Intelligence: success and fitness

David Lubinski

Department of Psychology and Human Development, Vanderbilt University, Nashville, TN 37203, USA

Abstract. This chapter presents the consensus among psychometricians regarding the construct of general intelligence ('*g*') and its measurement. More than any other construct, *g* illustrates the scientific power of construct validation research. To date, *g* is carried by more assessment vehicles and saturates more aspects of life than any other dimension of human variation uncovered by psychological science. Phenomena most vital to the core of *g*'s nomological network are reviewed (e.g. abstract learning, information processing, and dealing with novelty). This is followed by coverage of relevant but more peripheral phenomena (e.g. crime, health risk behaviour, and income). Because *g* constitutes such a ubiquitous aspect of the human condition, its omission in social science research often results in underdetermined causal modelling. Frequently, this constitutes a longstanding error in inductive logic, namely, the Fallacy of the Neglected Aspect. Attending to Carnap's Total Evidence Rule can help to forestall neglected aspects in scientific reasoning.

2000 The nature of intelligence. Wiley, Chichester (Novartis Foundation Symposium 233) p 6–36

This symposium is about the nature of intelligence and its evolutionary significance. Some might find it surprising, then, that Sigmund Freud's famous remark that life consists of loving and working, *lieben* and *arbeiten*, sets the stage for my presentation. Yet, these two important domains are good starting points for illustrating the scope of individual differences in general intelligence, which is the topic of my contribution. Assortative mating coefficients for general intelligence (or '*g*') approximate 0.50 (Plomin & Bergeman 1991); and *g*'s predictive validity for work performance surpasses this value as occupations become more fluid in terms of their complexity and novelty—that is, as they become more conceptually demanding (Schmidt & Hunter 1998). The evolutionary concept of fitness, however one chooses to define it, certainly would involve at least these two major components: mating and resource acquisition. To be successful in these endeavours requires learning and, in particular, as industrialized cultures have developed and evolved into the information age, learning demands have concentrated on solving abstract problems. In his award-winning book, *Will we be smart enough?*, Hunt (1995) refers to people especially able at learning abstract

relationships (and solving abstract problems) as 'symbol analyzers'. Historically, the psychology of individual differences has simply referred to them as intellectually gifted or talented (Benbow & Stanley 1996). The reason *g* accounts for more variance than any other personal characteristic in Freud's two chief realms of human endeavour is probably because, at least as much as anything else, *g* reflects individual differences in rate of learning abstract relationships (Carroll 1997).

Freud's statement identifies significant purviews of human activity. Functioning effectively within the dominant spheres of achievement and interpersonal relations is critical for general psychological well being, as well as for general biological survival. If cross-fertilization between differential psychology and evolutionary psychology has the potential to produce scientifically viable offspring, targeting psychological realms that both disciplines care about and are prominent features of the human condition will facilitate the process. In this regard, it is hard to imagine better arenas than mating and resource acquisition. None the less, the centrality of *g* is equally pervasive in many other facets of human life (Jensen 1998, Lubinski 2000). That needs to be emphasized.

For example, throughout this century, several personality theorists have pointed out that *g* constitutes an important dimension of psychological diversity relevant to molar behaviour (i.e. general personological functioning). When personality is viewed as a system of longitudinally stable behavioural tendencies that operate across situations, *g* clearly constitutes a substantively significant feature of the total personality (Lubinski 2000). Raymond B. Cattell (1950), arguably Spearman's most famous student, thought so; and Stark Hathaway, inventor of the most widely used personality inventory, the Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley 1940), thought so as well. Hathaway was a brilliant diagnostician whose clinical acumen was legendary (Nichol & Marks 1992). As one of his students recalls, Hathaway would always tell his clinical advisees: 'We tend to think of general intelligence in isolation, as if it only operated in educational and vocational contexts; yet, it is a salient aspect of personality that saturates almost everything we do.' (P. E. Meehl, personal communication, July 1993.)

The preamble above serves as an important introduction to my discussion of general intelligence. It is intended to forestall concerns that we are discussing a molecular strand of human diversity (e.g. book learning). To the contrary, the nomological network of the *g* construct is broader and deeper than any other systematic source of individual differences uncovered by psychological science to date. Moreover, its conceptual underpinnings were embryonically embedded in differential psychology's origin. Given this, and because differential and evolutionary psychology germinated from common soil, it might be useful to review certain key antecedents to modern treatments of the *g* factor (Gottfredson 1997, Jensen 1998).

Some background

Early on, Francis Galton, the father of differential psychology, held that a *general* dimension was central to many academic achievements as well as subsequent developmental trajectories throughout life but especially in the world of work. He also appears to have believed that psychological assessment should focus on attributes that operate widely. In Galton's (1869/1961) words:

> In statesmanship, generalship, literature, science, poetry, art, just the same enormous differences are found between man and man; and numerous instances recorded in this book, will show in how small degree, eminence, either in these or any other class of intellectual powers, can be considered as due to purely special powers. They are rather to be considered in those instances as the results of concentrated efforts, made by men who are widely gifted. People lay too much stress on apparent specialties, thinking over-rashly that, because a man is devoted to a particular pursuit, he could not possibly have succeeded in anything else. They might just as well say that, because a youth has fallen desperately in love with a brunette, he could not possibly have fallen in love with a blonde. (p 7)

In addition to postulating a general cognitive ability, Galton stressed that measures of intellective functions should forecast something important outside of assessment contexts. After examining, for example, Cattell's (1890) classic, *Mental tests and measurements*, wherein the term 'mental test' was first introduced, Galton (1869/1961) appended two pages of profoundly influential remarks.

> One of the most important objects of measurement is hardly if at all alluded to here and should be emphasized. It is to obtain a general knowledge of the capacities . . . by sinking shafts, as it were, at a few critical points. In order to ascertain the best points for the purpose, the sets of measures should be compared with an independent estimate. . . . We thus may learn which of the measures are the most instructive. (p 380)

Looking back, these remarks set the stage for subsequent *construct validation research* (Cronbach & Meehl 1955): the phrase 'independent estimate' anticipated external validation, whereas 'most instructive' depicts an empirically based form of competitive support. To Galton (1890), like subsequent participants of the mental measurement movement, scientific measures were not all seen as equally important (Lubinski 1996). Importance was to be calibrated against the breadth

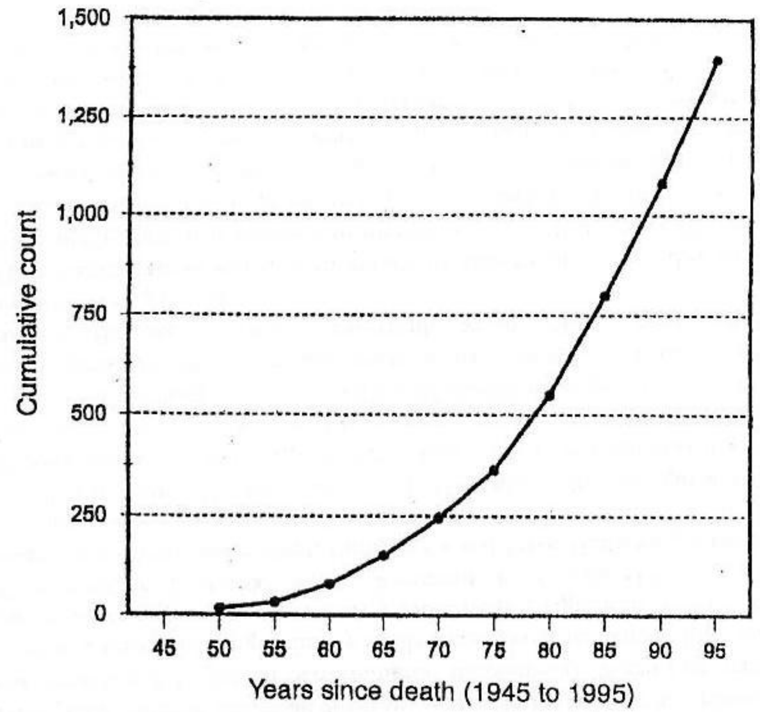


FIG. 1. Spearman's cumulative citation count, plotted in five-year blocks, beginning with the year of his death 1945. (From Jensen 2000.)

and depth of external relationships achieved outside of assessment contexts. Furthermore, scientifically significant sources of human variation should manifest predictive validity over extended temporal gaps (Lubinski 2000, Schmidt & Hunter 1998). *Predictive* validity was seen as an important aspect of construct validity because, by generating impressive forecasts, the former helped confirm the latter. Evaluating the psychological character of individuals or groups involves assessing longitudinally stable attributes. Longitudinally stable attributes are the ones that lend themselves to *proximal selection* (*phenotypes* for securing mates and resources through social selection) and *ultimate selection* (*genotypes* for perpetuating adaptive characteristics through natural selection).

The *g* factor: instrumentation and measurement

Shortly after Galton's (1890) remarks, Spearman (1904) published one of the psychological landmarks of the twentieth century: "'General intelligence,' objectively determined and measured", which supported Galton's view. Here, the concept of a general factor was spawned. At the phenotypic level, modern tests of general intelligence index essentially the same construct uncovered herein — albeit with much more efficiency and precision (Carroll 1993). We also know a great deal more about the correlates of g amassed from 95 years of research (Gottfredson 1997, Jensen 1998). That Spearman's thinking continues to influence modern thought is revealed by his cumulative citation count, compiled in five-year blocks, beginning with the 1945–1950 interval (Fig. 1). The positive acceleration of Spearman's current impact is remarkable for a psychologist who died in 1945.

Spearman–Brown prophesy formula

Important psychological constructs saturate multiple assessment vehicles because they operate in multiple aspects of life. Important constructs are ubiquitous. They are typically 'with us'. But because human behavioural determinants are both multilevel and multifaceted, evidence for even general psychological constructs needs to be secured through technical scientific instrumentation. This is not unique to psychological assessment, as E. O. Wilson (1998) observed in *Consilience*,

> Without the instruments and accumulated knowledge of the natural sciences — physics, chemistry, and biology — humans are trapped in a cognitive prison. They invent ingenious speculations and myths about the origin of the confining waters, of the sun and the sky and the stars above, and the meaning of their own existence. But they are wrong, always wrong, because the world is too remote from ordinary experience to be merely imagined. (p 45)

The discipline of psychometrics has developed instruments for dealing with psychological phenomena remote from personal experience. Psychological constructs are 'removed' from experience because they co-occur with other phenomena. Multiple behavioural episodes are necessary to detect them. Psychometric procedures sample responses to discrete performance opportunities (responses to items) and uncover dominant dimensions running through them (through aggregation). Each behavioural sample (test item) usually contains a tiny amount, say, 4% *construct-relevant* variance and, hence, a preponderance (96%) of *construct-irrelevant* variance. Most of any particular response is largely noise relative to signal. But Spearman's celebrated formula, discovered contemporaneously with Brown (hence, Spearman–Brown prophesy formula), interchanges these percentages (i.e. it interchanges the preponderance of noise to signal) through aggregation:

$$r_{tt} = (k)r_{xx}/[1 + (k-1)r_{xx}]$$

where: r_{tt} = the proportion of common variance in a composite of items, k = the number of items, and r_{xx} = the average inter-item correlation.

Psychometric procedures like Spearman–Brown provide psychological windows on human variation akin to the microscope in biology and the telescope in astronomy. With an average inter-item correlation of 0.20, a mere 40 item scale can generate a composite whose common (reliable) variance is 91%. This is how psychometrics distils dimensions of common variance for submission to construct validation procedures.

Labelling scales

Once a reliable source of individual differences has been established (e.g. $r_{xx} = 0.91$) attention naturally turns to its psychological nature — or, construct validity. That the same construct may run through ostensibly distinct assessment vehicles and generate functionally equivalent external relationships is implicit in *convergent validity* (Campbell & Fiske 1959). Otherwise, construct validation would not work. Construct validity implies multiple vehicles (convergent validity) and heterogeneous criterion families (for establishing nomological networks).

To illustrate how the same construct may run through varying mediums, Table 1 presents three measures of verbal ability all assessed with different item types: reading comprehension, literary information and vocabulary. Yet, in the context of a heterogeneous collection of external criteria, they behave as functionally equivalent measures of the same underlying construct. They can be used interchangeably, yet, superficially, they appear to be measuring different qualities. Note also how these cognitive measures co-vary with distinct measures of mathematical and spatial ability as well as a variety of information tests. Careful sampling at this level of generality (technically, *systematic heterogeneity*) is how the construct of general intelligence is uncovered (Lubinski & Dawis 1992, Lubinski & Humphreys 1997).

Figure 2 illustrates how g is distilled by systematically aggregating content distinct groupings of ability mediums (quantitative–numerical, spatial–pictorial and verbal–linguistic). In this illustration, each of the three ability measures manifests 90% reliable variance (i.e. r_{xx}=0.90). However, the preponderance of each scales reliable variance is specific (unshared with the other two). Yet, when all three are aggregated in a composite, the amount of specificity associated with each is attenuated, and the resulting amalgam primarily consists of what they have in common (viz., a general factor, 'g'). This is how psychometricians distil *general factors* from assessment vehicles whose reliable variance is primarily specific.

Probably the most impressive review of indicators of g is Carroll's (1993) book on the 20th century's factor analytic work (Fig. 3). This figure nicely illustrates the hierarchical aggregation developed here: items → scales → general constructs.

TABLE 1 Extrinsic convergent validation profiles across three measures having verbal content

	Literature	*Vocabulary*	*Reading comprehension*
Aptitude tests			
Mechanical reasoning	0.43	0.52	0.54
2-D visualization	0.25	0.32	0.35
3-D visualization	0.35	0.43	0.47
Abstract reasoning	0.45	0.53	0.61
Arithmetic reasoning	0.54	0.63	0.63
High-school maths	0.57	0.59	0.57
Advanced maths	0.42	0.43	0.39
Information tests			
Music	0.67	0.68	0.62
Social studies	0.74	0.74	0.71
Mathematics	0.62	0.63	0.57
Physical sciences	0.64	0.67	0.60
Biological sciences	0.57	0.61	0.56
Interest			
Physical sciences	0.24	0.25	0.22
Biological sciences	0.26	0.25	0.22
Public service	0.16	0.12	0.12
Literary-linguistic	0.37	0.32	0.32
Social service	0.07	0.06	0.07
Art	0.32	0.30	0.29
Music	0.23	0.20	0.20
Sports	0.12	0.12	0.13
Office work	−0.35	−0.29	−0.27
Labour	−0.08	−0.06	−0.06

These correlations were based only on female subjects (male profiles are parallel). $N=39\,695$. Intercorrelations for the three measures were the following: literature/vocabulary = 0.74, literature/reading comprehension = 0.71 and vocabulary/reading comprehension = 0.77. (From Lubinski & Dawis 1992.)

Among the more important points illustrated here is the abstract nature of *g*. When heterogeneous collections of cognitive tests are aggregated, they form a general factor relatively free of any particular content or product, which accounts for approximately 50% of their common variance. This occurs when pictorial, quantitative and verbal item types are administered individually or in group, orally or by paper and pencil. Because it is general, *g* can be assessed in many

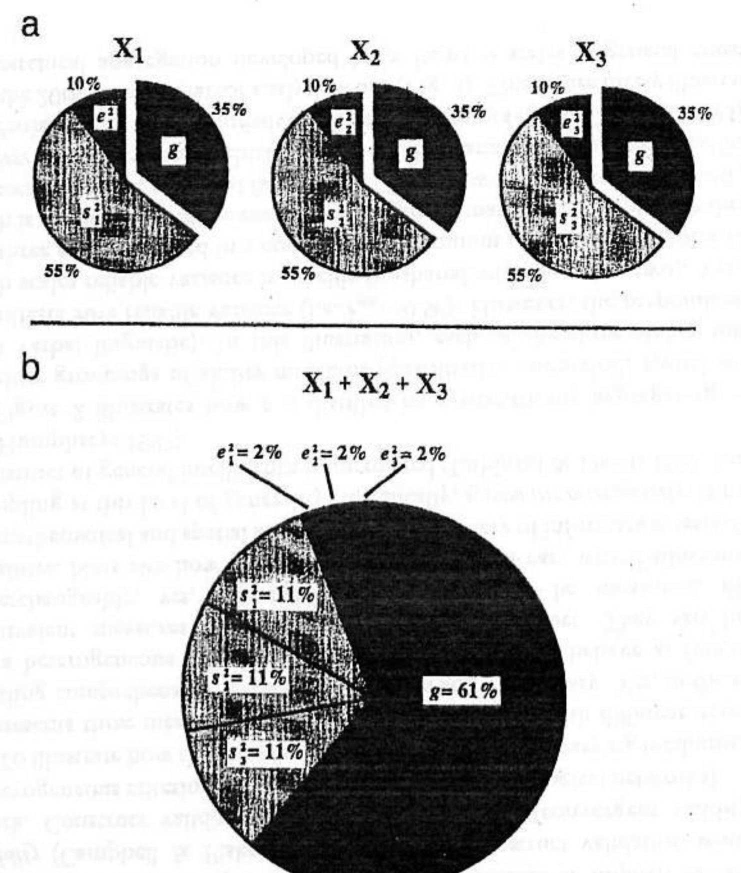


FIG. 2. (a) Three variable variate. Three hypothetical variables having the same amount of common, specific and error variance. As individual components of a predictor variate, most of the variance of each component is specific variance. (b) Three variable composite. When the three components found in (a) are aggregated, most of the composite's variance is variance shared with a general factor common to each. Moreover, the influence of any one form of specificity is considerably reduced. (From Lubinski & Dawis 1992.)

different ways; for the same reason, *g* extends to many different kinds of life events. It is important to keep in mind, however, that multiple ways of assessing general intelligence typically converge on a common core, just as the verbal measures in Table 1 do. That is, ostensibly disparate assessment procedures can eventuate in functionally equivalent measuring operations.

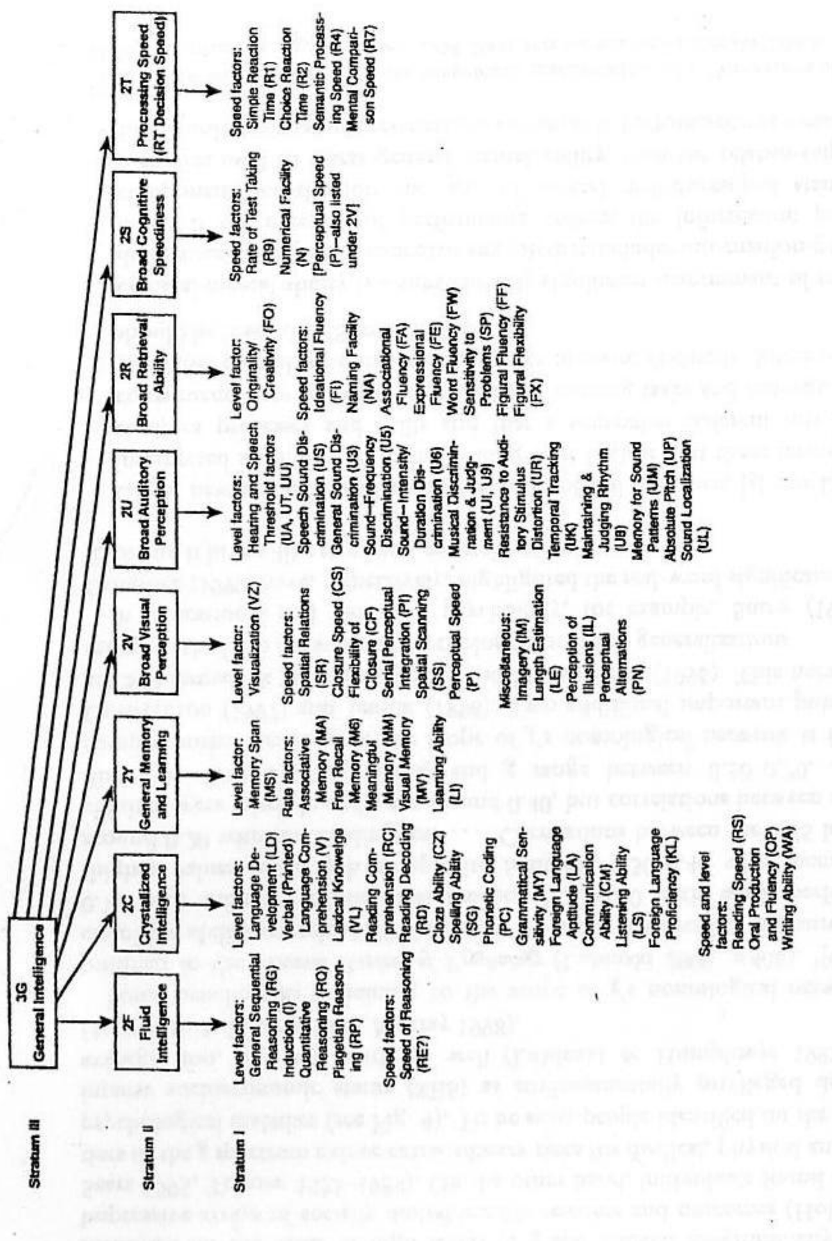
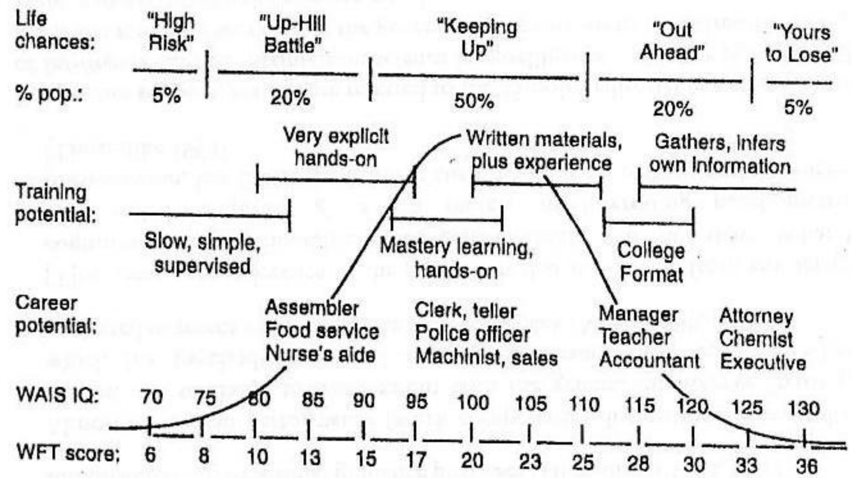


FIG. 3. The three-stratum structure of cognitive abilities. Reproduced from Carroll (1993).

Construct validity: *g*'s nomological network

The scientific meaning of constructs accrues from the role they play in nomological networks. Construct validation proceeds by establishing functional relationships between assessment vehicles and external criteria with the ultimate aim of tracing the causal directionality of these functions.



High school dropout (%)	55	35	6	0.4	0
Lives in poverty (%)	30	16	6	3	2
Chronic welfare recipient (mothers) (%)	31	17	8	2	0
Had illegitimate children (women) (%)	32	17	8	4	2
Went on welfare after first child (women) (%)	55	21	12	4	1
Ever incarcerated (men) (%)	7	7	3	1	0

FIG. 4. Overall life chances at different ranges of IQ. WAIS IQ, intelligence quotients obtained from the Weschler Adult Intelligence Scale (Weschler 1981); WPT, Wonderlic Personnel Test, Inc. (1992). Reproduced from Gottfredson (1997) with permission.

That *g* displays functional relationships with many different kinds of important outcomes and events is well known (Gottfredson 1997, Jensen 1998). People identified on the bases of high levels of *g* and tracked longitudinally display impressive arrays of socially desirable achievements and outcomes (Holohan & Sears 1995, Terman 1925–1959). On the other hand, individuals found at lower tiers of the *g* spectrum evince extraordinary risks for medical, physical and social-psychological maladies (see Fig. 4). To be sure, people identified on the bases of intense socioeconomic status (SES) as environmentally privileged do above average, too, but not nearly as well (Lubinski & Humphreys 1992, 1997, Herrnstein & Murray 1993, Murray 1998).

Some benchmarks pertaining to the scope of *g*'s nomological network are outlined in the *Annual Review of Psychology* (Lubinski 2000, p 408), '[G]eneral cognitive ability covaries 0.70–0.80 with academic achievement measures, 0.40–0.70 with military training assignments, 0.20–0.60 with work performance (higher values reflect job complexity families), 0.30–0.40 with income, and around 0.20 with law abidingness. . . . ' Correlations between the SES level that children were raised in and *g* are around 0.40, but correlations between achieved SES (i.e. their SES as adults) and *g* range between 0.50–0.70. A more comprehensive detailing of the scope of *g*'s nomological network is found in Gottfredson (1997) and Jensen (1998). Two additional important publications are Snyderman & Rothman (1987) and Neisser et al (1996). This network has served as the basis for some of psychology's broadest generalizations.

In educational and industrial psychology, for example, Snow (1989) and Campbell (1990) have, respectively, highlighted the real-word significance of *g* by featuring it in law-like empirical generalizations.[1]

> Given new evidence and reconsideration of old evidence, [*g*] can indeed be interpreted as 'ability to learn' as long as it is clear that these terms refer to complex processes and skills and that a somewhat different mix of these constituents may be required in different learning tasks and settings. The old view that mental tests and learning tasks measure distinctly different abilities should be discarded (Snow 1989, p 22).

> General mental ability is a substantively significant determinant of individual differences in job performance for any job that includes information-processing tasks. If the measure of performance reflects the information processing components of the job and any of several well-developed standardized measures used to assess general mental ability, then the relationship will be found unless the sample restricts the variances in performance or mental ability to near zero. The exact size of the relationship will be a function of the range of talent in the sample and the degree to which the job requires information processing and verbal cognitive skills (Campbell 1990, p 56).

[1]This review will be restricted to the behavioural manifestations of *g*. For reviews of the many biological correlates of *g*, see Jensen (1998, 2000, this volume) and Lubinski (2000).

Because of the foregoing considerations, *g* has achieved the status of differential psychology's most central dimension. Clearly, other things matter; complex human behaviour is almost always multiply determined. In fact, differential psychology has uncovered a number of 'deep shafts' in the realm of human ability (group factors) distinct from *g* (Achter et al 1999, Humphreys et al 1993, Messick 1992). And in conceptually distinct domains, such as vocational interests and personality (proper), scientifically valuable dimensions of human variation also are prevalent and well known (Lubinski 1996, 2000); yet, their psychological significance pales when contrasted to that of general intelligence (Gottfredson 1997, Jensen 1998). Indeed, in the words of some of the most distinguished psychometricians of all time:

> The general mental test stands today as the most important technical contribution psychology has made to the practical guidance of human affairs (Cronbach 1970, p 197).

> [A general] intelligence test is the single most important test that can be administered for vocational guidance purposes (Humphreys 1985, p 211).

> Almost all human performance (work competence) dispositions, if carefully studied, are saturated to some extent with the general intelligence factor *g*, which for psychodynamic and ideological reasons has been somewhat neglected in recent years but is due for a comeback (Meehl 1990, p 124).

> [T]he great preponderance of the prediction that is possible from any set of cognitive tests is attributable to the general ability that they share. What I have called 'empirical *g*' is not merely an interesting psychometric phenomenon, but lies at the heart of the prediction of real-life performances. (Thorndike 1994).

For further support, readers are referred to the 25-point editorial in a special issue of *Intelligence* entitled 'Mainstream science on intelligence', which is signed by 52 academic scientists working in the general intelligence arena (Gottfredson 1997). Here, general intelligence was described as:

> . . . a very general mental capability that, among other things, involved the ability to reason, plan, solve problems, think abstractly, comprehend complex

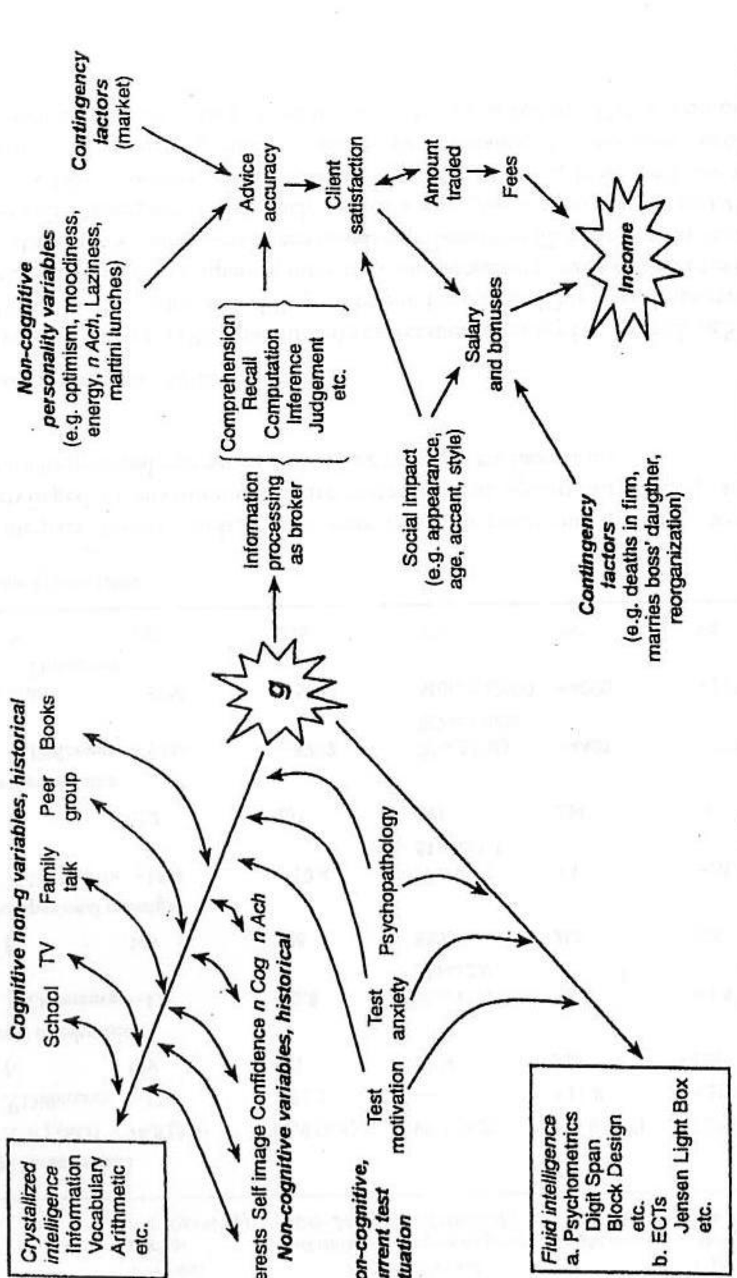


FIG. 5. Some stochastic relationships between *g* and external phenomena at various degrees of remoteness. Although formal analytic procedures for demarcating regions within theoretical networks are currently unavailable, near relationships are closer to the essence of constructs under analysis ('*g*') and constitute what Carnap (1950) referred to as 'meaning postulates'. Reproduced with permission from P. E. Meehl (personal communication, 1999).

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. (p 13)

This is a reasonable statement. And it is consistent with Spearman's (1927) 'eduction of relations and correlates'. Yet as Meehl (1998) points out, verbal definitions (Sternberg & Detterman 1986) are always problematic because they lack consensus. A scientific understanding of a construct is achieved by placing measurement operations of purported attributes into nomological networks and examining their functional relationships with meaningful external criteria. Meehl (personal communication, 1999) has offered a sketch illustrating some core phenomena central to the 'meaning' of *g* as well as more remote phenomena (Fig. 5).

Notice how, moving from 'pure' information processing capability and abstract learning to more distal phenomena, the relationship between *g* and more remote phenomena breaks down. This is understandable, because large temporal gaps open up possibilities for many different kinds of intervening events (Humm 1946). This can be played out a bit more. Given our knowledge of the many different kinds of external and internal influences operating within the *g* nexus, an important consideration is how much covariation one should anticipate between vehicles assessing *g* and construct-relevant criteria. To the extent that investigators are disappointed by covariational patterns generated by this ability construct (or the *g*-factor 'take' on human ability), the implication is that ability should account for more variance; and hence, non-ability attributes should account for less. Nevertheless, the question should be asked. Upon examining the functional relationships in *g*'s nomological network (and given the role that energy, health, interests, personality, psychopathology and chance factors, etc., play as determinants of complex behaviours and outcomes—to mention but a few non-ability attributes), 'Is this a reasonable amount of covariation for one variable to achieve?'

Is *g* a causal entity?

A concern about the above correlates is that they do not imply causation. Because *g* covaries with other purported causal determinants of social outcomes, for example SES, the causal antecedents of its correlates are equivocal. In Terman's (1925–1959) famous longitudinal study of intellectually gifted participants, for example, subjects were appreciably above the norm on several educational and vocational criteria. On various indices of physical and psychological health, they were also significantly better off than normative base rate expectations would lead one to

TABLE 2 **Paired sibling sample comparisons**

	Cognitive class				
	Very dull siblings (*<10th %ile*)	*Dull siblings* (*10th–24th*)	*Normal reference group* (*25th–74th*)	*Bright siblings* (*75th–89th*)	*Very bright siblings* (*≥90th %ile*)
IQ characteristics					
$\overline{X}$ IQ (SD)	74.5 (5.4)	85.9 (2.5)	99.1 (5.9)	114.0 (2.7)	125.1 (5.6)
$\overline{X}$ Difference	−11.2	−21.1	—	+11.8	+21.8
N	199	421	1074	326	128
Years of education					
$\overline{X}$ Difference	−1.6	−0.8	$\overline{X}$=13.5 SD=2.0	+1.3	+1.9
N	149	326	850	266	109
Occupational prestige					
$\overline{X}$ Difference	−18.0	−10.4	$\overline{X}$=42.7 SD=21.5	+4.1	+10.9
N	102	261	691	234	94
Earned income					
$\overline{X}$ Difference	−9462	−5792	$\overline{X}$=23703 SD=18606	+4407	+17786
Mdn Difference	−9750	−5000	Mdn=22000	+4000	+11500
N	128	295	779	257	99

From Murray (1998).

anticipate. Nevertheless, critics were quick to point out that they were more privileged in environmental circumstances and opportunity (SES); and thus, launching causal inferences emanating from *g* were hazardous.

Isolating cognitive abilities from SES

Recently, Murray (1998) has offered a clever methodology for untangling SES from ability–performance and ability–outcome functions. This is an important study because, of all the competing interpretations attached to causal inferences assigned to the *g*-factor, the hypothesized causal significance of SES has been by far the most prevalent competitor. So much so that it even has a name: *the sociologist's fallacy*. (This fallacy occurs when causal inferences emanating from SES are ventured without considering other possible determinants, for example, endogenous personal attributes like *g*.) Here is how Murray cracked the IQ/SES conundrum.

TABLE 3 **Paired sibling sample comparison**

	Cognitive class				
	Very dull siblings (*<10th %ile*)	*Dull siblings* (*10th–24th*)	*Normal reference group* (*25th–74th*)	*Bright siblings* (*75th–89th*)	*Very bright siblings* (*≥90th %ile*)
Bachelor's degrees					
For reference siblings without a BA					
Comparison siblings with a BA	1%	1%	(0%)	42%	59%
n	177	339	811	220	75
For reference siblings with a BA					
Comparison siblings with a BA	0%	18%	(100%)	76%	91%
n	19	55	198	78	46

From Murray (1998).

Murray (1998) analysed data from the National Longitudinal Survey of Youth (NLSY). NLSY consists of 12 686 subjects born between 1957–1964, with successive cohorts assessed, beginning in 1979, on the Armed Forces Qualifying Test (AFQT). AFQT was converted to IQ estimates normalized around a mean of 100 and a standard deviation of 15.[2]

Using NLYS 15-year longitudinal data, Murray (1998) studied outcome differences between biologically related siblings (reared together) but who differed in general intelligence (see Table 2). He compared various outcomes of 1074 sibling pairs. Each sibling-pair consisted of a member within the normal IQ range (25–74%) and a sibling in one of the following four IQ ranges: very dull (less than 10%), dull (10–24%), bright (75–89%), and very bright (greater than 89%). As ability differences between siblings increased, so did differences in socially valued outcomes. Table 2 illustrates only some of Murray's (1998) measured differences (years of education, occupational prestige and earned income), which mirror those in the general population at similar ability ranges.

Table 3 reports on the percentage of siblings who earned a Bachelor's degree, after blocking on the norm reference group (i.e. separating those who did and did

[2]For more detail on the NLSY, see Herrnstein & Murray (1993).

not earn a Bachelor's degree). It is clear that cognitive ability is predominantly related to securing a BA, but there are other ways to analyse these data to highlight the relationship between cognitive ability and this educational outcome.

For example, of the sibling pairs in Table 3, 228 pairs had different outcomes (i.e. one got a BA, while the other did not); interestingly, of these 228 discordant sibling pairs, the BA went to the higher IQ sibling 88% of the time. This indicates the profound advantage that higher cognitive abilities bestow.

Eliminating social deprivation and poverty

Subsequently, Murray (1998) created a 'utopian sample' from NLSY for informing policy researchers aiming to evaluate social interventions. He deleted all subjects who were either raised in the bottom quartile of income or in single-parent households. This gives researchers a purchase on the eventualities of eliminating poverty and single-parent homes on various social outcomes. Table 4 presents data from the Full NLSY sample with the Utopian sample. With respect to outcome differences across contrasts involving educational attainment, employment and earned income, and the childbearing characteristics of women, the Full NLSY and Utopian samples are not all that different. The largest difference is between the Full NLSY and Utopian sample for the Bright category (50% and 57%, respectively) for obtaining a BA.[3]

Collectively, when Murray's findings are teamed with recent advances from biometrically informative twin and adoption studies, the causal significance of SES, for most environments found in industrialized societies, becomes attenuated further. It also motivates the necessity of more general scientific tools.

Other tools for future research

In an early differential psychology text, Ellis (1928) introduced psychologists to the Fallacy of the Neglected Aspect:

> The logicians point out that a cause of much incorrect thinking is what is known as the fallacy of the neglected aspect. Early students of certain diseases

[3]As important as Murray's (1998) study is, however, a component is missing: it would be fascinating to reverse this analysis. It would be informative to select groups of biologically unrelated subjects at comparable ability levels, who are raised in different homes, which systematically vary in SES. If these subjects were studied longitudinally, they would complement the power of Murray's (1998) design, which controls for SES, by controlling for ability analogously. Both analyses would independently converge on a precise estimate of the relative influence of reared-in SES and general intelligence on various outcomes. Naturally, the relative influence of reared-in SES and general intelligence on remote outcomes may change over the life span, which is why Murray's (1998) decision to study these relationships longitudinally is so compelling. Other useful methods for establishing controls between *g* and SES are found in Lubinski & Humphreys (1992) for physical health, and Waller (1971) for social mobility.

TABLE 4 Utopian sample comparisons

	Cognitive class									
	Very dull (< 10th %ile)		*Dull (10th–24th)*		*Normal (25th–74th)*		*Bright (75th–89th)*		*Very bright (≥ 90th %ile)*	
Sample	*Utopian*	*Full NLSY*	*Utopian*	*Full NLSY*	*Utopian*	*Full NLSY*	*Utopian*	*Full NLSY*	*Utopian*	*Full NLSY*
Educational attainment										
$\overline{X}$ years of education	11.4	10.9	12.3	11.9	13.4	13.2	15.2	15.0	16.5	16.5
% obtaining BA	1	1	4	3	19	16	57	50	80	77
Employment & earned income										
$\overline{X}$ number of weeks worked	36	31	39	37	43	42	45	45	46	45
Mdn earned income (US$)	11000	7500	16000	13000	23000	21000	27000	27000	38000	36000
% with spouse with earned income	30	27	38	39	53	54	61	59	58	58
Mdn earned family income	17000	12000	25000	23400	37750	37000	47200	45000	53700	53000
Female childbearing characteristics										
Fertility to date	2.1	2.3	1.7	1.9	1.4	1.6	1.3	1.4	1.0	1.0
Mother's $\overline{X}$ age at first birth	24.4	22.8	24.5	23.7	26.0	25.2	27.4	27.1	29.0	28.5
% children born out of wedlock	49	50	33	32	14	14	6	6	3	5

From Murray (1998).

> considered them to be due to hot weather or excessive rain—neglecting the activities of the fly or the mosquito in spreading the bacteria. Neglecting aspects of problems often hides variable agencies that must be understood before the problem can be solved. Experiment has often been the only way out of this difficulty, and where experiment is not possible the problem may remain unsolved. (p 8)

Subsequently, Carnap (1950) formalized this fallacy as the *Total Evidence Rule*. The rule maintains that, when evaluating the plausibility of a particular hypothesis, or the verisimilitude of a theory, it is imperative to take into account all of the relevant information (Lubinski & Humphreys 1997). As commonsensical as this seems, it frequently is not done.

For example, investigators readily assume that the covariation between parent and child's general intelligence, verbal intelligence, personality, or vocational interests is due to the kinds of environmental stimulation parents provide (cf. Thompson's [1995] review of Hart & Risley [1995]). Yet, biometrically informed analyses reveal that covariation among the attributes studied by differential psychologists (abilities, interests, and personality) approach zero as adulthood is reached among biologically unrelated siblings reared together. As unrelated individuals who were reared together grow older, they appear to 'grow up and grow apart' (McCartney et al 1990), with respect to conventional dimensions of individual differences. It appears that an inconspicuous cause, namely shared genetic make-up, is responsible for the phenotypic covariation between biologically related parents and children. Parents do, indeed, have an influence on their children with respect to major dimensions of individual differences; however, this influence is transmitted through a different mechanism than many initially presupposed. This is also supported by a variety of kinship correlates, such as the intriguing finding that, on 'environmental measures' (e.g. HOME, cf. Plomin & Bergeman 1991), identical twins reared apart assess their reared-in home environments as being as similar as fraternal twins reared together do (Scarr 1996).

Now, to be sure, this is not to say that abusive environments are not detrimental to optimal development. Of course they can be, and they frequently are. What these studies do speak to, however, is that, in the large majority of environments, many families are functionally equivalent in terms of fostering the development of major sources of individual differences studied by differential psychologists (McGue & Bouchard 1998). Yet, much of psychology appears to *neglect* these findings when launching causal inferences about exogenous influences based on correlations between biologically related individuals. In many social science domains, for example, not considering g and incorporating general ability measures in empirical studies amounts to errors of omission, and misspecified or

underdetermined causal models, which constitute Fallacies of Neglected Aspects and violations of Total Evidence (Lubinski & Humphreys 1997). But furthermore, at times, scientists who 'know better' appear to conscientiously avoid placing their favourite constructs in competition with general ability measures (Coleman 1990, Humphreys 1991). In such instances (cf. Scarr 1998, p 231), should scientific malpractice perhaps be used to characterize their behaviour?

In conclusion, the construct of general intelligence is a pivotal feature of the human condition. The best evidence suggests that the observed covariation between measures of g and abstract learning and work performance, as well as social and vocational outcomes, is primarily causal rather than merely associative. Future work in evolutionary psychology and the social sciences more generally would do well by exploiting this construct more routinely. To the extent that g is ignored in biosocial sciences aiming to better understand broad human behaviour patterns, the comprehensiveness of purported frameworks are virtually guaranteed to fall short. This is especially likely to be found in the most general and familiar aspects of the human condition, such as mating and resource acquisition, which, according to Freud, and many other acute observers of human biosocial phenomena, constitute the two most central aspects of life.

Acknowledgements

Earlier versions of this article profited from suggestions by Camilla P. Benbow, Daniel L. Shea and Rose Mary Webb. The article was supported by an anonymous funding agency, the Kennedy Center at the College of Education and Human Development, Vanderbilt University, and a Templeton Award for Positive Psychology.

References

Achter JA, Lubinski D, Benbow CP 1999 Assessing vocational preferences among gifted adolescents adds incremental validity to abilities: a discriminant analysis of educational outcomes over a 10-year interval. J Educ Psychol 91:777–786

Benbow CP, Stanley JC 1996 Inequity in equity: how 'equity' can lead to inequity for high-potential students. Psychol Pub Policy Law 2:249–292

Campbell DT, Fiske DW 1959 Convergent and discriminant validation by the multitrait-multimethod matrix. Psychol Bull 93:81–105

Campbell JP 1990 The role of theory in industrial and organizational psychology. In: Dunnette MD, Hough LM (eds) Handbook of industrial and organizational psychology, 2nd edn, vol 1. Consulting Psychology Press, Palo Alto, CA, p 39–74

Carnap R 1950 Logical foundations of probability. University of Chicago Press, Chicago, IL

Carroll JB 1993 Human cognitive abilities: a survey of the factor-analytic literature. Cambridge University Press, Cambridge

Carroll JB 1997 Psychometrics, intelligence, and public perception. Intelligence 24:25–52

Cattell JM 1890 Mental tests and measurements. Mind 15:373–380
Cattell RB 1950 Personality: a systematic theoretical and factual study. McGraw-Hill, New York
Coleman JS 1990 The Sidney Hook Memorial Award address: on the self-suppression of academic freedom. Acad Quest 4:17–22
Cronbach LJ 1970 Essentials of psychological testing, 3rd edn. Harper and Row, New York
Cronbach LJ, Meehl PE 1955 Construct validity in psychological tests. Psychol Bull 52:281–302
Ellis RS 1928 The psychology of individual differences. Appleton, New York
Galton F 1890 Remarks. Mind 15:380–381
Galton F 1961 Classification of men according to their natural gifts [Originally published in 1869]. In: Jenkins JJ, Paterson DG (eds) Studies in individual differences. Appleton-Century-Crofts, New York, p 1–16
Gottfredson LS 1997 Intelligence and social policy. Intelligence 24:1–320
Hart B, Risley TR 1995 Meaningful differences in the everyday experience of young American children. Brookes, Baltimore
Hathaway SR, McKinley JC 1940 The Minnesota multiphasic personality inventory manual. Psychological Corporation, New York
Herrnstein RJ, Murray C 1994 The bell curve: Intelligence and class structure in American life. Free Press, New York
Holahan CK, Sears RR 1995 The gifted group in later maturity. Stanford University Press, Stanford, CA
Humm D 1946 Validation by remote criteria. J Appl Psych 30:333–339
Humphreys LG 1985 General intelligence: an integration of factor, test, and simplex theory. In: Wolman BB (ed) Handbook of intelligence. Wiley, New York, p 201–224
Humphreys LG 1991 Limited vision in the social sciences. Am J Psychol 104:333–353
Humphreys LG, Lubinski D, Yao G 1993 Utility of predicting group membership and the role of spatial visualization in becoming an engineer, physical scientist, or artist. J Appl Psychol 78:250–261
Hunt E 1995 Will we be smart enough? A cognitive analysis of the coming workforce. Russell Sage Foundation, New York
Jensen AR 1998 The *g* factor: the science of mental ability. Praeger, Westport, CT
Jensen AR 2000 The *g* factor: psychometrics and biology. In: The nature of intelligence. Wiley, Chichester (Novartis Found Symp 233) p 37–46
Jensen AR 2000 Charles Spearman: the discoverer of *g*. In: Kimble GA, Westheimer M (eds) Portraits of pioneers in psychology. Vol IV. American Psychological Association, Washington DC, and Erlbaum, NJ, in press
Lubinski D 1996 Applied individual differences research and its quantitative methods. Psychol Public Policy Law 2:187–203
Lubinski D 2000 Scientific and social significance of assessing individual differences in human behavior. 'Sinking shafts at a few critical points'. Annu Rev Psychol 51:405–444
Lubinski D, Dawis RV 1992 Aptitudes, skills, and proficiencies. In: Dunnette MD, Hough LM (eds) Handbook of industrial and organizational psychology, 2nd edn, vol 3. Consulting Psychology Press, Palo Alto, CA, p 1–59
Lubinski D, Humphreys LG 1992 Some bodily and medical correlates of mathematical giftedness and commensurate levels of socioeconomic status. Intelligence 16:99–115
Lubinski D, Humphreys LG 1997 Incorporating general intelligence into epidemiology and the social sciences. Intelligence 24:159–201
McCartney K, Harris MJ, Bernieri F 1990 Growing up and growing apart: a developmental meta-analysis of twin studies. Psychol Bull 107:226–237
McGue M, Bouchard TJ 1998 Genetic and environmental influences on human behavioral differences. Annu Rev Neurosci 21:1–24

Meehl PE 1990 Appraising and amending theories: the strategy of Lakatosian defense and two principles that warrant it. Psychol Inquiry 1:108–141
Meehl PE 1998 The power of quantitative thinking. American Psychological Society, New York
Messick S 1992 Multiple intelligences or multilevel intelligence? Selective emphasis on distinctive properties of hierarchy: on Gardner's Frames of Mind and Sternberg's Beyond IQ in the context of theory and research on the structure of human abilities. Psychol Inquiry 3:365–384
Murray C 1998 Income, inequality, and IQ. American Enterprise Institute, Washington, DC
Neisser U, Boodoo G, Bouchard TJ Jr et al 1996 Intelligence: knowns and unknowns. Am Psychol 51:77–101
Plomin R, Bergeman CS 1991 The nature of nuture: genetic influence on 'environmental' measures. Behav Brain Sci 14:373–427
Scarr S 1996 How people make their own environments: implications for parents and policy makers. Psychol Public Policy Law 2:204–228
Scarr S 1998 On Arthur Jensen's integrity. Intelligence 26:227–232
Schmidt FL, Hunter JE 1998 The validity and utility of selection methods in personnel psychology: practical and theoretical implications of 85 years of research findings. Psychol Bull 124:262–274
Snow RE 1989 Aptitude treatment interaction as a framework for research on individual differences in learning. In: Ackerman PE, Sternberg RJ, Glasser RG (eds) Learning and individual differences. WH Freeman, New York, p 13–59
Snyderman M, Rothman S 1987 Survey of expert opinion on intelligence and aptitude testing. Am Psychol 42:137–144
Spearman C 1904 'General intelligence,' objectively determined and measured. Am J Psychol 15:201–292
Spearman C 1927 The abilities of man: their nature and measurement. Macmillan, New York
Sternberg RJ, Detterman DK 1986 What is intelligence? Contemporary viewpoints on its nature and definition. Ablex, Norwood, NJ
Terman LM 1925–1959 Genetic studies of genius (I–V). Stanford University Press, Stanford
Thompson T 1995 Children have more need of models than critics: early language experience and brain development. J Early Interven 19:264–272
Thorndike RL 1994 *g* (Editorial). Intelligence 19:145–155
Waller JH 1971 Achievement and social mobility: relationships among IQ score, education, and occupation in two generations. Soc Biol 18:252–259
Weschler D 1981 WAIS-R manual: Wechsler adult intelligence scale revised. Psychological Corporation, New York
Wilson EO 1998 Consilience: the unity of knowledge. Knopf, New York
Wonderlic Personnel Test 1992 Wonderlic personnel test and scholastic level exam: user's manual. Wonderlic, Libertyville, IL

DISCUSSION

Harnad: I want to make a comment on causality, which will no doubt be a recurring theme in these discussions. Causality is, in my opinion, the key scientific question in connection with intelligence and IQ. Suppose we had a battery of athletic skill measures: swimming, javelin throwing, long-distance running and so on. Imagine that we then factor-analysed them and came up with a factor, which turned out to be a *g* factor, involving a basis such as the ratio of slow twitch to fast twitch muscle fibres. The question is, where would that have advanced us if we didn't know the causal mechanisms of muscle use, and movement and motor patterns? The counterparts of this in cognition are the cognitive mechanisms underlying cognitive capacity. To what extent does the correlation pattern that *g* seems to exemplify help us understand the real causal mechanisms underlying cognitive capacities? I'm inclined to say that *g* has provided approximately zero benefit in helping us with cognitive modelling.

Lubinski: One thing I wanted to emphasize in my paper is that there is a lot more to differential psychology than *g*. There are group factors, interests and personality, and you can get at all of them through the Spearman–Brown formula (Lubinski 1996, Lubinski & Humphreys 1997). In terms of how *g* can help us understand low-order mechanisms, I think the jury is still out (but see Lubinski 2000). Science moves slowly sometimes. I think it can help us find and isolate different populations, look for genetic markers and look for individual differences in the CNS that can provide clues to the underlying structure. To be clear, the *g* construct is not a 'thing'; it's an abstraction like horsepower. There are different components to horsepower, such as carburettors and cylinders, but still there's a general property. The overall functioning of this general property can be increased by tinkering with the components individually, tinkering with the whole system, or tinkering with fuel: there are a variety of different variables underlying 'horsepower' as there undoubtedly are with *g*. It helps us to know where to look, just as Skinner's principle of reinforcement helped us identify different areas of the brain for positive reinforcement centres and so on.

Detterman: To a large extent I agree with Stephen Harnad's comment: I think he was exactly right until he said that *g* has made zero contribution to cognitive modelling. We could think of *g* as being the gold standard that we need to compare cognitive models to: if we look at it that way then it is the cognitive models that have let us down, not *g*.

Harnad: Wouldn't you say that the gold standard of cognitive models for cognitive capacity is cognitive capacity itself?

Detterman: As related to *g*, since *g* is known to correlate with all these social variables.

Maynard Smith: I'm already baffled. I am not clear about what kind of claim is being made for *g*. Let's take the analogy of athletic performance that Stephen Harnad drew, which I think is quite a good one. One's athletic ability will be influenced by factors such as heart volume, leg length, muscle development and motivation: if you were to measure each of these factors independently, you would be in quite a good position to predict how people might perform in athletic contests. You would probably find high heritabilities in these features. But there isn't a thing called 'athletic ability': it is just a performance category. If you were then to analyse this you wouldn't necessarily expect to find that everybody with a high athletic ability had one set of traits: you would expect to find these different components, such as heart, muscle and nerves, contributing to it. You wouldn't find one thing; you would find many contributors, just as you would if you analysed the horsepower of an engine. Are we required to think that there exists a 'thing' which will have a specific neurological component? Because, if not, let's not talk about it as if it were a 'thing'. It is just a correlate that statistically is quite good at predicting one's performance. That is fine. There is no reason why it shouldn't have a high heritability. But somehow the discussion about it seems not to be in those terms: it is discussed as if there were an object like a heart in there that we could identify.

Jensen: No one who has worked in this field has ever thought of *g* as an entity or thing. Spearman, who discovered *g*, actually said the very same thing that you're saying now, and Cyril Burt and Hans Eysenk said that also: just about everyone who has worked in this field has not been confused on that point.

Maynard Smith: From reading the abstracts of the papers in this meeting and listening to you, this is not the impression I got. I'm delighted to hear workers in the field do not think of *g* as a thing.

Detterman: There are some people who believe there may be a single underlying variable that explains *g*.

Deary: Can I comment on what you, David Lubinski, were saying about the McKeen Cattell paper from 1890 and Galton's (1890) response, because I think therein lies the reason for the situation we're in today. The phenotypes that ability tests seem to get at have proved relatively easy to measure and mental tests are predictively quite successful. The problem lies in understanding the causes of mental ability differences. The roots of this problem were already present in these early papers. McKeen Cattell didn't present any correlations, but simply described some tests off the top of his head. These tests were much less good than those he had developed under Wundt (Deary 1994). Galton's comments on the McKeen Cattell suggestions were twofold. One is that there was no theory driving the tests, which worried him, and second was the point that psychologists should strive to sink a few critically placed shafts into the brain. What he meant was just what has been raised here: we should try to identify a few key processing parameters that are actually the partial bases of mental ability differences. Galton was lamenting then something that we still don't have. I think that little interchange in 1890 is very much the same sort of thing we are going to see replayed here at this meeting. We can describe ability phenotypes and use test scores in prediction but we still don't know the cognitive or biological bases of human mental ability differences.

Flynn: David Lubinski, I picked up your comments on Murray. The sociologists have had a serious setback. Murray's utopian experiment simulated improving people in terms of SES plus other factors, and it looks like if we were to improve those environmental factors, people would not benefit as much as we would hope.

However, you could say that Murray's own thesis is subject to an even more devastating disappointment: we actually have made massive IQ gains over time, without any of the progress that he would anticipate from that. That is, during the period in which solo-mother homes have risen and crime has risen, there have been massive IQ gains. One could say, therefore, that his position has been subjected not merely to a simulated refutation, but to a devastating real-world refutation. There are ways out of that, of course: you can claim that the massive IQ gains over time are really just a result of test sophistication — that they are artefacts. There is strong evidence against that; I won't go into that now, but I will offer an alternative interpretation. It could well be that IQ gives enormous advantages within a group in which there is intragroup competition, particularly for the reason you've given — and that this results in a matching between genes for IQ and quality of environment. An intelligent kid with good genes makes a lot more out of a library than an unintelligent kid. Therefore, his genes get the credit for potent environmental factors: the latter are hidden behind a genetic mask in a competitive intragroup situation. When you look at IQ gains over time — there of course we think that there is effectively genetic parity between the generations — you see the potent force of the environment revealed in all of its naked power. There is no longer a covariance between genes and environment that hides the explosive force of environment. In sum, it could be argued against Murray that his hopes that raising IQ will bring progress are suspect. Let me give a simple analogy. It may be that self-seekingness is very profitable in competition at a particular place and time, in terms of getting you higher on the ladder. It's not at all clear that were the human race to become more self-seeking, that there would be group progress over time — that is, from one generation to another. In other words, a trait that pays enormous dividends within a competitive situation at a given place and time, doesn't necessarily solve social problems over time. At present, it looks like increases in IQ are totally feeble in this regard—perhaps even more feeble than SES.

Brody: Examine Table 3. If you look at the performance of normal siblings who have graduated college and those who have not, you see for the same range of IQ differences among siblings enormous differences in the probability of higher education. For example, the dull siblings of those siblings who did not graduate college have a 1% probability of graduating college, and the dull siblings of those normal siblings who did graduate college have an 18% probability of graduating college. This raises an interesting issue. As determinative as IQ is, there surely is a great deal of variation among children with equivalent IQ, and this must be some kind of a family influence.

Suddendorf: You ended your talk by stating that social sciences might commit the 'fallacy of neglected aspect'. How would you respond to the reverse criticism, raised by people like Robert Sternberg (1999), that psychometricians might be committing a 'confirmation bias' by restricting the range of participants, tasks and the situational context in which testing takes place?

Lubinski: I would say the opposite. Differential psychologists tend to be very sensitive to restriction of range and try hard to study the full range of human talent and, if not, they will correct for that. Sternberg studies Yale undergraduates.

Deary: That wasn't the point being made: the point was about the restricted range of tests that we use, not the attenuation of the sample variance. Sternberg (1999) makes the point that by looking at what he calls 'analytic reasoning' we are restricting the range of talents and abilities we examine. We don't look at what he would call 'creative and practical reasoning'.

Lubinski: That is the general factor that factors into a broader array of criteria that we're interested in predicting when we talk about the concept of intelligence. To my knowledge Sternberg's creative and practical tests haven't added any incremental validity to those predictions in the full range of talent in any study that I know.

Suddendorf: Isn't the point that he's making that there is no correlation between practical intelligence the way he has tried to measure it and *g*?

Brody: I'm going to address that in my paper. I will present data indicating that the correlation between creative intelligence as Sternberg assesses it and *g* is very high. Indeed, when Sternberg actually tries to measure something — not when he is being conceptual — the data show that *g* is a good predictor of the three kinds of intelligence that Sternberg postulates. I know of no data that show that Sternberg is able to measure components of intelligence that are independent of *g*.

Whiten: In relation to the worry raised by John Maynard Smith about whether *g* is being treated as a 'thing', I'm a bit concerned about your reference to 'construct validation'. Deriving a construct *g* as a statistical property describing the relationships amongst a battery of scores on tests thought appropriate for measuring intelligence is fine, but an effort to *validate* the construct could easily lead to reifying it as a natural 'thing' that equates with intelligence. The results of this could perhaps be presented as achieving a better definition of intelligence than the one existing before this research programme was undertaken — intelligence might be claimed to be better understood and measured. But both the internal correlations generating *g*, and the external, 'validating' correlations with external factors, obviously depend on what test results are used in the first place. If some important aspect of intelligence is not tested to begin with, one might still run through this whole research process and claim at the end to have validated the construct *g* — as a factor of general intelligence — because it correlates with certain outcomes one would expect to be associated with intelligence. But both the tests and the outcomes might be based on similar — and similarly limited — notions of intelligence, making the process somewhat circular, misleadingly appearing to justify those notions of intelligence as 'the real thing'.

One potential omission that naturally concerns me is social intelligence, that is the subject of my later paper. You talked of Sternberg's work. In one paper he and his colleagues actually went to railway stations and other public places and asked people what they thought 'intelligence' was, and what it meant to them (Sternberg et al 1981). When the study went into a second phase and asked people to rate these various notions in relation to everyday intelligence and academic intelligence, a factor analysis did indeed throw up a first factor that looked a bit like *g*—general problem solving. But the second factor was social intelligence, like tact and managing your social affairs. Such social intelligence could be one omission in the process that has been used to extract *g* and validate it. It's probably very difficult, if not impossible, to measure in quick tests—especially just using pen and paper—that then have predictive power.

So, two questions. First, what exactly is the 'construct validation' validating? Is it some claim about intelligence, and if so what? Second, is it possible the whole process could misleadingly appear to reify a notion of intelligence that is not in fact as rich or broad as that people apply in everyday life?

Lubinski: These are good questions, and Cronbach & Meehl's (1955) treatment of construct validation is still a must read for people interested in these topics (but see also Meehl's recent refinements [personal communication, 1999]). Construct validity seeks to validate measures of a postulated attribute. 'Horsepower' is a postulated attribute, you can't 'see' horsepower, but you can construct indicators that co-vary with meaningful criteria that reflect our concept of horsepower and make it a conceptually powerful and useful concept. Just as horsepower is an abstract property of complex combustion engines, *g* is an abstract property of complex biological systems. Moreover, constructs can be postulated intuitively, or inferred from families of empirical relationships, either is fine. In the final analysis what's important is, does the measure behave in accordance with our theoretical expectations about the postulated attribute it purports to assess? Perhaps it would be helpful to review how Cronbach & Meehl (1955) illustrated how construct validation works for psychological measures.

When Cronbach & Meehl (1955) introduced the logic of the construct validation process, they exemplified the process by systematically compiling a heterogeneous collection of empirical phenomena all related to the *psychopathic deviate* ('*Pd*') scale of the Minnesota Multiphasic Personality Inventory (MMPI). How, they asked, could a scale initially developed to isolate criminals and delinquents from the general population, also evince the following network of empirical relationships: elevated scores for Broadway actors, high school drop-outs, deer hunters who accidentally shoot people, police officers and nurses who were rated by their supervisors as not especially afraid of psychotic patients? (Note this was before wide use of psychoactive drugs, and patients commonly experienced florid psychotic episodes.) *Pd* also correlated negatively with peer ratings of trustworthiness. What possible mechanism could possibly underlie this family of empirical relationships? They reasoned that the communality (abstract property) cutting across all of these findings involved low anxiety.

Two years later, Lykken (1957) published support for a hypothesis that appears to have withstood the test of time. What these groups have in common is that, relative to the norm, they are fearless or in possession of a 'low anxiety IQ'. He tested this idea, using a Pavlovian conditioning paradigm, and showed that, as a group, particularly hardened criminals (high *Pd* scorers), when contrasted with the general prison population, were 'retarded' when it came to developing conditioned responses to neutral stimuli paired with an unconditioned aversive stimulus, shock. Replicated now in a number of laboratories, a differential sensitivity to the development of conditioned responses to aversive stimuli seems to be a robust parameter of individual differences. This fits with all of the aforementioned empirical findings. (It also fits with the idea that low anxiety can be an asset or a liability, depending on the constellation of other personal attributes that a person possesses, and the purview one is operating in.) Hence, here, construct validity has clarified and helped us to better understand this measure, and to generate valid inferences (about the internal make up, for example, of spies, paratroopers, politicians, deep sea divers, stunt men, astronauts, etc., as well as a host of other phenomena) about events, people, and outcomes that have not been studied empirically.

Like the *Pd* scale of the MMPI, Binet's initial test was designed to ('simply') isolate a particular group of people ('educable' children) from the general population. But the validity of this measure soon generalized to school performance, amount of education, work performance and occupational groupings, and a variety of other domains involving abstract learning, information processing and responding to novel abstract problems in effective ways (Gottfredson 1997). What seems to be operating here is rate of learning abstract material, which is what conventional *g* measures assess. (This is why leaders in a variety of information-dense occupations rich with novelty manifest high levels of *g*. This also is why I quoted Galton's earlier on *general* ability.) But, to be clear, there is more to the intellectual repertoire than this. For example, the importance of mathematical, spatial and verbal abilities has actually been documented in some of my own work and that of others, which underscores why it is important to assess abilities beyond *g* (Achter et al 1999, Humphreys et al 1993, Lubinski & Dawis 1992); all differential psychologists agree that there is more to cognitive abilities than *g*. (I do not know of any exceptions.)

Your mentioning of social intelligence is, of course, not new; a number of early differential psychologists discussed social intelligence, practical intelligence and abstract intelligence (essentially *g*) early in the 20th century. People interested in the multitude of psychometric measure that have been developed to get at these (and other) unique aspects of cognitive functioning should consult Messick's (1992) excellent review (and see Lubinski & Benbow 1995). But the problem with social intelligence is that no one appears to have developed a valid measure that adds incremental validity to measures of *g* or verbal ability; nothing new has been developed (to my knowledge) over and above what conventional measures of *g* and verbal ability already give us. Although many measures have been proposed (Messick 1992), and the concept of social intelligence is clearly attractive and intuitively appealing, we have not yet been able to derive an assessment procedure to measure individual differences in this purported attribute.

Actually, this happens all of the time in psychology, a measure is proposed that sounds appealing and 'validated' without ever considering whether it adds anything to what we already have. A well-known measure of moral reasoning, for example, generated hundreds of studies and at least three books. But it was never evaluated in competition with general and verbal ability measures. While the Defining Issues Test (DIT) was correlated with conventional ability measures, it was never evaluated for its unique predictive properties (incremental validity) in the context of relevant external criteria, over and above conventional measures of *g* and verbal ability. When the appropriate analysis was done (Sanders et al 1995), all of the DIT's validity was absorbed by verbal ability and, moreover, verbal ability typically accounted for a great deal more criterion variance. Yet, three decades of research with this instrument never involved this simple analysis, and research continues to appear as if the Sanders et al (1995) study didn't exist. (Contemporary work on 'emotional intelligence' is encountering similar problems [Davies et al 1998].) It is fine to theorize about new constructs and to build new assessment tools but, to make a scientific advance, innovative procedures need to provide us with something new.

Finally, it is not scientifically problematic when postulated inferences about constructs under analysis change through empirical research (construct validation). How constructs are conceptualized typically changes as a result of empirical research; indeed, if this did not happen, there would be little need for empirical research. Measurement operations also evolve as empirical evidence accrues. Developing scientific tools is an ongoing process, and material changes in theoretical concepts and their measures are always to some extent evolving (e.g. see Tellegen 1985). But this is to be expected, for example: with respect to the measures discussed here, modern research has indicated that Lykken's (1957) initially rough idea (low 'anxiety IQ') has been refined (and split) into two components: physical anxiety and social anxiety (Lykken & Katzenmeyer 1973). Similarily, Binet's initially rough concept of mental age is essentially the core, dominant-dimension in Carroll's (1993) hierarchy of cognitive abilities (Fig. 3). These refinements are important advances in the field of differential psychology, and they enable us to speak more precisely about anxiety-proneness (and the components of anxiety-provoking situations) and cognitive abilities (and different kinds of intellectual demands).

Hinde: One has to be awfully careful to make clear what we mean by adaptation. In my view, we must not talk about social success as being an adaptation: that will lead us right up the wrong path.

The dependent variables used in these studies are variables that you would expect to show this pattern in an intensely competitive society such as the USA. We should bear in mind that it is at least possible that different sorts of results would be obtained in a collectivist society such as East Asian societies. These variables could be rather culture-specific. This issue impacts on what James Flynn raised about changing the cultural environment over time, and even the question of social intelligence. Do you agree that many of these data may be very much culture bound? The USA is a very peculiar society.

Lubinski: It's even more clear-cut than that I think. Anyone who just speaks one language is culture bound to being tested using that medium, if you're going to use a language-based vehicle. This is why a lot of cross-cultural research uses things like the Raven matrices, which are just pictorial.

Detterman: There is strong evidence that *g* applies cross-culturally.

Hinde: But are the measures that you are taking of social success as consequences of the economic success comparable with similar measures in other societies?

Detterman: There have been studies in Warsaw, Poland, which was destroyed after World War II and reconstituted under a Communist regime. People were more-or-less randomly assigned to neighbourhoods. They looked at academic achievement in relation to IQ, and found the same relationship as was found in democratic societies (Firkowska et al 1978).

Hinde: Supposing the test was done with a hunter–gatherer society, where social success depended in part on hunting success: would you get the same answer?

Detterman: Yes, you would get the same answer.

References

Achter JA, Lubinski D, Benbow CP, Eftekhari-Sanjani H 1999 Assessing vocational preferences among gifted adolescents adds incremental validity to abilities: a discriminant analysis of educational outcomes over a 10-year interval. J Educ Psychol 91:777–786

Carroll JB 1993 Human cognitive abilities: a survey of the factor-analytic literature. Cambridge University Press, Cambridge

Cronbach LJ, Meehl PE 1955 Construct validity in psychological tests. Psych Bull 52:281–302

Davies M, Stankov L, Roberts RD 1998 Emotional intelligence: In search of an illusive construct. J Pers Soc Psych 75:989–1015

Deary IJ 1994 Sensory discrimination and intelligence: postmortem or resurrection? Am J Psychol 107:95–115

Firkowska A, Ostrowska A, Sokolowska M, Stein ZA, Susser M, Wald I 1978 Cognitive development and social policy: the contribution of parental occupation and education to mental performance in 11 year olds in Warsaw. Science 200:1357–1362

Galton F 1890 Remarks on 'Mental tests and measurements' by J. McK. Cattell. Mind 15:380–381

Gottfredson LS 1997 Intelligence and social policy. Intelligence 24:1–320

Humphreys LG, Lubinski D, Yao G 1993 Utility of predicting group membership: Exemplified by the role of spatial visualization for becoming an engineer, physical scientist, or artist. J Appl Psych 78:250–261

Lubinski D 1996 Applied individual differences research and its quantitative methods. Psychol Public Policy Law 2:187–203

Lubinski D 2000 Assessing individual differences in human behavior: 'Sinking shafts at a few critical points'. Annu Rev Psychol 51:405–444

Lubinski D, Benbow CP 1995 An opportunity for empiricism: Review of Howard Gardner's *Multiple intelligences: the theory in practice*. Contemp Psych 40:935–938

Lubinski D, Dawis RV 1992 Aptitudes, skills, and proficiencies. In: Dunnette MD, Hough LM (eds) The handbook of industrial/organizational psychology, 2nd ed. Consulting Psychologists Press, Palo Alto, p 1–59

Lubinski D, Humphreys LG 1997 Incorporating general intelligence into epidemiology and the social sciences. Intelligence 24:159–201

Lykken DT 1957 A study of anxiety in the sociopathic personality. J Abnorm Soc Psych 55:6–10

Lykken DT, Katzenmeyer C 1973 Manual for the activity preference questionnaire (APQ). In: Psychiatric research reports. University of Minnesota Press, Minneapolis, MN

McKeen Cattell J 1890 Mental tests and measurements. Mind 15:373–380

Messick S 1992 Multiple intelligences or multilevel intelligence? Selective emphasis on distinctive properties of hierarchy: On Gardner's *Frames of mind* and Sternberg's *Beyond IQ* in the context of theory and research on the structure of human abilities. Psych Inquir 3:365–384

Sanders CE, Lubinski D, Benbow CP 1995 Does the defining issues test measure psychological phenomena distinct from verbal ability? An examination of Lykken's query. J Pers Soc Psych 69:498–504

Sternberg RJ 1999 Successful intelligence: finding a balance. Trends Cognit Sci 3:436–442

Sternberg RJ, Conway BE, Ketron JL, Bernstein M 1981 People's conceptions of intelligence. J Pers Soc Psychol 41:37–55

Tellegen A 1985 Structures of mood and personality and their relevance to assessing anxiety, with an emphasis on self-report. In: Tuma AH, Maser JD (eds) Anxiety and the anxiety disorders. Lawrence Erlbaum, Hillsdale, NJ, p 681–706