CHAPTER 9

g *Theory*
How Recurring Variation in Human Intelligence and the Complexity of Everyday Tasks Create Social Structure and the Democratic Dilemma

Linda S. Gottfredson

Prologue to a Theory of *g*

As a new PhD in 1977, I had no particular interest in intelligence. I was, however, skeptical of my fellow sociologists' conception of it. Their status-attainment path models assumed that offspring IQ ("cognitive ability") is a product of parents' socioeconomic advantage. These models reflected the discipline's general consensus that intelligence, real or perceived, predicts school and work success only because it transmits the parent generation's social privileges to its offspring.

Where psychologists saw individual differences, sociologists saw social inequality. Where psychologists suspected genetic influences on cognitive competence, influential figures in sociology alleged an elite perpetuating itself under the guise of intellectual merit. Career-development psychologists asked how young people choose among different occupations; status-attainment researchers asked what bars the less privileged from entering the most desirable ones. Both theories of occupational attainment pointed to factors the other ignored. One classified occupations horizontally, by field of work; the other ordered them vertically, by prestige. One looked at the nature of work performed and interests rewarded in different occupations; the other only at the socioeconomic benefits flowing to workers in them. Both approaches had venerable histories and vast bodies of evidence, yet contradicted the other's most fundamental assumptions and conclusions.

I set out to reconcile the two disciplines' positions by testing the validity of their guiding assumptions. I began with the most basic premise in sociological explanations of social inequality: higher intelligence matters only for getting a good job, not for performing it well. Testing it required marshaling evidence on job duties, requirements, working conditions, and rewards for a large number of occupations from diverse sources, including vocational guidance research, employee selection psychology, and sociological studies of status attainment, plus civilian and military databases on jobs and employment testing, plus proprietary job analyses. My question: does the inherent nature of some work tasks and jobs require workers to do more difficult mental processing to carry out the tasks? Evidence that cognitive ability predicts job performance (then, mostly supervisor ratings) could not answer that question.

This initial research did double duty because my immediate concern was to fill a gap in vocational guidance: what types and levels of ability do different occupations inherently require of workers to do the job well? Ability profiles influence career choice, but I wanted to provide counselors and counselees a two-dimensional map of occupations, by the type and level of abilities they require, so they could assess and improve a counselee's odds of entering their preferred occupation and performing it well (Gottfredson, 1986, 2005b).

I mention this history because it would shape my intelligence research in distinctive ways. First, I studied populations of jobs, not persons. For instance, I factor analyzed the attributes of occupations, not workers. Second, I tested claims from one discipline with data from others. Third, I worked to solve nagging puzzles, unravel seeming paradoxes, and discern the deep patterns in superficially chaotic data. These strategies would lead me on a long journey through various disciplines to find evidence on the nature, origins, and societal consequences of human variation in intelligence.

Figure 9.1 schematizes my synthesis of replicated bodies of evidence: a cross-disciplinary theory, or explanation, of *g* in all its manifestations (Gottfredson, 2016). Seven levels of analysis are essential for pinning down what general intelligence is and does. They proceed clockwise from the most molecular processes (Genes) to most macro (Evolution). To be clear, by *g* I mean population variation in general intelligence and its ubiquitous, highly patterned effects in human affairs. The figure also highlights my conviction that we cannot understand this human trait or how it evolved until we know how the mundane tasks in daily life activate latent differences in *g*, make them visible, and magnify the practical advantages of having higher *g*. I will use Figure 9.1 to help answer the six questions to follow.

What Is Intelligence?

"Intelligence" is no longer a scientifically useful concept, in part because scholars have applied the word to ever more numerous and varied forms of

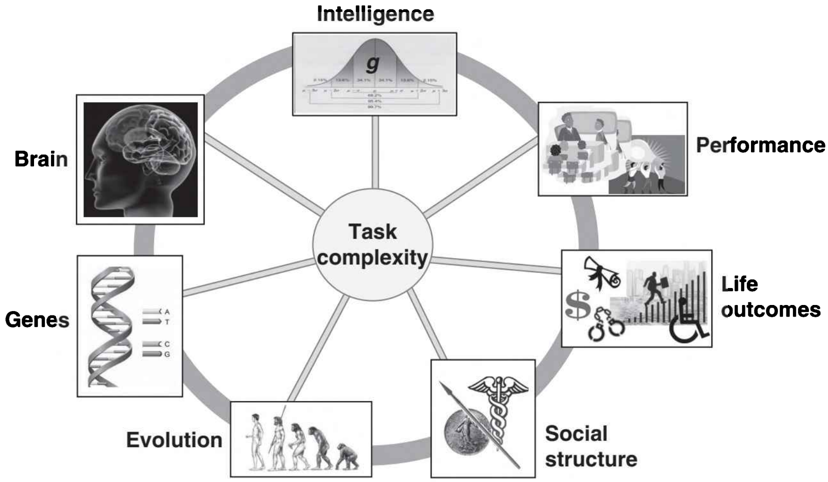


Figure 9.1 Network of evidence on *g* spanning different levels of analysis (Gottfredson, 2016).

aptness. No longer informative either are the long-standing debates over how best "to define" intelligence, as if a natural phenomenon could be formed or banished by expert consensus. The phenomenon in question is now best identified as *g*. It is a *trait*, a recurring dimension of human variation, and is best described by evidence converging from different levels of analysis.

The Problem with Intelligence

Carroll's (1993) empirically derived hierarchical model of human cognitive abilities helps illustrate why intelligence is no longer a useful theoretical construct. The three strata in his model order abilities by their observed generality of application, from highly general (Stratum III) to content specific (Stratum I). Carroll's laborious reanalysis of hundreds of factor-analytic studies confirmed there is only one domain-general (Stratum III) ability factor, *g*. Many intelligence researchers now prefer to focus on this more precisely specified and measured construct. They sometimes refer to it as *general* intelligence, by its more technical name (the general mental ability factor), or with an acronym less likely to provoke sensitivities over intelligence, such as GMA (general mental ability). Carroll identified eight broad abilities at the Stratum II level: fluid and crystallized *g* (G*f*, G*c*), memory and retrieval (G*y*, G*r*), visual and auditory processing (G*v*, G*u*), and two kinds of speed (G*s*, G*t*). All correlate highly with *g*, which means that each is basically *g* plus some domain-specific additive (e.g., memory, spatial visualization, speed of processing).

Other scholars use the unmodified noun "intelligence" more inclusively. Some include the entire set of human cognitive abilities, general and specific. For them, all the strata in Carroll's model collectively represent intelligence. Others extend the concept to abilities outside the cognitive realm, such as physical coordination (kinesthetic intelligence) and emotion perception and regulation (emotional intelligence). Yet others suggest that the term "intelligence" properly includes all traits and behaviors that contribute to effective performance (adaptive intelligence), or even life outcomes too (successful intelligence) – respectively Performance and Life Outcomes in Figure 9.1. All are important phenomena in their own right, but labeling some unspecified set of them as intelligence is more likely to confuse than inform research and theory.

The Meaning of g*, a Latent Construct*

As Figure 9.1 suggests, *g* can be described at different levels of analysis. Psychometrically, it is the first principal factor derived from factor-analyzing scores on professionally developed batteries of mental tests. *g* manifests itself in test and task performance as variation in a domain-general capacity for processing information. In lay terms it is

> a very general capacity that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience. It is not merely book learning, a narrow academic skill, or test-taking smarts. Rather, it reflects a broader and deeper capability for comprehending our surroundings – "catching on," "making sense" of things, or "figuring out what to do." (Gottfredson, 1997a, p. 13)

This description of *g* mirrors Horn and Cattell's construct of fluid intelligence (G*f*). In fact, factor analyses often find that the two *g*'s – Stratum III *g* and Stratum II G*f* – are essentially identical.

No mere chimera of factor analysis, *g* also manifests itself at the Gene and Brain levels of analysis (Gottfredson, 2016). For instance, the genetic structure of (covariation among) cognitive abilities mirrors its phenotypic structure, and the heritability of Stratum III *g* accounts for all but a fraction of the heritability of individual differences in Carroll's Stratum II abilities. Individual differences in *g* are radically polygenic, and no one allele (single nucleotide polymorphism, SNP) has more than a minuscule influence on phenotypic differences in *g*. Not only that, the heritability of *g* rises with age, meaning that phenotypes increasingly reflect differences in genotype (they correlate 0.9 by late middle age).

Virtually all differences in brain structure and function correlate to some extent with general cognitive ability measured in some fashion. Thus, far from being located in some particular part or process in the brain, the neural fundaments of *g* are radically dispersed throughout it. This fits with *g*'s radically polygenic origins. Psychometric *g* observed at the Intelligence level of analysis is an emergent property of the brain operating as a whole.

All human populations show a wide dispersion in general intelligence that replicates across generations. That this variation is so regular and recurring suggests that it is a biological fact, an evolved feature of our species. It is also a feature with wide and deep influence on human culture. The advantages of higher intelligence operate like a tailwind in virtually all life domains, weak in some but strong in others. It is a strictly cognitive trait, not affective, social, or physical. Data on test, school, and job performance all tell the same story: having higher *g* gives individuals a bigger edge in performing tasks well when the tasks (a) are instrumental (getting something done, correctly), (b) require more complex information processing, and (c) must be carried out independently. Conversely, low levels of *g* can create severe disadvantages for individuals, a stiff headwind, as they attempt to navigate our highly complex, bureaucratic, technological modern world.

The Auxiliary Role of Broad Stratum II Abilities

The broad Stratum II abilities seem specific either to content domains, such as language and spatial perception, or to different aspects of mental processing, such as speed, storage, and retrieval. Some are associated with particular Brodmann areas in the brain and some to the neural networks connecting them. Certain domain-specific Stratum II abilities, especially spatial perception, add to the prediction of performance (beyond *g*) in corresponding content domains, such as the visual arts and hard sciences. Yet, unless samples are highly restricted in range on *g* (e.g., students at elite colleges or in graduate programs), *g* always carries the freight of prediction in broad test batteries when predictors and outcomes are assessed reliably and objectively. The extent to which Stratum II or I abilities predict performance is owed primarily to their *g* component.

How Is Intelligence Best Measured?

It depends on your purpose.

Professional Practice

When the stakes are high for an assessed individual, standards call for a professionally developed, individually administered test of intelligence such as (in the United States) the Wechsler, Stanford-Binet, or Woodcock-Johnson. High-stakes assessment includes diagnosing the cognitive status of individuals referred for forensic or clinical evaluation, determining eligibility of children and adults for services, helping design individualized education programs (IEPs) for eligible students, and recommending treatment for individuals with cognitive impairments. IQ tests may be required, but are rarely sufficient for such purposes.

Group-administered assessments of general intelligence are more feasible and efficient when screening large numbers of individuals for jobs and training programs, especially the more cognitively demanding ones. Examples in the United States include the SAT and ACT college entrance exams, the Armed Services Vocational Aptitude Test Battery (ASVAB) for selecting and placing military recruits, and the Wonderlic Personnel Test (WPT) for screening job applicants (and major league football players). All these tests correlate highly with *g*, so function like an IQ test for ranking applicants. Only the 12-minute Wonderlic, however, is ever referred to as a test of intelligence. Again, scores on these tests are seldom the only factor in accepting or rejecting applicants. Intelligence tests have been designed to discriminate best in the range of IQ 70–130, so other assessments are required to assess the intellectual capabilities of individuals at the two extremes. For example, the SAT (normally administered for college entrance) is effective for identifying extremely gifted children.

Research

Researchers may not be able to use these highly vetted tests of intelligence in their work, perhaps for lack of access, resources, or participant time. Or the datasets they acquire do not include any of them. But Arthur Jensen dramatically increased our options (Gottfredson, 1998) when he reintroduced the concept of *g* and, more importantly, the insight that tests themselves can be assessed for how well they capture individual differences in *g* (their *g* loadedness). When empirical and theoretical precision is essential, say, in determining the neural architecture of general intelligence, the measure should be highly *g* loaded or at least its *g* loading should be known.

Standardized tests of academic achievement provide a close surrogate for IQ tests among native speakers of the same age, as do the U. S. Department of Education's national assessments of adult functional literacy (Gottfredson, 1997b). Other social indicators do not measure *g* directly, but help us view it operating unobtrusively in different life domains. When they differ in *g* loading, those differences can be exploited to test competing hypotheses about *g*. To illustrate, if personal health depends more on informed self-care than on wealth, then group disparities in health should line up closest with the most *g*-loaded indicators of socioeconomic inequality (education level, then occupational status, then income), but in reverse order if the "wealth-is-health" hypothesis is true (it is not; Gottfredson, 2004).

How Is Intelligence Best Developed?

Intelligence is a maximal trait – what a person *can* do when circumstances are favorable. That is what IQ tests are intended to measure, one's best. Developing intelligence can refer to raising one's maximum, working to one's maximum to develop specific skills and knowledge, or protecting it from preventable decline (Gottfredson, 2008). I see no compelling evidence that any educational, brain training, nutritional, or pharmacological intervention has yet been able to raise a person's maximal level of intelligence (*g*), either absolutely or relative to others their age. The apparent increases produced by education and training programs either do not generalize or they fade away. Nutritional interventions have produced mental and physical growth, but only among individuals with a nutritional (e.g., vitamin) deficiency; it is termed "catch-up growth."

Fully Exploiting Maximal Capacity

Many, perhaps most, individuals routinely function below their maximum. Thinking is hard work. If my students are any guide, many have never experienced working to their maximum (except on standardized tests) so do not even know what they are capable of until pushed. Exploiting one's intelligence more fully is a form of developing it: taking greater advantage of one's existing capacities to learn and accomplish more. Like other forms of capital, human capital is wasted if not invested.

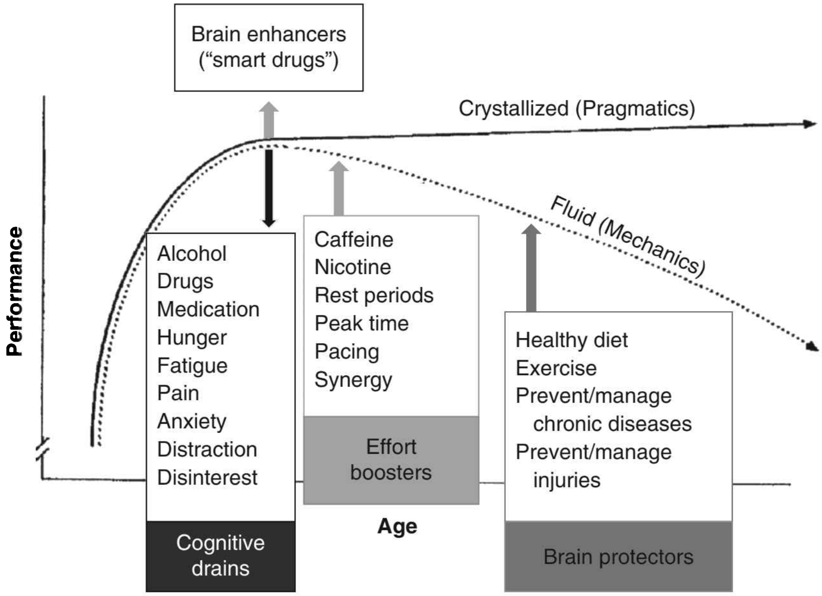


Figure 9.2 The fragility of maximal intelligence (Gottfredson, 2008).

The prospect of boosting maximal intelligence is enticing but remains remote. The closest we have come to smart pills are drugs that temporarily sustain effort and attention, such as Modafinil (for narcolepsy) or Ritalin (for attention deficit disorder). More widely used aids, listed in Figure 9.2, include caffeine and periodic rest periods. More consequential, in my view, are the many ways we unthinkingly squander the intellectual powers we already have. Alcohol, sleep deprivation, and distraction are some of the many ways we waste our available cognitive resources. They dull information processing.

Preventing Needless Decline in Maximal Capacity

Yet more insidious is our frequent failure to prevent needless cognitive damage and decline. Injury and chronic disease can damage the brain. Both are preventable with vigilance and effective self-care. Consider diabetes Type 2. It is epidemic but preventable, as are its debilitating complications. They include not just peripheral neuropathy, blindness, heart disease, and limb amputation, but also accelerated cognitive decline and increased risk of dementia.

What Are Some of the Most Interesting Empirical Results from Your Own Research and Why Are They Important to the Field?

My most novel work examines the interplay between the structure of human abilities (Intelligence in Figure 9.1) and a society's structure and inner workings (Social Structure). This work was essential for creating a more comprehensive theory of *g*, one that ties together all levels of analysis.

How Did Human Populations Evolve such a Finely Graded, g*-ordered Hierarchy of Occupations?*

When I got my PhD, sociologists had just developed a finely graded scale to quantify the standing of occupations according to perceived overall desirability or prestige. Not only did these ratings generalize across social groups in the United States, but across nations as well. The prestige of occupations correlates highly with their incumbents' mean levels of education, income, and IQ, as well as with U.S. Department of Labor ratings of work complexity. The occupational prestige scale provided a more tractable measure of occupational attainment than had broad categories of work (professional, semi-professional, and such). But no one asked how this astonishing regularity in human societies, or *social structure*, ever arose. Once upon a time there were only two occupations, hunter and gatherer.

If I was correct that the occupational hierarchy serves a functional purpose, not just the interests of a powerful elite, I needed to explain how it emerged from serving a society's needs. I also needed to answer reasonable objections, for instance, "If intelligence really is important on the job, why do years of education predict an individual's occupational level better than does IQ?"

Now, occupations are just recurring constellations of work tasks. They often split, disappear, or shift composition as technologies, industries, and cultures change. The evolution of these constellations is constrained, however, by a fixed feature of every society's labor pool: predictable, wide variation in *g*. My analyses of job attributes demonstrated that the core distinction among job demands (cognitive complexity, or *g* loadedness) mirrored the core distinction among human cognitive abilities (*g*). So, how might the constraints and opportunities created by recurring variation in *g* generate a cross-population *g*-ordered set of task constellations? The answer lay in the commonplace processes by which individuals become sorted, and sort themselves, to the different work tasks a society needs doing (Gottfredson, 1985).

When brighter individuals increasingly flow into a particular occupation, the tasks comprising it can re-assort, by *g* loading, to other jobs. We can observe in our own work settings how easier tasks migrate to easier jobs or less capable workers, while brighter workers are assigned or take on more cognitively demanding assignments. Thus do the two populations – occupations and workers – gradually align themselves along parallel continua: occupations by *g* loading and their incumbents by mean IQ/g. The hierarchy itself will expand or contract as workers become sorted more vs. less reliably by *g* to the occupational hierarchy. Other things obviously influence how occupations get structured and workers end up in them, but differences in *g* appear to be the most consistent factor. Stratum II abilities appear to distinguish among occupations only at the same cognitive level (Gottfredson, 1986).

What Makes Some Life Tasks and Outcomes More g *Loaded than Others?*

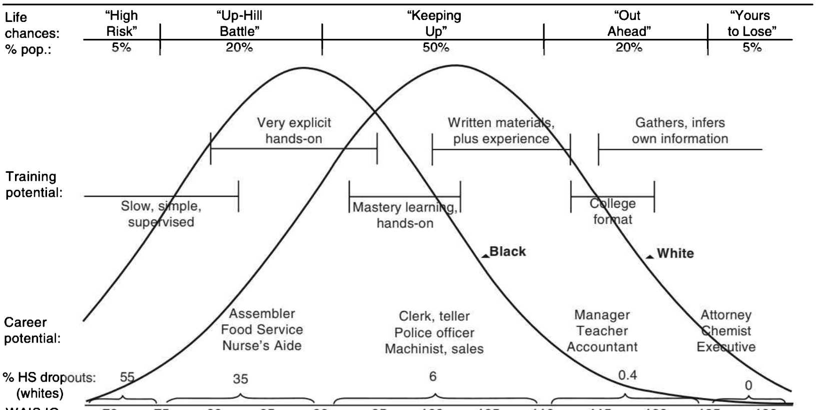
As noted earlier, the fact that IQ tests predict meaningful outcomes does not prove that whatever they measure actually caused those differences in outcomes. Moreover, IQ tests lack face validity; their items don't resemble anything familiar to the average person. Further, IQ scores are norm-referenced (calculated relative to a group average), so do not describe what people at different IQ levels can actually do in the real world. In this sense, mental test results are opaque, which limits their utility and fuels public doubt about what they really measure.

I did two things to enhance their interpretability. First, I collected what little information there was about individuals' trainability and life chances at different ranges of IQ in early adulthood. That information is summarized in the upper part of Figure 9.3. To explain this pattern, however, requires showing why *g* matters. So, second, I compiled data on jobs to see how the inherent demands of work might differ across tasks and jobs. What aspect of tasks would most strongly call forth the latent trait *g*, and to accomplish what?

Consider the definition of ability, as used to describe an attribute of individuals.

> [A]bility refers to the possible variations over individuals in the … levels of task difficulty … at which, on any given occasion in which all conditions appear favorable, individuals perform successfully on a *defined class of tasks.* (Carroll, 1993, p. 8, emphasis added)

Tasks define abilities. Task difficulty signifies ability level, that is, what a person *can* do. Carroll goes on to define *task* as an activity in which a



WAIS IQ:	70	75	80	85	90	95	100	105	110	115	120	125	130
Pop. percentile:		5th	10th	16th	25th	37th	50th	63rd	75th	84th	90th	95th	
Cumulative %													
White	2	4	7	13	22	33	46	60	72	82	90	95	97
Black	10	18	30	44	59	73	84	92	96	98	99	99+	99+
Hispanic	8	14	23	34	47	60	73	82	90	94	97	99	99+
Asian	1	2	4	8	14	23	34	47	60	73	82	90	94
Ratio in IQ range		IQ≤75			IQ≤90		IQ>100		IQ>110			IQ>125	
Black/white		5:1			3:1		1:3		1:7			1:30	
Hispanic/white		4:1			2:1		1:2		1:3			1:5	
Asian/white		2:3			2:3		6:5		4:3			2:1	

Figure 9.3 Life chances of young adults at different levels of the bell curve for general intelligence, by race (Gottfredson, 2005a, figure 18.2).
Adapted from Gottfredson (1997b, figure 3, p. 117) © Elsevier. Reproduced by permission of Elsevier. Permission to reuse must be obtained from the rights holder.

person engages to achieve specific objectives. It is purposeful work, be it using a map to reach a designated location or repairing a laptop.

Psychometricians, especially Arthur Jensen, had already pointed to the complexity of required information processing as the active ingredient, so to speak, in tests of intelligence. And so it is too in the world of work. As just noted, factor analyses of mental test scores revealed a general mental ability dimension, and factor analyses of job attributes demonstrated a corresponding distinction among jobs, namely, a complexity-of-work factor.

Factor loadings on the work complexity dimension indicate which particular mental processes and structural features of work contribute to a job's overall complexity. Starting with cognitive-processing tasks, loadings on the complexity dimension (essentially, their *g* loading) reflect the distinction between productive and reproductive thinking: higher for the importance of compiling (0.90), combining (0.88), and analyzing (0.83) information and lower for the importance of coding (0.68), transcribing

(0.51), remembering (0.40), and recognizing (0.36) it. Factor loadings of structural features reflected the importance of independent judgment and the ability to juggle more numerous and varied activities: importance of self-direction (0.88), lack of structure (0.77), lack of supervision (0.73), variety and change (0.41), and negatively with repetitive activities (-0.49; Gottfredson, 1997b).

Analyses of responses on national assessments of functional literacy also determined that item difficulty on scales of supposedly different constructs (Prose, Document, Quantitative, and Health literacies) rested on the same "processing complexity." At this item level of analysis, complexity is associated with more bits of information to integrate, more abstract concepts, more distant inferences, more irrelevant distracting information, and the need to select (not just implement) the correct arithmetic operation. Other findings reflected Spearman's "indifference of the indicator": neither the superficial content nor the readability (word length, sentence length) of task materials contributed to processing complexity. By design, literacy tests simulate everyday tasks, so have good face validity. Table 9.1 gives a gut-level feel for why differences in cognitive capacity matter in real life. I provide a specific example under Question 6.

How Could such a Highly General Information-Processing Ability, g*, Have Evolved so High so Fast in Pretechnological Human Groups?*

General intelligence varies widely within human populations, is dispersed throughout the brain, is a strictly cognitive tool, and has near-universal functional utility. Higher *g* is clearly an advantage for getting ahead and staying healthy in the modern world. But humans evolved their high intelligence in a pretechnological world of small, roving bands, and especially quickly (judging from skull size) beginning 500,000 years ago. No theory of *g* is complete without closing the circle in Figure 9.1, that is, without explaining both the accelerated rise in human intelligence and its sustained variation over the past half million years.

Evidence on *g* contradicted common hypotheses. *g* is strictly cognitive, which falsifies the social-brain hypothesis. The correlates of *g* are widely dispersed throughout the brain, which falsifies theories that intelligence evolved as a single module among many to meet specific adaptive challenges. Yet I struggled to find a plausible alternative. The problem lay in *g*'s most distinctive feature: its universal utility. Something in the human environment of evolutionary adaptation (EEA) had to be equally general and have consistently tilted the odds against less-bright individuals. What could that be?

Table 9.1 *Functional literacy of American adults, by age*

NALS difficulty level	Reading grade level*	%U.S. adults peaking at this level: Prose scale**				Simulated everyday tasks
		Age				Daily self-maintenance in modern literate societies
		16–59	60–69	70–79	80+	
5	16+	4	1	1	0	• Use calculator to determine cost of carpet for a room • Use table of information to compare two credit cards
4	16	20	8	5	1	• Use eligibility pamphlet to calculate SSI benefits • Explain difference between two types of employee benefits
3	12	35	27	19	6	• Calculate miles per gallon from mileage record chart • Write brief letter explaining error on credit card bill
2	7.2	25	33	22	27	• Determine difference in price between two show tickets • Locate intersection on street map
1	2.5	16	30	42	66	• Total bank deposit entry • Locate expiration date on driver's license

* Estimated by Carroll (1987).

** Results and sample items from the National Adult Literacy Survey (Kirsch, I. S., Jungeblut, Jenkins, & Kolstad 1993/2002). Literacy level reflects 80% probability of getting items at that level correct. The Prose, Document, and Quantitative scales on this assessment all correlate >0.9.

I finally realized that the tilt had to be tiny, inconspicuous, and typically affect one person at a time, because humans develop group defenses (such as food sharing) to protect individual members against obvious killers, like starvation. Studies of remaining (mostly) pretechnological human groups, such as the Ache of Paraguay, demonstrated a variety of such group defenses. I had already written about accident prevention being a quintessentially cognitive process (Gottfredson, 2004), and causes of death among the Ache illustrated the sometimes lethal consequences of cognitive error (not noticing poisonous snakes underfoot while hunting prey in the forest canopy, getting lost in the forest without a fire brand for cold nights).

Perhaps beginning with the invention of fire half a million years ago, human innovation began to generate evolutionarily novel hazards (risk of burns). Being novel, man-made hazards (such as falls from bridges, boats, and ladders; bites from domesticated animals) increased the relative risk of injury and death in the lower half of the intelligence distribution, as they still do worldwide. Ever more numerous and dangerous man-made hazards (weapons, poisons, vehicles) can explain humans' suddenly accelerated evolution of high intelligence. The mind's eye became ever more important to spot lurking hazards, imagine consequences, and avoid "accidents waiting to happen." All that evolution required to ratchet up our species' intelligence was for these novel hazards to increase the relative risk of crippling and fatal injuries among individuals of below-average ability, in turn resulting in them leaving relatively fewer genetic descendants behind (Gottfredson, 2007b).

The steady influx of man-made hazards into the human environment created a giant, increasingly *g*-loaded intelligence test administered to our species over hundreds of generations. No one type of accident or injury correlates noticeably with *g*. Each is like a barely *g*-loaded item on an IQ test. Non-*g* influences generally matter more in precipitating any particular event, but these non-*g* influences differ across events. However, *g* – using one's mind's eye to avoid injury – remains a consistent influence. When unintentional injuries are aggregated across all types of injury, over whole populations, and over long stretches of time, like adding items to a mental test, other influences on performance cancel out while the variance due to *g* grows. As the Spearman-Brown Prophecy Formula for test reliability tells us, even lightly *g*-loaded test items will, in sufficient number, create a highly *g*-loaded test when no other source of variance is as consistent as *g*.

What Do You See as the Most Important Educational or Social Policy Issue Facing the Field of Intelligence Today?

Public reluctance to entertain human variation in *g*, and the misinformation and fallacies promulgated to enforce it.

I have written often about this social phenomenon and how it can distort policy, practice, and science (e.g., Gottfredson, 1994a, 2007a, 2009). Persisting variation in *g* among a society's members creates the *democratic dilemma* (Gottfredson, 1996c). Free and democratic societies cannot simultaneously satisfy two guiding principles that intelligence differences put in conflict: equal opportunity and equal results. Politicians, academics, and pundits tend to firmly deny any such conflict or trade-off, often by denying the variation itself.

Denying a Consequential Biological Fact Does More Harm than Good

Some argue that to openly acknowledge intelligence differences, especially by race, would harm the body politic. They overlook the fact that denying human variation in intelligence does nothing to neutralize its inexorable, pervasive, observable effects in human affairs. Social policies and practices that deny it often create more problems, rancor, and suspicion of institutional discrimination than they dispel. Despite good intentions and high hopes, *g*-oblivious policies invariably disappoint and confound when *g* actually matters. Worse, interventions that aim to reduce social disparities in education and health usually increase them instead (Gottfredson, 2004).

Using Knowledge of g *to Predict Policies that Will Fail, and How*

The No Child Left Behind (NCLB) Act of 2001 exemplifies the waste and futility of *g*-denying social policy, especially in public education. In no other realm of public life is *g* so tightly linked to differences in performance. The NCLB required all public schools to make "adequate yearly progress" in getting all their students (including disadvantaged and special education students) leveled up by 2014 to the same state proficiency standards in reading and mathematics.

Data in the lower part of Figure 9.3 show why many schools were doomed to fail unless they gamed the system, as many did. Racial groups differ in their distribution of IQ. Based on their means and standard deviations in IQ test standardization samples in the United States, I estimated that the percentages of black, Hispanic, white, and Asian American students scoring below IQ 100 would be, respectively, 84, 73, 46, and 34. Group differences are even more striking at the two tails of the distribution.

Educationists had long argued that all students can learn equally well if their teachers are competent and their schools well funded. But now educators were protesting that some students are harder to teach than others. The states sought and the federal government began granting more waivers (may now exclude results of students in special education), learning opportunities withered for brighter students (whose good performance can widen performance gaps), schools creatively reclassified students (dropouts) to avoid reporting all low scores, and states dumbed down their proficiency tests to demonstrate progress in leveling up proficiency.

Spotting and Confronting the Use of Deceptive Science

There was tremendous political and legal pressure on employers in the 1980s and 1990s to use "nondiscriminatory" tests, meaning ones having no disparate impact (different pass rates by race or gender; Gottfredson & Sharf, 1988). Efforts to increase test reliability and validity had boomeranged because they tended to increase, not reduce, disparate impact by race by better measuring *g*. Adding personality tests to a selection battery hardly dented the disparate impact. The temptation to "psychomagic" grew.

Some selection professionals began advocating testing practices that reduced disparate impact by, in effect, reducing the reliability and validity of tests. The U.S. Employment Service (USES) began race-norming the General Aptitude Test Battery (GATB), which it used to refer better-qualified job applicants to employers. Race-norming calculates an individual's score relative to the average for his or her own race, which eliminates the normally large mean racial differences in test results. This in turn allowed USES to refer equal proportions of each race to participating employers. It had adopted the practice because setting different standards for different races did less damage to the validity of its referrals than being prohibited from using the test at all.

The National Research Council (NRC) created a blue-ribbon committee to assess the appropriateness of this practice. Although race-norming is an outright racial quota, the NRC committee gave a convoluted rationale for endorsing it in 1988 as "scientifically justified." Psychometrician Lloyd Humphreys described it in *Science* (1989, July 7, p. 14) as "statistical legerdemain."

In 1990, I received over-the-transom documents on the draft civil rights bill then under consideration in the U.S. Congress. The NRC's language had been slipped into the bill, which would have legally required employers to race-norm their selection tests. In addition, government agencies had already started threatening major companies if they failed to race-norm. Once these efforts to mandate race-norming were revealed (Blits & Gottfredson, 1990; Gottfredson, 1990, 1994b), Congress banned it instead. Now unable to race-norm the GATB, USES stopped using it.

Using Knowledge of g *to Identify and Explain Successes too Good to Be True*

No longer able to promote race-norming, the U.S. Department of Justice (DOJ) teamed up with nine eminent industrial-organizational psychologists to achieve the impossible: create a valid police selection test for Nassau County, NY, that had virtually no disparate impact – this despite the team documenting that police work is cognitively demanding. The DOJ immediately began forcing the team's "state-of-the-art," "innovative" test on police jurisdictions nationwide. I learned of the test only after receiving an author-blinded copy of the team's technical report, as did two other experts. Close examination of the foot-high report revealed that the test development team had dropped all subtests showing disparate impact *after* they had administered their large experimental battery to 25,000 applicants in Madison Square Garden.

This elite team had gerrymandered test battery content, post hoc, to eliminate all cognitive demands except reading above the first percentile (Gottfredson, 1996a, 1996b). Its technical report provided a labyrinth of questionable statistical procedures to claim, implausibly, that the new test was *more* valid than previous ones. But denuding a test of cognitive demands, ones actually experienced on the job, guts mental standards for all applicants. It leads to high rates of failure in training and subpar policing in all racial groups, as was observed.

DOJ scrapped the test in the ensuing scandal, but other consultants stood by eager to satisfy it.

What Are the Most Important Questions about Intelligence that Future Research on Intelligence Should Address?

If we take the journal *Intelligence* as our guide, basic research on intelligence falls mostly into two categories.

1. *Individual differences in phenotypic and genotypic intelligence:* for instance, how their expression changes from birth to old age; how both covary with observed individual differences in brain, cognitive processing, achievements, beliefs, attitudes, social behavior, cumulative life outcomes; and whether these relations generalize across all levels of *g* (Genes through Life Outcomes levels of analysis in Figure 9.1).

2. *Group differences in mean phenotypic intelligence (genetic differences by group are still taboo):* including, race and sex differences in the distribution and structure of intelligence; possible origins of intergenerational shifts in IQ scores (Flynn effect); how the mean IQ (or IQ surrogate) of geographic units (counties, states, nations) covaries with societal-level indicators of health, wealth, and social organization (Intelligence through Societal levels of analysis).

In contrast, virtually no intelligence research looks at how task environments activate this latent trait, *g*, to produce the observable differences in behavior that intelligence researchers study. Yet we cannot understand *g*'s remarkably pervasive and systemic, but varied, influence across the social landscape until we know how task environments are distributed, by complexity, across that landscape. Knowing the distribution of task environments by *g* also has practical value, especially as the distribution of *g*-loaded tasks shifts over time.

Modernity and technology have clearly made life more complex, even as they have made it better. The side effect, however, has been to put lower-*g* individuals at increasing disadvantage. For instance, treatment regimens for diabetes and other chronic diseases (now 7 of the 10 major causes of death in the United States) are becoming steadily more complex and hence more difficult for patients of all abilities levels to implement, but especially those in the lower reaches of IQ. Nonadherence to treatment is a huge problem. Patient errors are common, especially among low-literacy and older patients. Both groups have high rates of emergency room care and hospitalization.

Taking the sociological perspective, policy makers assume that "disparities" in health result from social inequality. Their solution is therefore to equalize financial, cultural, and physical access to medical care. As past experience and *g* theory predict, however, equalizing access to resources only increases the disparities it would eradicate. Disparities increase because higher-*g* individuals are better able to exploit newly available resources. This solution leaves less able individuals behind because it does nothing to increase their *cognitive* access of care.

Table 9.1 helps explain why cognitive access is crucial. It gives the percentage of adults who score at each of five levels of functional literacy, separately for four age groups. In health settings, Levels 1 and 2 are designated low literacy. The percentage of adults who can function no higher than Level 2 ranges from 41% at ages 16–59 to 93% among individuals age 80 and above. As Table 9.1 shows, these individuals cannot routinely perform tasks more difficult than "determine the difference in price between two show tickets" listed on a card. However, critical self-care tasks for patients with chronic diseases are often at Levels 3–5.

Consider diabetes self-management. The patient's job, all day every day, is to keep his or her blood glucose levels within safe limits and avoid health-damaging complications. This requires a lot of independent learning and judgment, planning and foresight, quickly spotting problems and taking corrective action, and adjusting self-care as circumstances change. All are attributes of complex occupations. And, worse, diabetes patients get little training, feedback, or supervision in performing this job.

Using insulin (or oral hypoglycemic) makes a patient's job even more complex. All patients with Type 1 diabetes must use insulin, and many with Type 2 do too as their condition worsens. Many need to adjust how much insulin they inject before each meal depending on the meal's carbohydrate content, their current blood glucose level, and the level of physical activity they anticipate. Insulin and oral hypoglycemics also make the patient's job more hazardous, because they form one of three classes of prescription medications with high rates of adverse drug events (the others being anticoagulants and opioids/benzodiazepines; U.S. Department of Health and Human Services, 2014). If patients miscalculate their insulin dose, administer it incorrectly, or fail to eat soon enough after they inject, they risk an insulin reaction (plunging blood glucose). Dangerously low blood glucose levels (severe hypoglycemia) can send patients to the emergency room or to their graves. This risk grows as age- and disease-related comorbidities, functional decline, and dysregulation of homeostatic mechanisms make self-care more difficult while reducing a patient's ability to self-manage.

It is not feasible, fair, or prudent to prescribe lower-ability and older patients such a complex, hazard-laden job. Error rates for individuals peaking at literacy Levels 1–2 range from 50% to 90% when carrying out tasks at complexity Levels 3–5 (Gottfredson & Stroh, 2016). Physicians are starting to recognize that treatment plans should be simplified and made safer for elderly patients and others at risk of adverse events (Mathur, Zammitt, & Frier, 2015; Munshi et al., 2016). Intelligence researchers can help healthcare providers accomplish this.

To illustrate, I am collaborating with certified diabetes educators and other health professionals to assay the complexity of self-care tasks and how they invite patient error among patients having low literacy or experiencing cognitive decline. We aim to develop two strategies. One, for paring back a patient's regimen to the tasks most essential for meeting his or her particular medical needs, while not overtaxing his or her functional capabilities. The second, for sequencing and pacing instruction to bring mastery of these tasks within the patient's cognitive reach.

Physicians and other healthcare professionals apparently receive no instruction in individual differences, so they tend to overestimate what less able individuals can do. They do not realize that what is obvious to them may not be obvious to their patients. Instruction must therefore assure that patients master basic facts (the meaning and relevance of carbohydrates, that "g" on a nutrition label means grams, that time-release pills work differently, not all "insulin" works the same way) before they are taught how to act on those facts: for instance, use a nutrition label to count grams of carbohydrate (*not* g of sugars or serving size), do not chew time-release pills, do not mix up your long- and short-acting insulins or use someone else's.

This is how *g* theory and research on task complexity can be used to improve health outcomes among our most vulnerable citizens and help contain the ballooning costs of health care. As Kurt Lewin (1943) said, there is nothing so practical as a good theory.

References

Blits, J. H., & Gottfredson, L. S. (1990). Employment testing and job performance. *The Public Interest*, Winter, 98, 18–25.

Carroll, J. B. (1987). The national assessments in reading: Are we misreading the findings? *Phi Delta Kappan*, 424–430.

(1993). *Human cognitive abilities: A factor-analytic study*. Cambridge: Cambridge University Press.

Gottfredson, L. S. (1985). Education as a valid but fallible signal of worker quality: Reorienting an old debate about the functional basis of the occupational hierarchy. In A. C. Kerckhoff (Ed.) *Research in sociology of education and socialization, Vol. 5* (pp. 119–165). Greenwich, CT: JAI Press.

(1986). Occupational aptitude patterns map: Development and implications for a theory of job aptitude requirements (Monograph). *Journal of Vocational Behavior,* 29, 254–291.

(1990, December 6). When job-testing "fairness" is nothing but a quota. *Wall Street Journal*, p. A18.

(1994a). Egalitarian fiction and collective fraud. *Society*, 31(3), 53–59.

(1994b). The science and politics of race-norming. *American Psychologist,* 49(11), 955–963.

(1996a). Racially gerrymandering the content of police tests to satisfy the U.S. Justice Department: A case study. *Psychology, Public Policy, and Law*, 2(3/4), 418–446.

(1996b, October 24). Racially gerrymandered police tests. *Wall Street Journal*, p. A16.

(1996c). What do we know about intelligence? *The American Scholar*, Winter, 15–30.

(1997a). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence*, 24(1), 13–23.

(1997b). Why g matters: The complexity of everyday life. *Intelligence*, 24(1), 79–132.

(1998). Jensen, Jensenism, and the sociology of intelligence. *Intelligence*, 26(3), 291–299.

(2004). Intelligence: Is it the epidemiologists' elusive "fundamental cause" of social class inequalities in health? *Journal of Personality and Social Psychology*, 86(1), 174–199.

(2005a). Implications of cognitive differences for schooling within diverse societies. In C. L. Frisby & C. R. Reynolds (Eds.) *Comprehensive handbook of multicultural school psychology* (pp. 517–554). New York: Wiley.

(2005b). Using Gottfredson's theory of circumscription and compromise in career guidance and counseling. In S. D. Brown & R. W. Lent (Eds.) *Career development and counseling: Putting theory and research to work* (pp. 71–100). New York: Wiley.

(2007a). Applying double standards to "divisive" ideas. *Perspectives on Psychological Science*, 2(2), 216–220.

(2007b). Innovation, fatal accidents, and the evolution of general intelligence. In M. J. Roberts (Ed.), *Integrating the mind: Domain general versus domain specific processes in higher cognition* (pp. 387–425). Hove, UK: Psychology Press.

(2008). The fragility of maximal performance. Presented at the conference, "How can we improve our brains?" The Banbury Center, Cold Spring Harbor, NY, September 16. Retrieved from www1.udel.edu/educ/gottfredson/reprints/2008fragility.pdf.

(2009). Logical fallacies used to dismiss the evidence on intelligence testing. In R. Phelps (Ed.), *Correcting fallacies about educational and psychological testing* (pp. 11–65). Washington, DC: American Psychological Association.

(2016). Hans Eysenck's theory of intelligence, and what it reveals about him. *Personality and Individual Differences*, 103, 116–127. doi:10.1016/j.paid.2016.04.036.

Gottfredson, L. S., & Sharf, J. C. (Eds.) (Special Issue) (1988). Fairness in employment testing. *Journal of Vocational Behavior*, 33 (3).

Gottfredson, L. S., & Stroh, K. (2016). How to select or create materials your patients will actually understand. Pre-conference workshop conducted at the annual meeting of the American Association of Diabetes Educators, San Diego, August 11.

Kirsch, I. S., Jungeblut, A., Jenkins, L., & Kolstad, A. (1993/2002). *Adult literacy in America: A first look at the findings of the National Adult Literacy Survey.* U.S. Department of Education, National Center for Education Statistics, Washington, DC. Retrieved from http://nces.ed.gov/pubs93/93275.pdf.

Lewin, K. (1943). Psychology and the process of group living. *Journal of Social Psychology*, 17, 113–131.

Mathur, S., Zammitt, N. N., & Frier, B. M. (2015). Optimal glycaemic control in elderly people with Type 2 diabetes: What does the evidence say? *Drug Safety*, 38, 17–32. doi:10.1007/s40264-014-0247-7.

Munshi, M. N., Slyne, C., Segal, A. R., Saul, N., Lyons, C., & Weinger, K. (2016). Simplification of insulin regimen in older adults and risk of hypoglycemia. *JAMA Internal Medicine*, 176(7), 1023–1025. doi:10.1001/jamainternmed.2016.2288.

U.S. Department of Health and Human Services. (2014). *National action plan for adverse drug event prevention*. Washington, DC: Office of Disease Prevention and Health Promotion. Retrieved from https://health.gov/hcq/pdfs/ade-action-plan-508c.pdf.