it presents trials automatically, records the subject's performance on each of *n* trials, and yields five measures (based on *n* trials): (*i*) *median RT* in milliseconds (interval between onset of the reaction stimulus and releasing the home button; (*ii*) *median MT* (movement time, the interval between onset of the reaction stimulus and touching the target light/button); (*iii* and *iv*) *intraindividual variability* of RT and of MT (measured as the standard deviation of the subject's RTs and MTs over n trials, labeled RTSD and MTSD); and (v) the number of errors in n trials.

My first use of this apparatus was to see if its measures of RT for 0, 1, 2, and 3 bits of information conformed to Hick's law, which predicts that RT increases as a linear function of the amount of information measured in bits (i.e., the base 2 logarithm of the number of light/buttons in the array on the subject's response console). The RT data indeed conformed beautifully to Hick's law; it looked more like data from a physics experiment than from a psychological experiment.

In a series of studies using the apparatus with groups selected from different segments of the bell-curved IQ distribution, I replicated the results of an earlier study by Roth (1964) based on a similar procedure, except that Roth's apparatus did not use a "home" button and therefore measured an amalgam of RT and MT in unknown proportions. Roth had reported a significant negative correlation between IQ and the *slope* of the Hick function. The intercept of the Hick function had a consistently higher correlation with IQ than did the slope, which has much lower reliability than the intercept. (Much of this work is summarized in 201, 209, 219, 220, 282, 286.)

I also discovered what had not been anticipated by any of the previous RT literature, namely, that intraindividual variability in RT has the highest (negative) correlation with IQ of any of the RT parameters (327).

My procedure of measuring RT and MT separately showed that these two variables are very different in their nature. RT is more strongly and consistently related to IQ than is MT; in factor analyses of chronometric and psychometric tests, RT and RTSD load significantly on the *g* factor, but MT and MTSD do not. In fact, when a considerable number of speed-of-information paradigms are factor analyzed, RT and MT clearly load on different orthogonal factors.

Another theoretically important finding is that the RT measures correlate only with the g factor of a psychometric battery (such as the Wechsler Adult Intelligence Scale or the Armed Services Vocational Aptitude Battery); when g is partialled out of the various subtests, their multiple correlation with RT is not significantly greater than zero. The multiple correlation between a number of different RT measures derived from various ECTs and g seems to asymptote at about .70, because a substantial part of the total reliable variance of all RT tasks comprises a group factor that is specific to RT and independent of g. (This specific RT factor probably reflects largely the purely sensory and motor aspects of RT.)

In the 1970s, a few cognitive scientists, such as Alfred Baumeister, Earl Hunt, and R.J. Sternberg, began systematically investigating the relationship of individual differences in mental speed to traditional measures of cognitive ability. Despite the interesting and promising results of their studies, many psychologists (including some of my colleagues) insisted that this line of research had been tried and had failed. When I began using chronometric techniques to study the nature of g, my efforts were disparaged as being impossible, trivial, laughable, or at best eccentric. A typical reaction: "RT is obviously much too simple to reflect individual differences in anything so marvelously complex as human intelligence." Hadn't I studied the history of psychology? Hadn't I read the basic textbooks? For over half a century nearly every introductory psychology text discredited the idea that RT is related to "intelligence," while recounting the unsuccessful attempts of Galton and James McKeen Cattell (and his student Wissler) to measure "intelligence" with chronometers and other "brass instrument" devices. Now, two decades after the revival of

mental chronometry in the 1970s, there is massive evidence for a substantial relationship between RT and g. Indeed, one of the joys of research is seeing a long-shot intuition finally pay off empirically in the face of unimaginative, tradition-bound nay saying. Eminent scientists, Galton among them, have played long-shot, but eventually fruitful intuitions; while more conservative researchers have consistently played it safe, spending their professional lives confirming the prevailing conventional wisdom.

Galton's original intuition has now been fully confirmed. The fact that a number of RT variables, singly or in an optimally weighted combination, do not correlate as highly with psychometric g as the reliability of measurement would allow does not diminish the theoretical importance of the established fact that "mental speed" is a substantial component of g. A causal basis of the observed between "mental speed" and g has been hypothesized (338); though consistent with existing data to date, it has been neither proved nor refuted.

"Mental speed" is in quotes because we still don't know just what "mental speed" consists of. It can be defined operationally in terms of the measurements derived from various chronometric techniques, or as the response latency in various ECTs. But what is the *mechanism* by which "mental speed" is correlated with g? Is it a cause of g, or just an epiphenomenon? One of the challenges to contemporary neuroscience is discovering the nature of "mental speed." Are individual differences in our chronometric measures attributable to individual differences in brain nerve conduction velocity (335), in axonal conduction, in the ionic concentration at the axonal membrane in synaptic latency, in the design features of the neural circuitry, in the number of cortical neurons, in amount of neural redundancy, in chemical neurotransmitters, or in some combination of these and other variables (340, 348, 374)? As yet, no one knows. This is clearly territory that cries out for exploration. We have reached the point where certain psychological phenomena, such as intelligence, cannot be explained in purely psychological terms. A meaningful explanation awaits understanding in terms of the underlying neurophysiology.

Non-psychometric Correlates of g

Critics of the g construct have argued that g is merely an arbitrary artifact of the way psychometric tests are constructed and inherent in the mathematical procedure of factor analysis. If this were truly the case, I reasoned, the g factor should not be related to variables other than psychometric tests and should tend to disappear when using different factor analytic methods, assuming, of course, that such methods (like varimax rotation) are not specifically devised to scatter the g variance among a number of uncorrelated group factors.

Method Invariance of g. First, I treid to determine whether the g factor of a correlation matrix of diverse psychometric tests is more or less invariant when the g factor is extracted by any of the several quite different methods that have been used by various researchers throughout the history of factor analysis, from Spearman to the present day. By applying each of the main methods of factor analysis to real data and also to artificial data for which the g loadings of the "tests" were known exactly, it was found that g is remarkably similar across all of the different methods, as shown by congruence coefficients averaging over +.99 (360).

Meta-analyses of Physical Correlates. In 1930, long before the invention of metaanalysis, Donald G. Paterson published his classic work, Physique and Intellect, which reviewed all of the then existing studies on the correlation between physical features and measures of intelligence. Most of the correlational studies were based on rather small samples, and as the correlations between physical measures and IQ are typically small, they were usually nonsignificant statistically. Paterson simply compared the number of significant and nonsignificant correlations and usually concluded that the null hypothesis could not be rejected. His conclusions of essentially "no physical correlates of IQ" have become a dogma in psychology textbooks, and the vast majority of psychologists even today will tell you, for example, that there is no correlation between head size or brain size and IQ. As I generally doubted many of Paterson's conclusions, I decided to review all of the studies of physical correlates done since 1930 and, when possible, to combine the results of various studies by the methods of meta-analysis. The result was that the null hypothesis (i.e., no correlation) could be rejected at high levels of confidence for most of the physical characteristics that had been examined in relation to IQ; these include body size, head size, brain size, blood types, ocular characteristics, and other anatomical and physiological variables (341). The significant correlations between psychometric scores and such a wide variety of physical traits argues forcefully that the population variance on standard mental tests, such as IQ, reflects latent traits that are profoundly enmeshed with organismic variables in complex ways.

An innovative feature of my meta-analytic review was that, where possible, it examined three different types of correlation of a given physical trait with IQ: (*i*) the correlation within individuals (W1), (*ii*) within families (WF), and (*iii*) between families (BF). This methodology, based on sibling data (202), is an analytically important tool that helps determine the probable cause of the observed correlation. For example, failure to find a WF correlation, even when there is a significant WI correlation in the general population, rules out pleiotropy (i.e., two or more distinct phenotypic characteristics being the result of the same gene). Height and IQ show a WI (and BF) correlation, but they do not show a WF correlation (200). Head size and IQ show a WF correlation (358), as do myopia and IQ (299), suggesting that these two physical traits are pleiotropically related to IQ. These findings aid the search for the specific gene loci responsible for variance in IQ or g and may also provide clues to the precise physical basis of IQ variance.

The Method of Correlated Vectors. IQ is highly g-loaded, but it is typically saturated with other factors as well. To determine whether a given nonpsychometric variable is related to g per se, rather than to any other factors or specific sources of variance in test scores, I invented the method of correlated vectors. Essentially, it consists of factor analyzing a large psychometric battery of highly diverse subtests to obtain the g loadings of each subtest. This column vector of the subtests' g loadings is termed V_g . Each subtest is then correlated with some non-psychometric variable, X. The column vector of these correlations is V_X . Controlling for differences in the subtest reliabilities (by disattenuating or partialing out the subtests' reliability coefficients), a significant correlation between the parallel column vectors V_g and V_X shows that g and X are related. It tells us that the larger a subtest's true g loading, the larger is its correlation with variable X.

I have examined a number of variables (X) by this method. The correlation between V_{ρ} and the various measures (i.e., V_X) is shown in brackets: in brackets): spouse correla-

tion [.90]: heritability [.60-.80]; inbreeding depression [.80]; cerebral glucose metabolic rate [-.79]; brain intracellular pH [.63]; head size [.60-.70]; choice and discrimination RT [.70-.80]; average evoked potential habituation amplitude [.80]; and AEP waveform complexity complexity [.95]. All of these correlations are significant; the particulars on the studies of each variable are given in (226, 258, 282, 356, 383 [Chapters 6-9]). No other factor shows anywhere near the same degree of relationship to non-psychometric variables as does g. My research shows conclusively that psychometric g, far more than any other factor, reflects individual differences in certain biological and developmental properties of the brain that govern its speed, consistency, and capacity for information processing. Though manifested overtly in many ways that can be described in behavioral terms, g itself cannot be described or explained in behavioral or psychological terms. The g factor per se does not reflect any *particular* achievements, knowledge, or skills, but rather the information processing capacity for acquiring and using the knowledge and skills necessary for achievement.

I have pointed out a crucial conceptual distinction, namely, that the *construct* (in this case g) and the *vehicle* used for measuring the construct (in this case, a psychometric test and the scores it yields) are not one and the same: they are conceptually and empirically distinct. Though the *rank order* of individuals' scores on any highly g-loaded test can be accounted for largely in terms of individual differences in the level of g, the *absolute level* of the individuals' raw scores on any such test also reflects the particular composition of the test items (332). This fact has important consequences for the interpretation of test scores and the secular trend in the population mean for any particular vehicle of g (319, 368, 383, Chapter 10).

In addition to showing that g is correlated with various biological variables, I have also amassed empirical evidence (based largely on the method of correlated vectors) to show that it is g itself that accounts for most of the practical predictive validity of tests used in educational placement and selection and in personnel selection in industry and the Armed Forces. When the validity coefficient is based on a multiple correlation, typically the increment in predictive power contributed by all other factors (and by test specificity) independent of g is remarkably small (383, Chapter 9).

Future Directions

I see basic research on human mental ability, particularly g and the major group factors, as advancing in each of two directions, which I think of as the *horizontal* and the *vertical*. Both are necessary and each can be scientifically rigorous.

Horizontal research on g explores the whole nexus of behavioral, social, and economic correlates and consequences of individual and group differences in the level of g. I believe g plays a greater role in these spheres than most educators, sociologists, criminologists, economists and social policy analysts presently realize. But serious consideration of this probability seems to be strongly resisted in some circles. The generally nihilistic reaction of the mass media to *The Bell Curve* by Herrnstein and Murray (1994), which, more than any previous scholarly work, examined the relationship between g and a number of social variables of national concern, is a case in point. It is a reasonable supposition that in the global economy and the competitive technological and information-intensive world of the twenty-first century, a nation's chief natural resource will be its population's overall level

of educability, in which the distribution of g-loaded abilities will inevitably be a major determining factor. Researching the extent and the ramifications of the g factor in the nexus of societal variables is the province of a budding new field named the "sociology of intelligence" by sociologists Robert Gordon and Linda Gottfredson. (See the a special issue of *INTELLIGENCE*, 1997, 24, No. 1, for example.)

Vertical research digs down in search of the causal basis of g. Being analytical and reductionistic by nature, I personally find this is the more interesting aspect and the one to which I expect to devote my efforts in the future. By definition, an underdeveloped field is one in which many of the findings, and even some of the classic experiments, have not been replicated and many of the supposed facts not fully consolidated. I am happy therefore to see research on "mental speed" in relation to psychometric g being actively pursued in a number of laboratories around the world. The new information that I see coming in, almost every month, is most valuable. However, there is still confusion, contradiction, and many unanswered questions. More standardized apparatus and procedures are called for (as the same standardized reagents are used in every chemical laboratory) and much more importance must be accorded to replicating the theoretically crucial findings across different laboratories. The measurement of individual differences in g by means of mental chronometry is as close to the interface between brain and behavior as we are likely to come. The consolidation of the knowledge gained at this interface is important for vertical advancement, that is, identifying the basis of g in the structural and functional features of the brain itself.

A few hypothesis-generating steps have already been taken by showing g-correlates of direct brain measurements obtained with evoked potentials, magnetic resonance imagining (MR1), positron emission tomography (PET scan), and intracellular brain pH levels. Many researchers in the neurosciences, however, have ignored g as a subject for brain research, mostly, I fear, because a crucial distinction has not been made sufficiently clear.

In recent articles (374, 384), I have emphasized the distinction between (i) the neural circuitry or design features of the brain possessed by all neurologically intact members of a species that are responsible for that species' characteristic behavioral capacities, and (ii) the properties of the brain of a given species that cause intraspecies variation (i.e., individual differences) in that species' characteristic behavioral capacities. We know from research in behavior genetics that intraspecies variation in many behavioral capacities is not entirely the result of experiential differences and learning. But I have found no good reason to believe that the design features of the brain (which are undoubtedly crucial determinants of *interspecies* variation in behavioral capacities) are necessarily involved in intraspecies variation. The latter may well be due to an entirely different set of causes than neural circuitry or other designs features of the brain, but rather involve such within-species factors as: differences in blood supply (via the richness of the capillary network), the degree of myelination of axons (which affects nerve conduction velocity), the neuroglial cells (which nourish neurons), the brain chemistry of neurotransmitter (which affect synaptic transmission), and individual differences in the number of neurons involved in the various brain modules. We now know quite conclusively from MRI studies, for example, that IQ is correlated with brain size, but we still don't know what precisely it is about brain size that causes this correlation.

All but an exceedingly few neuroscientists today are interested in intraspecies variation in behavioral capacities. They may well find discovering the brain's general operating

principles daunting enough, without having to explore the causes of individual variation in the functional efficiency of the essential design features of the brain and their general operating principles. As Francis Crick (1994) has pointed out, neuroscience cannot yet explain even how the human brain sees things in the environment, much less how it performs the complex functions we call intelligence. The very existence of the g factor (like all other psychometric factors) is only revealed by examining *intra*species individual *differences* in each of the broad and diverse class of functions we regard as constituting intelligence discrimination, generalization, learning, memory, insight, abstraction, problem solving, and the like. But before we can begin to research the physical basis of g, do we first need to discover all of the brain's design features that make these functions possible? I don't think so. The question of what causes the various cognitive functions of the brain to be positively correlated is a very different question from that of understanding the specific operating mechanisms of each of these functions.

As a heuristic proposition to encourage research in this "vertical" search for the neurological causes of g, 1 propose the following working hypothesis: Individual differences in human behavioral capacities do not result from differences in the brain's structural operating mechanisms per se, but rather are the result of other aspects of cerebral physiology that modify the sensitivity, efficiency, and effectiveness of the basic information processes that mediate the individual's responses to certain aspects of the environment.

I'm placing my bets on the search for those aspects of brain physiology responsible for g as most likely to generate the next path-breaking discoveries in differential psychology and human biology. I have been told by experts that the technology to do this already exists. So, looking ahead, I see my principal endeavor to be sparking the interest of qualified scientists in the brain sciences and helping them solicit the necessary resources to pursue this "vertical" investigation of g.

NOTE

1. Numbers in parentheses refer to the items listed in Jensen's bibliography (Appendix).

REFERENCES

Crick, F. (1994). The astonishing hypothesis. London: Simon & Schuster.

- Herrnstein, R.J., & Murray, C. (1994). The bell curve: Intelligence and class structure in American life. New York: Free Press.
- Hull, C.L. (1942). Principles of behavior. New York: Appleton-Century-Crofts.
- Roth, E. (1964). Die Geschwindigkeit der Verarbeitung von Information und ihr Zusammenhang mit Intelligenz. Zeitschrift für Experimentelle und Angewandte Psychologie, 11, 616–622.
- Skinner, B.F. (1938). The begavior of organisms. New York: Appleton-Century-Crofts.

Snyderman, M., & Rothman, S. (1988). The IQ contraversy. New Brunswick, NJ: Transaction Books.

Spearman, C.E. (1970/1927). The abilities of man: Their nature and measurement. New York: AMS Press.

Watson, J.B. (1929/1919). Psychology from the standpoint of a behaviorist. Philadelphia, PA: Lippincott.

Wigdor, A.K., & Garner, W.R. (Eds.) (1982). Ability testing: Uses, consequences, and contraversies/Part 1: Report of the Committee/Part 2: Documentation section. Washington, D.C.: National Academy Press.