two from the *School and College Ability Tests* (Verbal and Quantitative).

A two-factor solution was extracted, again using a procedure that was probably suboptimal. In this case, both factors were rather general, in that the loadings were uniformly positive and fairly high. The first factor was the more verbal and the second the more quantitative, a characterization that held over all testing occasions, and the differentiation into verbal and non-verbal increased somewhat across the grades. In short, the solution was reasonably coherent, although not consistent with a single g factor.

To my knowledge, this approach has not been explored further; not knowing what might be wrong with it, I pursued other areas of psychology. Yet the approach presents another way of looking at the problem of rotation, and also addresses the problem of how factors change over time. In this second respect, it seems to me to bear more directly on issues concerning innate versus environmental influences in intelligence than does any consideration of black–white differences.

The practical significance of black–white differences in intelligence

Linda S. Gottfredson

Department of Educational Studies, College of Education, University of Delaware, Newark, Del. 19716; Center for Social Organization of Schools, Johns Hopkins University, Baltimore, Md. 21218

Black–white differences in mental test performance often differ from one test to another, but Jensen (1985) has shown that the magnitude of mean black–white differences in performance is highly correlated with the g loadings of the tests (see also Gordon 1985). Jensen has thereby provided additional persuasive evidence that the black–white difference in tested intelligence is in fact a difference in g rather than in mastery of content specific to each test.

Several commentators on Jensen's target article referred explicitly to the educational, occupational, and other social consequences of individual or group differences in g (Carlson 1985; Cattell 1985; Eysenck 1985; Rushton 1985; Sternberg 1985), as did Jensen himself. The controversy over Jensen's work on black–white differences in mental test performance has probably raged so fiercely because most scientists and lay people alike have assumed that intelligence is of great practical importance in human affairs. Now, however, in the face of indications that a large and persistent average black–white difference in g may indeed exist, more people are beginning to dispute the assumption that differences in intelligence are, or need be, of much practical significance, especially outside school settings (Gottfredson 1986). Commentators Rabbitt (1985) and Stanovich (1985) reflect this growing trend, with their argument that better training of specific skills negates the importance of underlying differences in g. Sternberg's reaction of "So what?" to Jensen's target article epitomizes the reaction of many people who have accepted that a nontrivial black–white difference in g does in fact exist.

Despite this undercurrent of opinion about the practical significance of differences in g, there has been almost no explicit discussion of those consequences and how we might deal with them most constructively. (For exceptions, see Hunter et al. [1984] on worker productivity and Gottfredson [1985; 1986] on social policy with regard to education and work.) In this commentary, I review recent evidence that is relevant to this issue but that is not well known outside industrial psychology.

Recent meta-analyses of decades of personnel selection research (Hawk 1970; Hunter 1983a; 1983b; Hunter & Hunter 1984; Hunter et al. 1984; McDaniel 1986; Schmidt & Hunter 1981; Thorndike 1985) have shown the following: (a) Intel-

ligence tests predict performance in training and on the job in all kinds of work, (b) job performance is more correlated with test performance in higher level (more complex) jobs than in lower level ones, (c) the relation of tested intelligence to job performance is linear, meaning that there is no threshold above which higher levels of intelligence are not associated with higher mean levels of job performance, (d) it is almost entirely the g factor in psychometric tests that accounts for their validity for predicting job performance, (e) the predictive validity of intelligence tests remains largely the same but that of experience fades among workers with higher mean levels of experience, (f) intelligence tests predict job performance even after controlling for differences in job knowledge, and (g) intelligence tests predict job performance equally well for blacks and whites, whether performance is measured objectively or subjectively.

Analyses of extensive job analysis data (Gottfredson 1984) have shown that intellectually more demanding jobs are not only rated as more generally desirable (prestigious) by workers but also as jobs in which good performance is especially critical to employers. They are also typically characterized by more arduous training demands, higher levels of decision making and data analysis, continual updating of job knowledge, and other presumably highly g-loaded tasks. Not surprisingly, large-scale studies from World War I through the 1970s have shown that occupations differ considerably in the mean intelligence levels of their incumbents – means correlating extremely highly with the prestige level of the occupations (Matarazzo 1972, Ch. 7; U.S. Department of Labor 1970, Table 9-2).

It is often suggested that differences in school or work performance, even those due to differences in intelligence, can be reduced or eliminated by lengthier or more effective education and training in the specific skills that schools and jobs require. This belief is implied, for example, by Stanovich's commentary (1985, p. 243), in which he refers to training for intelligence as a "tired debate" now outmoded by research showing that specific cognitive skills (e.g., reading comprehension) can be improved. Along similar lines, Rabbitt (1985, p. 239) suggests that predicting the time taken to learn unfamiliar tasks in strange social settings is "not very informative" and that knowing differences in asymptotic performance levels is more important.

The aforementioned research findings suggest, however, that differences in trainability may be important on the job, quite aside from the differential financial costs they impose in training programs per se. Specifically, jobs appear to remain g loaded despite prior training and despite years of experience. This is especially so for higher level jobs, perhaps because of their demands for continual learning, decision making, and other highly g-loaded activities. Jobs do not come in completely predictable bundles of standardized tasks for which complete training can be provided, any more than do most of life's demands – hence the pervasive impact of g on job performance (Gottfredson, in press).

Instead, the foregoing findings suggest that occupations themselves can be viewed as analogous to differentially g-loaded mental tests. The g loading of an occupation, considered most broadly, is a function of the predictive value of intelligence with respect to entry, good performance, and remaining in that occupation. This implies, as Cattell (1985) suggested in his commentary, that the proportion of blacks to whites will be smaller in the more intellectually demanding jobs.

A substantial correlation (−.5) has, in fact, been found between one measure of the overall intellectual complexity of work and the percentage of workers who are black, within either sex (Gottfredson 1984). However, this correlation fails to convey any real sense of the occupational significance of the black–white difference in g. What ratio of blacks to whites would be expected among physicians in a racially blind but appropriately g-loaded selection system: 1 to 20? 1 to 4? 1 to 2? How would these ratios be expected to differ for other occupations – say, teacher, electrician, or truck driver? And, what would the

Table 1(Gottfredson). *Actual black-white ratios of male employment in nine occupations versus employment ratios expected on the basis of black–white differences in intelligence*

z scores for IQ range from which most workers are recruited[a]	(1) Prestige level[b]	(2) Major IQ range of recruitment (in SB metric)[c]	(3) % of population above minimum IQ required[d] B	(4) % of population above minimum IQ required[d] W	(5) % of population in IQ range[d] B	(6) % of population in IQ range[d] W	(7) B/W ratio in range — range for blacks is .5 SD higher	(8) B/W ratio in range — same range	(9) range for blacks is .5 SD lower	(10) B/W ratio for % of male workers employed in each occupation[e] 1970[f]	(11) B/W ratio for % of male workers employed in each occupation[e] 1980[g]
0.8+		114+	1.1	23.0	1.1	23.0	.01	.05	.22		
Physician	88									.23	.30
Engineer	66									.12	.25
0.5 to 2.0		108–134	3.3	35.2	3.3	32.7	.02	.10	.35		
Secondary teacher	63									.59	.59
Real estate sales	48									.18	.23
−0.5 to 1.0		91–117	28.4	74.5	27.8	56.9	.20	.49	.87		
Fire fighter	41									.27	.65
Police officer	37									.69	.87
Electrician	44									.33	.50
−0.8 to 0.7		86–112	42.5	83.1	40.8	56.4	.36	.72	1.07		
Truck driver	29									1.59	1.48
Meat cutter	32									.98	.98

[a]Range extends roughly from minimum intelligence level (GATB aptitude G) required as determined in GATB studies of job performance (U.S. Department of Labor 1970, Table 9-1, job titles 131, 142, 155, 179, 246, 297, 356, 404, and 413) to 1.5 SDs above that minimum. C cutting points are not available for fire fighter and meat cutter, but $\bar{X}_G$ and SD_G are the same as for police officer and truck driver, respectively, so the former were grouped with the latter. A full sample SD of 1.5 above the minimum was chosen for the major range of recruitment because it generally represents within-occupation ranges of at least 2.0 SDs. The range for physicians and engineers was adjusted downward to be consistent with studies reviewed by Matarazzo (1972, Ch. 7) for physicians and other professionals.

[b]Source: Gottfredson and Brown (1978).

[c]IQ ranges in the 1937 Stanford-Binet metric were estimated by assuming that $\bar{X}_G = 100$, $SD_G = 20$, $\bar{X}_{IQ} = 100$, and $SD_{IQ} = 17$. The latter two parameters were estimated from data in footnote d and assuming that the total population is 88% white and 12% black.

[d]Calculated assuming that $\bar{X}_{IQ}$ and SD_{IQ} are 83.4 and 13.4 for blacks and 101.8 and 16.4 for whites. These parameters were calculated by transforming WISC data for a nationally representative sample of 12–17 year-olds (Hitchcock 1976, Tables 1 & 2) into the 1937 Stanford-Binet metric. Percentages of the population within any specific IQ range were then estimated from the cumulative normal distribution (Dixon & Massey 1969, Table A-4).

[e]Data are for the experienced civilian labor force and so include the experienced unemployed as well as the employed. Occupational categories for 1970 and 1980 are not perfectly comparable with each other or with the GATB job titles. More Spanish-origin persons were included in the white category in 1970 than 1980 (U.S. Bureau of the Census 1983, pp. 3–4).

[f]Calculated from data in U.S. Bureau of the Census (1973, Table 2) for occupational categories 006–023, 065, 144, 270, 430, 631, 633, 715, 961, and 964 (public police only, U.S. Bureau of the Census 1971).

[g]Calculated from data in U.S. Bureau of the Census (1983, Table 2) for occupational categories 044–059, 084, 157, 254, 416–418, 575, 686, and 804 (U.S. Bureau of the Census 1982).

Source: Gottfredson (1986).

consequences be for productivity and social harmony of implementing ratios that progressively exceed or fall short of those expected ratios? These are the sorts of questions that face judges, employers, and unions, but for which we have no answers.

Table 1 provides data pertinent to some of these questions by estimating what occupational patterns would look like if blacks and whites were both recruited to jobs from the same range of IQ that occupational incumbents appear to have been drawn from in the past. The first step in making these estimates was to determine the IQ ranges from which workers have most often been recruited to different occupations (column 2). Such recruitment ranges were estimated for nine diverse occupations

that use large numbers of males and for which there were intelligence data available for incumbents. The second step was to use nationally representative mental test data to determine the proportions of blacks and whites falling within each of those IQ recruitment ranges (columns 5–6). (Additional detail on procedures is provided in the notes to Table 1.) The third step was to compute the ratio of blacks to whites who would therefore be eligible for each occupation on the basis of intelligence alone, using the same IQ standards for both races (column 8). Black–white parity in employment is represented by a ratio of 1.0.

For purposes of comparison, ratios were calculated for recruitment ranges that are half a standard deviation higher for blacks than for whites (column 7) and also for ranges half a

standard deviation lower for blacks (column 9). Finally, black–white ratios for actual employment in 1970 and 1980 were calculated (columns 10–11).

As would be expected, smaller estimated black–white ratios are found in the more intellectually demanding jobs, whether or not blacks are recruited from the same IQ ranges as are whites. The ratios differ considerably depending on whether recruitment ranges for blacks are assumed to be higher than, the same as, or lower than those for whites; however, ratios differ even more by job level *within* each of those hypothetical recruitment ranges for blacks. In addition, expected black–white ratios are very far from parity in the more intellectually demanding occupations no matter which of the three recruitment ranges for blacks is considered, and ratios approach parity only among the lowest level jobs listed in Table 1. For example, even when recruitment standards are half a standard deviation *lower* for blacks, the ratios are only 1 to 5 for physican and engineer and 1 to 3 for secondary school teacher and real estate salesperson. Different assumptions about the distribution of intelligence in the black and white populations and about the intelligence requirements of occupations would produce somewhat different estimated black–white ratios for individual occupations, but the overall *pattern* of ratios would probably be the same under any set of reasonable assumptions: Racially blind worker selection can be expected to produce especially striking deviations from black–white parity in higher level jobs.

Data on *actual* employment ratios in 1970 and 1980 reveal the same inverse relation between black–white ratios and the intellectual demands of occupations. Surprisingly, however, actual employment ratios are most similar to estimated recruitment ratios when the latter are calculated from recruitment ranges that are *lower* for blacks than for whites. If anything, then, black–white differences in employment are smaller than those expected on the basis of intelligence alone. This finding is consistent with data showing that mean IQs are lower for blacks than for whites in the same occupational category (e.g., Stewart 1947) and for black versus white applicants for the same jobs (Jensen 1977).

Several commentators (Bardis 1985, Das 1985, Sternberg 1985) implied, as Jensen's critics often do, that social conscience and humanitarian concern dictate that scientists *not* persist in studying black–white differences in intelligence. The evidence seems to suggest, however, that the real-life impact of black–white differences in *g* may be larger than expected, raising unsettling questions about the widespread goal of racial parity in employment, particularly in higher level jobs.

Jensen has shown great scientific integrity and personal courage in continuing his important research on black–white differences in *g*, despite an inhospitable social climate that leads others to maintain a discreet silence. Other researchers must follow Jensen's example by opening a responsible scientific dialogue on the societal consequences of black–white differences in *g* under alternative social policies. As was noted in a *BBS* commentary on Jensen's *Bias in Mental Testing* (Gordon 1980, p. 344), "Large IQ differences and changing demographics place racial-ethnic groups on a sociopolitical collision course; before lead-time and goodwill are squandered altogether, it might behoove us to begin giving the various potential scenarios reasoned consideration."

Jensen's data on Spearman's hypothesis: No artifact

William Shockley
Alexander M. Poniatoff Professor (Emeritus) of Engineering and Applied Sciences, Stanford University, Stanford, Calif. 94305

Schönemann's 1985 commentary on Jensen's 1985 target article on Spearman's hypothesis, were it correct (which it is *not*),

would have established that most of the discussions of Spearman's hypothesis in the June 1985 issue of *BBS* were pointless – a waste of print. Schönemann's commentary claimed to prove that Jensen's "impressive empirical evidence in support of 'Spearman's hypothesis'" was not support at all but instead arose from a "psychometric artifact" – a finding implied, possibly facetiously, by its seemingly inappropriate title "Artifical Intelligence." Schönemann's alleged proof was compact: Only 5 lines of displayed equations imbedded in 34 lines of text; there were no mathematical errors. *However*, not recognized in Schoënemann's commentary, nor in Jensen's response to it, was the fact that Schönemann had *not* followed Jensen's prescription for his factor-analytic test of Spearman's hypothesis. The prescription was to calculate *separately* for blacks and for whites the two correlation-coefficient matrices based on the $N \times p$ score matrix of N individuals on p tests; this would exclude data on black–white differences in average test scores. Instead, Schönemann retained these black–white differences by pooling the black and white test scores in a $(2N) \times p$ matrix, thus producing an automatic – but irrelevant – correlation between the g-factor loadings and the black–white differences in average test scores. This irrelevant correlation is Schönemann's "psychometric artifact." Thus the support of Spearman's hypothesis by Jensen's data is meaningful and not an artifact. In a letter to me, Dr. Schönemann states that my "point is well taken" concerning this error and welcomes my quoting his view that "the first order of business is to straighten out flaws in reasoning on all sides, precisely because the potential social implications are so enormous."

Author's Response

Further evidence for Spearman's hypothesis concerning black–white differences on psychometric tests

Arthur R. Jensen
School of Education, University of California, Berkeley, Calif. 94720

Before responding to the commentaries by Brody, Corballis, Gottfredson, and Shockley, I will present new evidence regarding two points that were raised in the first round of commentaries (Jensen 1985). Answers to these questions depended on empirical analyses that could not be completed in the time available for preparation of my initial response (Jensen 1985r).

Gordon (1985). Gordon proposed that Spearman's hypothesis is best examined in terms of the coefficient of congruence between (a) the loadings of the n tests in a battery on the first unrotated principal component of the battery and (b) the point-biserial correlations of each of the n tests with the black–white variable (quantized as 0 and 1). These point-biserial correlations would represent, in effect, a black–white difference factor; that is, the point-biserial correlations would be identical to the factor loadings that would be obtained by including the black–white variable along with the n tests in the principal components analysis and by rotating the axes, or components, so that the black–white variable would have a loading of 1 on one axis. Gordon based his suggestion on the claim by Gorsuch (1974, p. 253) that, mathematically,