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

ıould be expected, smaller estimated black–white ratios nd in the more intellectually demanding jobs, whether or cks are recruited from the same IQ ranges as are whites. tios differ considerably depending on whether recruit-anges for blacks are assumed to be higher than, the same ower than those for whites; however, ratios differ even y job level *within* each of those hypothetical recruitment for blacks. In addition, expected black–white ratios are ır from parity in the more intellectually demanding tions no matter which of the three recruitment ranges for is considered, and ratios approach parity only among the level jobs listed in Table 1. For example, even when ment standards are half a standard deviation *lower* for the ratios are only 1 to 5 for physican and engineer and 1 r secondary school teacher and real estate salesperson. ınt assumptions about the distribution of intelligence in ıck and white populations and about the intelligence ›ments of occupations would produce somewhat different ted black–white ratios for individual occupations, but the *pattern* of ratios would probably be the same under any easonable assumptions: Racially blind worker selection expected to produce especially striking deviations from white parity in higher level jobs.

on *actual* employment ratios in 1970 and 1980 reveal the ıverse relation between black–white ratios and the intel-demands of occupations. Surprisingly, however, actual ment ratios are most similar to estimated recruitment vhen the latter are calculated from recruitment ranges e *lower* for blacks than for whites. If anything, then, white differences in employment are smaller than those ed on the basis of intelligence alone. This finding is ent with data showing that mean IQs are lower for blacks r whites in the same occupational category (e.g., Stewart ınd for black versus white applicants for the same jobs ı 1977).

ıral commentators (Bardis 1985, Das 1985, Sternberg ıplied, as Jensen's critics often do, that social conscience manitarian concern dictate that scientists *not* persist in ıg black–white differences in intelligence. The evidence to suggest, however, that the real-life impact of black–differences in *g* may be larger than expected, raising ing questions about the widespread goal of racial parity in ment, particularly in higher level jobs.

en has shown great scientific integrity and personal cour-continuing his important research on black–white dif-›s in *g*, despite an inhospitable social climate that leads to maintain a discreet silence. Other researchers must Jensen's example by opening a responsible scientific e on the societal consequences of black–white dif-›s in *g* under alternative social policies. As was noted in a ›mmentary on Jensen's *Bias in Mental Testing* (Gordon . 344), "Large IQ differences and changing demograph-e racial-ethnic groups on a sociopolitical collision course; lead-time and goodwill are squandered altogether, it ›ehoove us to begin giving the various potential scenarios ›d consideration."

en's data on Spearman's hypothesis: rtifact

ı Shockley

er M. Poniatoff Professor (Emeritus) of Engineering and Applied s, Stanford University, Stanford, Calif. 94305

›mann's 1985 commentary on Jensen's 1985 target article ›arman's hypothesis, were it correct (which it is *not*), would have established that most of the discussions of Spear-man'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, possi-bly facetiously, by its seemingly inappropriate title "Artificial 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 pre-scription 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 "psy-chometric 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, Co-ballis, 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 hy-pothesis 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 compo-nents, 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,

ıe congruence coefficient, r_c, ıian correlation, r, between ıctor scores of two factors. Va ınventionally regarded as indi r components) both represen ordon noted that the values hite point-biserial correlation r *g*) loadings in each of the 12 ppendix of Jensen's target a .91 to +.99, with a mean of According to the standards b ıuated, the average [congrue ıte that the black–white fact Because the *g* loadings availa ı principal factors (or comm ıan on principal components aim, and because individuals lculated for the test batteries eemed desirable to check C ın's inference based on it) ıngruence coefficient betw ıint-biserial correlations in a adings of the tests on the mponent, and by obtaining tween (c) factor scores deri rrelations and (d) factor sco incipal component.

The raw data for this analy aglieri from a study of 86 bla e 4th and 5th grades, match cioeconomic status (Naglier 987). All children were tes Vechsler Intelligence Sca WISC-R) and the 13 subtests attery for Children (K-ABC ıe usual test of Spearman gure 1, in which the mean b its, based on raw scores) on otted as a function of the sub iman; 1957, hierarchical *g* al factor analysis). The sta ences on the various tests a 78, ρ = +.75) with the tests earman's hypothesis. The med on the *averaged* cor ıthin the black and white gr ctor loadings themselves ca mean test score difference portant point of methodol ockley's cogent commenta orrect criticism of my earman's hypothesis.)

The same 24 × 24 correlati principal components ana relations were compute ck–white variable and sco e congruence coefficient dings and 24 point-biseria ding to Gorsuch (1974, p same as the Pearson corre factor scores derived fron t and from the point-b res were calculated by th h variables. The Pearson the congruence coefficient, r_c, is equivalent to the Pearsonian correlation, r, between the principal component factor scores of two factors. Values of r_c above $+.90$ are conventionally regarded as indicating that the two factors (or components) both represent one and the same factor. Gordon noted that the values of r_c between the black–white point-biserial correlations and the principal factor (or *g*) loadings in each of the 12 test batteries listed in the Appendix of Jensen's target article ranged from about $+.91$ to $+.99$, with a mean of $+.97$. Gordon concluded, "According to the standards by which factors are usually equated, the average [congruence] coefficients . . . indicate that the black–white factor is *g*" (p. 231).

Because the *g* loadings available to Gordon were based on principal factors (or common factor analysis) rather than on principal components, as required by Gorsuch's claim, and because individuals' facfor scores could not be calculated for the test batteries in the target article, it was deemed desirable to check Gorsuch's claim (and Gordon's inference based on it) directly, by obtaining the congruence coefficient between (a) the black–white point-biserial correlations in a battery of tests and (b) the loadings of the tests on the unrotated first principal component, and by obtaining the Pearsonian correlation between (c) factor scores derived from the point-biserial correlations and (d) factor scores derived from the first principal component.

The raw data for this analysis were provided by Jack Naglieri from a study of 86 black and 86 white children in the 4th and 5th grades, matched on age, school, sex, and socioeconomic status (Naglieri 1986; Naglieri & Jensen 1987). All children were tested on 11 subtests of the Wechsler Intelligence Scale for Children-Revised (WISC-R) and the 13 subtests of the Kaufman Assessment Battery for Children (K-ABC), for a total of 24 subtests. The usual test of Spearman's hypothesis is shown in Figure 1, in which the mean black–white differences (in σ units, based on raw scores) on the 24 various subtests are plotted as a function of the subtests' *g* loadings (a Schmid-Leiman; 1957, hierarchical *g* factor derived from a principal factor analysis). The standardized black–white differences on the various tests are positively correlated ($r = +.78$, $\rho = +.75$) with the tests' *g* loadings, as predicted by Spearman's hypothesis. The factor analysis was performed on the *averaged* correlation matrices obtained within the black and white groups *separately;* hence the factor loadings themselves cannot possibly be affected by the mean test score differences between the groups. (This important point of methodology is precisely the gist of Shockley's cogent commentary on Schönemann's, 1985, incorrect criticism of my empirical substantiation of Spearman's hypothesis.)

The same 24×24 correlation matrix was also subjected to a principal components analysis, and the point-biserial correlations were computed between the quantized black–white variable and scores on each of the 24 tests. The congruence coefficient between the 24 component loadings and 24 point-biserial correlations is $+.965$. According to Gorsuch (1974, p. 253), this value should be the same as the Pearson correlation between the two sets of factor scores derived from the first principal component and from the point-biserial correlations. Factor scores were calculated by the same standard method for both variables. The Pearson r between the two sets of

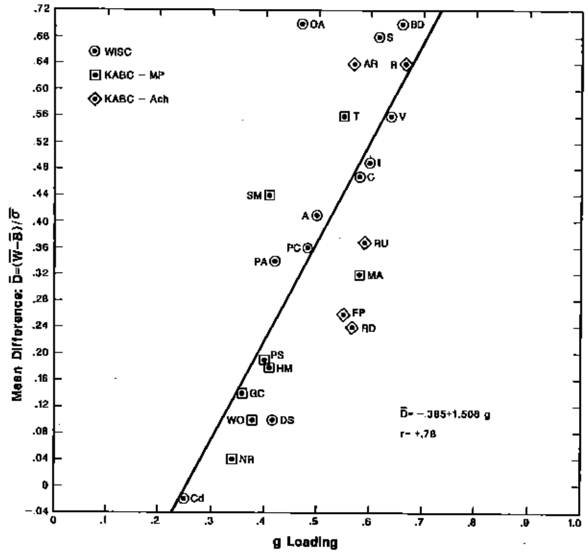

Figure 1. Mean black–white difference (expressed in standard deviation units; see footnote to Table 2) on WISC-R and K-ABC subtests as a function of the subtests' loadings on *g*. WISC-R subtests: I, Information; S, Similarities; A, Arithmetic; V, Vocabulary; C, Comprehension; DS, Digit Span; PC, Picture Completion; PA, Picture Arrangement; BD, Block Design; OA, Object Assembly; Cd, Coding. K-ABC Mental Processing (MP) subtests: HM, Hand Movements; GC, Gestalt Closure; NR, Number Recall; T, Triangles; WO, Word Order; MA, Matrix Analogies; SM, Spatial Memory; PS, Photo Series. K-ABC Achievement (Ach) subtests: FP, Faces and Places; Ar, Arithmetic; R, Riddles; RD, Reading/Decoding; RU, Reading/Understanding. (From Naglieri & Jensen 1987.)

factor scores turned out to be $+.996$, which is higher than the congruence coefficient of $+.965$. The calculations, done by computer, were carefully rechecked. Evidently, there is need of further mathematical clarification of the precise relationship between the congruence coefficient between factors and the Pearson correlation between factor scores, if, indeed, there is an exact mathematical relationship. In any case, the present result is clearly consistent with Gordon's conclusion that "the black–white factor is *g*."

Borkowski and Maxwell (1985). These commentators argued that although a relationship between tests' *g* loadings and the size of the black–white differences had been demonstrated in the target article, it had not been shown that the black–white difference is either exclusively or more highly related to *g* than to any other common factors, besides *g*, that could be extracted from the various batteries of tests which were used to examine Spearman's hypothesis.

The most direct way to investigate this question is to enter the point-biserial correlation of the black–white dichotomous variable (quantized as 0 and 1) with each of the tests in a battery into the matrix of all the test intercorrelations and to subject the matrix to a principal factor analysis. This was done on 10 of the 11 data sets used in the target article, and it is important to note that it was done *separately* for the white (W) and black (B) correlation matrices, which, again, is the point of Shockley's commentary, namely, that the intercorrelations among all the tests are not at all affected by the black–

Table 1. *Factor loadings (decimals omitted) of the black–white dichotomy on the g factor and on the 3 largest non-g factors when the B–W variable is factor analyzed (separately in white and black correlation matrices) among various test batteries in 10 independent studies*

Data set[a]	Number of tests	Matrix group	g factor	non-g factors[b] 1>	2>	3	Ratio of g/non-g variance
Jensen-Reynolds (WISC-R)	13	W	59	25	23	08	2.90
Jensen-Reynolds (WISC-R)	13	B	57	21	19	10	3.54
Nichols	13	W	51	57	52	01	0.42
Nichols	13	B	51	72	37	02	0.39
Dept. of Defense (ASVAB)	10	W	59	08	07	04	27.23
Dept. of Labor (GATB)	8	W	41	35	27	26	0.63
Veroff	6	W	56	68	02	01	0.69
Veroff	6	B	52	39	19	07	1.42
Hennessy	10	W	53	23	15	13	3.07
Hennessy	10	B	48	22	18	06	2.78
Nat'l Longitudinal Study	12	W	68	41	17	12	2.18
Nat'l Longitudinal Study	12	B	70	23	11	07	7.21
Mercer (WISC-R)	12	W	61	18	14	04	6.98
Mercer (WISC-R)	12	B	59	14	13	10	7.67
Sandoval (WISC-R)	12	W	60	34	24	01	2.11
Sandoval (WISC-R)	12	B	60	16	14	08	6.81
Kaufman (K-ABC)	13	W	40	43	41	10	0.44
Kaufman (K-ABC)	13	B	40	31	17	13	1.06
Average[c]			55	37	24	10	4.31

[a]Tests in each data set are listed in Appendix (Tables 4 and 5) of target article (Jensen 1985).
[b]The three largest factor loadings besides g in the B–W dichotomy are given in their order of magnitude regardless of the order of the eigenvalues of the factors on which they occur.
[c]The average of factor loadings is the root mean square.

white differences. (One data set, Reynolds and Gutkin, was omitted, because these data are a subset of the Jensen–Reynolds data set.) Hence, the analysis was performed on 18 independent subject samples (8 black, 10 white). Table 1 shows the loadings of the black–white variable on the *g* factor and the loadings of the black–white variable on the next three largest unrotated principal factors, with the loadings arranged in order of magnitude, regardless of the size of the eigenvalues of the factors on which they occur. (Note: A factor loading of +.45 is approximately equivalent to a mean black–white difference of 1 σ on the corresponding factor scores.) In the majority of cases, the factors for which the non-*g* loadings of the black–white variable are reported are not even significant ones, that is, their eigenvalues are less than 1 and they contain fewer than two tests with loadings greater than .40. Hence the cards are strongly stacked in favor of finding the largest non-*g* loadings of the black–white variable, even if the loadings do not occur on any factors that would conventionally be extracted in a factor analysis. In every data set, four factors were extracted, regardless of the sizes of their eigenvalues. No data set had more than three unrotated primary factors with eigenvalues greater than 1. As can be seen in Table 1, the average *g* loading is considerably larger than the average of the largest non-*g* factors, and the average ratio of *g* to non-*g* variance is greater than 4 to 1. Detailed examination of the 18 analyses indicates two non-*g* factors that show a rather consistent pattern: Blacks are relatively weak on a spatial visualization factor and are relatively strong on a memory span factor. It is noteworthy that the verbal ability factor (independent of *g*) shows a negligible black–white difference.

It can be concluded from these analyses that the black–white difference is indeed predominantly a difference in the *g* factor, as Spearman hypothesized.

Brody. Spearman's hypothesis concerns the relation between the size of the black–white differences on various cognitive tests and the loadings of those tests on the *g* factor. It makes no statement about "intelligence." The fact that certain tests or a composite of scores on a battery of tests that happen to be called "intelligence" tests, or IQ tests, are virtually always highly *g* loaded, or that individual differences on such highly *g*-loaded tests generally appear to reflect common sense notions of "intellectual" ability is merely incidental. Brody, or anyone else, can define "intelligence" as they please. This is not the scientific issue. The *g* factor, in contrast, is an operational construct and can be treated objectively, quantitatively, and empirically.

A good deal is known about psychometric *g*. I have recently detailed this information elsewhere (Jensen 1986; 1987a). Besides the fact that *g* is the main source of variance in the practical predictive validity of virtually all aptitude tests for a great variety of criteria (Thorndike

1985; 1986), *g* is also correlated with a number of variables outside the conceptual realms of factor analysis, psychometrics, and psychology, such as certain physical traits, inbreeding depression, heterosis, blood and body chemistry, myopia, and evoked brain potentials. Psychometric *g* is only the tip of the iceberg, the extent of which many investigators have now begun to explore.

The *g* loadings of various cognitive tests and the *g* factor of any collection of tests, like any other statistical parameters, are subject to variability associated with various factor analytic methods of computing *g*, population sampling error, sampling variation from the total domain of cognitive tests, and measurement error. Yet, despite these potential sources of variability, tests' *g* loadings, and consequently the *g* factor itself, are found to be remarkably stable across various subject samples and various collections of tests, and, like any other statistic, the reliability of *g* increases with increased subject samples and psychometric test samples. The stability of tests' *g* loadings is clearly demonstrated in a recent study by Thorndike (in press) in which 17 target tests were each factor analyzed separately among 6 nonoverlapping test batteries, each composed of 8 diverse tests. The average correlation between the 17 tests' *g* loadings across the 6 test batteries was +.83. The fact that psychometric sampling error is decreased by averaging a test's *g* loadings obtained in 2 or more different batteries means that we can speak of a hypothetical "true" *g* for a given test in the same statistical sense that we can speak of a hypothetical "true" score on a test. The variability between different tests in their *g* loadings is a real phenomenon and is not mainly attributable to the various kinds of sampling errors mentioned previously.

The *g* loading of cognitive tasks seems to be related to the complexity of the information processing they require. Even within the limited range of task complexity seen in simple and choice reaction time (RT) tasks involving 1, 2, 4, or 8 choice alternatives (corresponding to 0, 1, 2, and 3 bits of information) there is a slight, but highly significant, linear increase in the correlations between RT and IQ (or other highly *g*-loaded test scores) as a function of the number of response alternatives scaled in bits (Jensen, 1987b; Jensen & Vernon 1986, Figure 6). Meta-analyses of such data clearly indicate that the *g* loadings of even the simplest cognitive tasks vary in a systematic and continuous fashion which is not attributable to sampling errors of various kinds, when these errors are greatly reduced by being averaged out in a large-scale meta-analysis.

What Brody seems to claim, however, is that various mental tests are either measures of intelligence, in which case they are highly *g* loaded, or they are not measures of intelligence, in which case their *g* loadings are next to nil, and any *g* loadings in between are merely flukes, due possibly to the various types of error mentioned earlier. Brody's position essentially amounts to a denial of the continuity of *g* loadings across the wide variety of tests of mental abilities. He believes that after the nonintelligence tests are culled, the remaining tests of intelligence are "all approximately equivalent or equally good measures of *g*." He further states, "Among the large and diverse class of good measures of intelligence the differences in their loading on *g* may be too small and evanescent to constitute a meaningful dimension. I believe that the Spearman hypothesis as presented by Jensen . . . is without empirical or conceptual foundation."

Brody claims that the black population, on average, scores about one standard deviation below the white on measures of intelligence. Presumably, Brody also regards the variation of the mean black–white differences across various tests as not a continuous variable, and that any black–white differences that fall between 0 and approximately 1 σ are sampling errors.

Let us examine the degree of continuity of *g* loadings and black–white differences on all of the tests that were used in the analyses of the target article. Figures 2 and 3 show the cumulative relative frequency distributions of tests' *g* loadings in the black and white samples. Differences on the horizontal axis would indicate discontinuity. We see no pronounced discontinuities of this kind, and the few small discontinuities are more or less equally dispersed throughout the full range of *g* loadings. These data certainly contradict the impression given by Brody's empirically and theoretically unsupported conjecture.

Another way of looking at the distribution of *g* loadings is in terms of their relative positions *within* all of the studies. This can be accomplished by standardizing the *g* loadings within each study (using *T* scores with a mean of 50 and standard deviation of 10), and plotting these standardized values (i.e., *T* scores) as a function of their rank order of magnitude. Since the number of ranks is

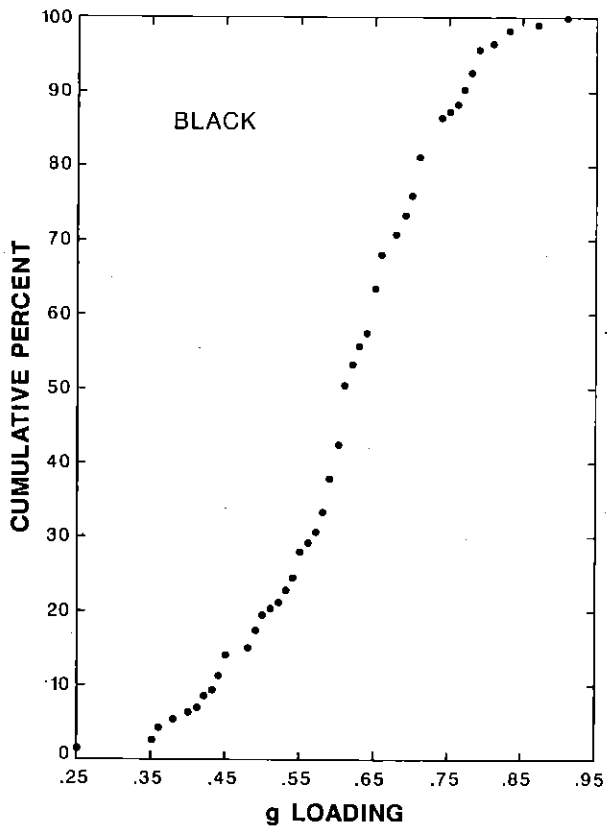

Figure 2. Cumulative relative frequency distribution of *g* loadings obtained in factor analyses of 9 test batteries given to independent black samples.

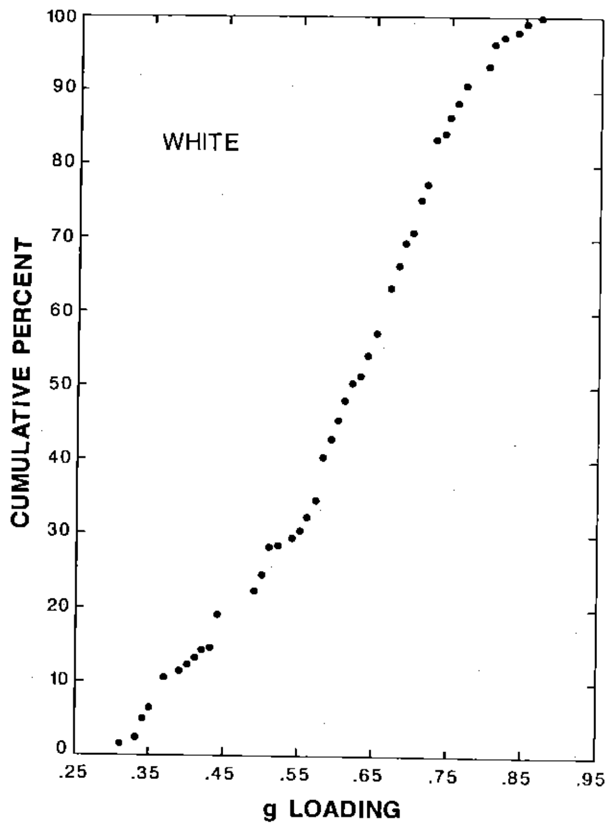

Figure 3. Cumulative relative frequency distribution of g loadings obtained in factor analyses of 10 test batteries given to independent white samples.

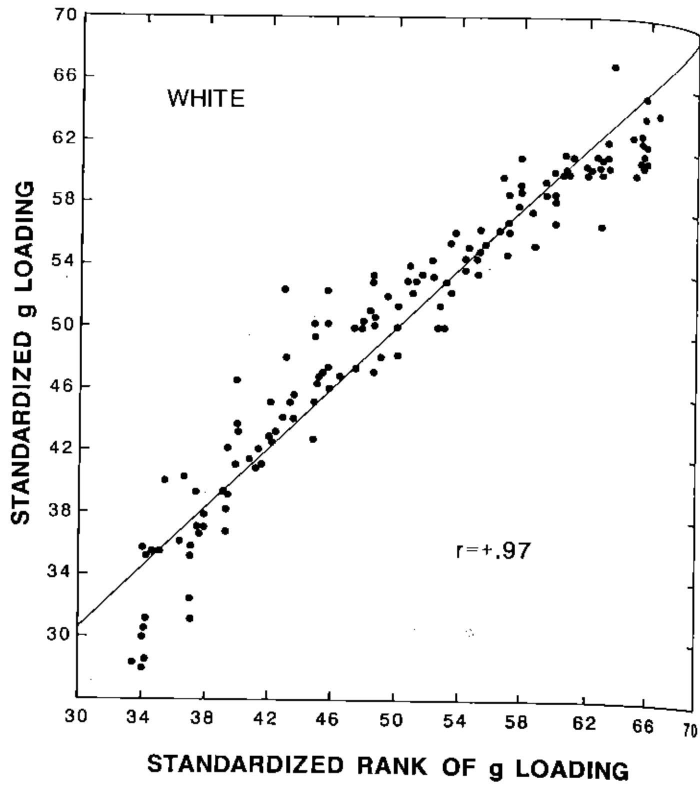

Figure 4. Standardized ($\bar{X} = 50$, $\sigma = 10$) g loadings (obtained in 11 test batteries given to white samples) plotted against their standardized rank order of magnitude within each battery.

equal to the number of tests in a given battery, and differs across studies, we must use standardized ranks also, expressed as T scores. In this way, the relative g loadings of all studies can be plotted on the same coordinates, with standardized g loadings on the ordinate and the standardized ranks of the g loadings on the abscissa. This plot is shown in Figure 4 for the g loadings of the white samples in the 10 data sets used in the target article. In addition, the 24 g loadings of the white sample of Naglieri's data shown in Figure 1 are included in Figure 4. This plot is especially important in terms of Brody's conjecture. If it were true that what Brody would consider tests of intelligence are all highly g loaded and more or less equivalent in g loadings, differing only by random errors, and if those tests that do not measure intelligence have zero or negligible g loadings due to random errors, we should expect the plot in Figure 4 to have an S-shaped form. In the lower-left quadrant there would be a near-plateau of small g loadings, rising only slightly as a function of their rank, and in the upper-right quadrant there would be a similar picture. But the plotted data points contradict this expectation and show a highly linear trend, with $r = +.97$, that is, 94% of the total variance in standardized g loadings is accounted for by their linear regression on their rank order of magnitude. Again, there is no suggestion of discontinuity in g loadings, least of all at the high end. An almost identical picture is found for the g loadings in the black samples, in which r is also $+.97$.

The cumulative relative frequency distribution of the standardized mean black–white differences on the 121 tests in 11 studies is shown in Figure 5. There is only one rather marked discontinuity in this distribution: 8 tests show B–W differences of less than 0.20 σ and the remaining 113 differences range continuously between 0.30 σ and 1.30 σ. It is instructive to look at the 8 tests with the smallest B–W differences and at their g loadings, shown in Table 2. The median B–W difference on all 121 tests is 0.71 σ and the median g loadings for all tests are .62 and .61 in the white and black samples, respectively. The tests with the smallest B–W differences have below-average g loadings, but it may seem surprising that the g loadings are not even lower. These are tests, however, which have substantial loadings on other factors on which black–white differences are either nil or opposite to the g differences, factors involving mainly motor abilities and memory span. Blacks would probably outscore whites on these tests were it not for their moderate g loadings.

If we examine Spearman's hypothesis with only those tests that have g loadings *above* the median g of all the tests in the target article, the correlations between g loadings and B–W differences are +.33 and +.34 in the black and white groups, respectively. The corresponding correlations for those tests with g loadings *below* the median g are +.43 and +.34. The approximate symmetry of these correlations obtained with tests above or below the median g is consistent with the continuity and linearity of the bivariate scatterplot of B–W differences and g loadings. (The correlations are quite low, of course, because of the restriction of range resulting from cutting the distribution of g loadings in half.)

A clear test of Spearman's hypothesis is seen in Figure 1, in which the bivariate plot is based on 24 tests adminis-

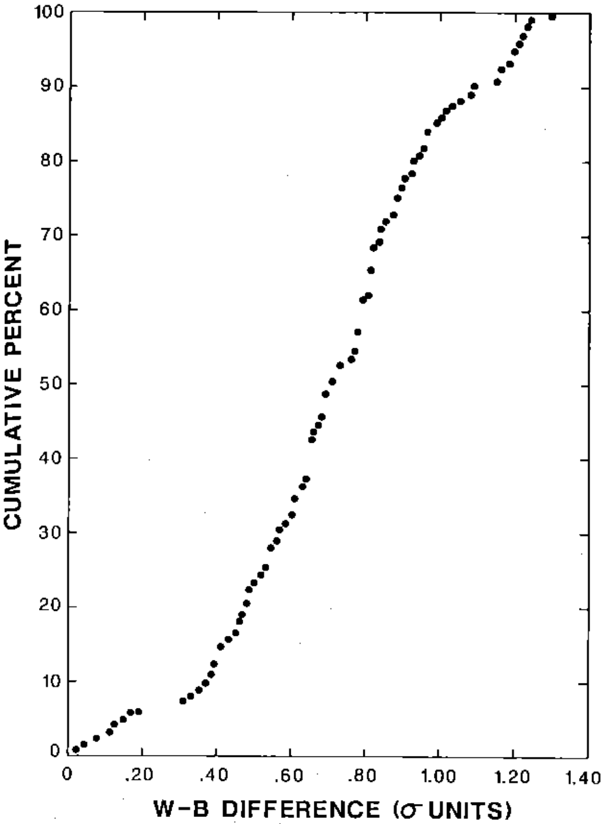

Figure 5. Cumulative relative frequency distribution of mean black–white differences (expressed in standard deviation units) on 121 tests.

Table 2. *The 8 tests (out of 121 tests) with black–white differences of less than 0.20 σ*

Test	Difference[a]	g Loading: White	g Loading: Black
GATB: Motor Coordination	0.02 σ	.52	—
K-ABC: Number Recall	0.04 σ	.59	.51
GATB: Manual Dexterity	0.08 σ	.37	—
Draw-a-Man	0.11 σ	.44	.49
WISC-R: Digit Span	0.12 σ	.44	.59
K-ABC: Word Order	0.15 σ	.62	.66
WISC-R: Coding	0.17 σ	.31	.25
CGP: Intersections	0.19 σ	.31	.25
Average	0.12 σ	.46	.48

[a]The mean difference is expressed in standardized units: $(\bar{W} - \bar{B})/\bar{\sigma}$, where $\bar{W}$ and $\bar{B}$ are the raw score means of the white and black samples and $\bar{\sigma}$ is the average standard deviation of raw scores within the two samples. $\bar{\sigma} = [(N_w\sigma_w^2 + N_B\sigma_B^2)/(N_w + N_B)]$, where σ_w and σ_B are the standard deviations in the white and black samples and N_w and N_B are the numbers in the white and black samples.

tered to samples of 86 black and 86 white children matched for the socioeconomic status of their parents. The bivariate data points reveal no discontinuity or significant departure from linearity.

It should be emphasized that the correlation between B–W differences and *g* loadings is not an artifact of factor analysis, as some critics have argued. The *g* loadings are the *average* of the loadings of each test when the battery is factor analyzed *within* the black and white groups *separately*. Hence, as **Shockley** has noted, the tests' *g* loadings could not be at all influenced by the black–white differences; the *g* loadings and black–white differences are experimentally independent. Yet they are correlated +.78. Hence the criticism that the demonstration of Spearman's hypothesis is artifactual or tautological is not valid.

I believe the total evidence I have amassed amply demonstrates that the black–white differences in performance on mental tests is intimately connected with *g*. Advances in understanding black–white differences will depend in large part on gaining a greater knowledge of the nature of *g*. To dissociate these two phenomena, as Brody does, seems scientifically unproductive.

Corballis. Corballis is one of the developers of "longitudinal factor analysis," a method that permits the analysis of changes in factor structure or factor loadings of given tests throughout the course of development from early childhood to late maturity. The three main findings from his comparison of principal factor analysis and longitudinal factor analysis to the Primary Mental Abilities (PMA) tests obtained on the same group of children tested in grades 5 and 8 are noteworthy in terms of *g* theory: (1) The first principal factor and the first longitudinal factor are essentially the same *g* factor, with congruence coefficients (r_c) of +.995 and +.979 at grades 5 and 8, respectively; (2) the longitudinal *g* is slightly more congruent (r_c = +.996) between grades 5 and 8 than is the first principal factor *g* (with r_c = +.976); (3) in both types of factor analysis the first factor, or *g*, has much greater stability over the three-year interval than has the second factor, orthogonal to *g*, which has congruence coefficients between grades 5 and 8 of +.817 and +.877, for the principal factors and longitudinal factors, respectively. Congruence coefficients below +.90 are conventionally regarded as indicating that the factors are not the same. The much greater stability of the *g* factor than of other factors over the course of development is an important item of evidence in *g* theory. One hopes to see the future application of Corballis's method to more extensive longitudinal data from much larger batteries of diverse cognitive tests.

I find myself in disagreement with Corballis on one point, namely, his idea that Thurstone's concept of "simple structure" and *g* are in opposition ("*g* versus simple structure"). In my view, the concept of simple structure is basic to *g* in the factor analysis of human abilities. A hierarchical *g* factor is the logical consequence of maximizing the approximation to simple structure, which, because of the existence of the positive manifold, can be achieved in the factor analysis of ability tests only by means of oblique (i.e.. correlated) factors. Factor analysis of oblique first-order factors creates a hierarchical factor structure, at the apex of which is *g*. This is usually attained at the second or third order of factorization. The basic theoretical justification for hierarchical analysis, and for *g* itself, is based on Thurstone's concept of simple structure. Simple structure is such a strongly compelling rationale for the rotation and interpretation of factors that, to my knowledge, it has never been seriously challenged.

Gottfredson. Gottfredson's data clearly substantiate Cattell's (1985) surmise that "the percentage of blacks [in various occupations] should be inversely correlated with the mean intelligence levels of persons holding the occupations" (p. 228). That the mean IQ of persons employed in various occupations ranges widely reflects the fact that occupations differ in their g demands. As Gottfredson has pointed out, and has impressively elaborated elsewhere (Gottfredson 1986), this inescapable fact has important social, economic, and political consequences when there are substantial statistical differences in the distribution of g among various segments of the population that are identifiable along racial or cultural lines. Because of the well-established variation in the g demands of different occupations and of the differences in the distribution of g in different subpopulations, it seems unlikely that departures from parity in the representation of different racial or cultural populations in all occupations is entirely, or even predominantly, a result of prejudice and racial discrimination per se. Sociologists should follow Gottfredson's lead in applying similar methods of analysis to other g-correlated variables of social significance, of which there are probably many more than we presently foresee.

Shockley. Schönemann's (1985) commentary attempted to prove mathematically that the methodology of my empirical test of Spearman's hypothesis necessarily produced a result that is purely a psychometric artifact. In my reply (Jensen 1985r, p. 248) to Schönemann, I unfortunately failed to point out the precise error in Schönemann's reasoning that was responsible for his false conclusion. Hence I am most grateful for Shockley's pointed refutation of Schönemann's misconceived argument. **Brody** should perhaps note the refutation of Schönemann's argument, and so should the few earlier commentators who suggested (but without attempting a formal "proof") essentially the same methodological artifact, or "circular reasoning," as claimed by Schönemann. Their arguments, too, would seem to fall with Schönemann's. The striking phenomenon demonstrated by my investigation, the social implications of which Schönemann apparently believes to be "enormous," is indeed not a psychometric artifact, but an empirical fact.

References

Anderson, S. B. & Maier, M. H. (1963) 34,000 pupils and how they grew. *Journal of Teacher Education* 14:212–16. [MCC]

Bardis, P. D. (1985) Jensen, Spearman's g, and Ghazali's dates: A commentary on interracial peace. *Behavioral and Brain Sciences* 8:219–20. [LSG]

Borkowski, J. G. & Maxwell, S. E. (1985) Looking for Mr. Good-g: General intelligence and processing speed. *Behavioral and Brain Sciences* 8:221–22. [rARJ]

Brand, C. (1985) Jensen's compromise with componentialism. *Behavioral and Brain Sciences* 8:222–23. [MCC]

Brody, N. (1985) The validity of tests of intelligence. In: *Handbook of intelligence*, ed. B. Wolman. Wiley. [NB]

Carlson, J. S. (1985) The issue of g: Some relevant questions. *Behavioral and Brain Sciences* 8:224–25. [LSG]

Cattell, R. B. (1985) Intelligence and g: An imaginative treatment of unimaginative data. *Behavioral and Brain Sciences* 8:227–28. [rARJ, LSG]

Corballis, M. C. (1973) A factor model for analysing change. *British Journal of Mathematical and Statistical Psychology* 26:90–97. [MCC]

Corballis, M. C. & Traub, R. E. (1970) Longitudinal factor analysis. *Psychometrika* 35:79–98. [MCC]

Cronbach, L. J. & Snow, R. (1977) *Aptitudes and instructional methods*. Irvington. [NB]

Das, J. P. (1985) Interpretations for a class on minority assessment. *Behavioral and Brain Sciences* 8:228. [LSG]

Dixon, W. J. & Massey J., Jr. (1969) *Introduction to statistical analysis*. McGraw-Hill. [LSG]

Eysenck, H. J. (1985) The nature of cognitive differences between blacks and whites. *Behavioral and Brain Sciences* 8:229. [LSG]

Gordon, R. A. (1980) Implications of valid (and stubborn) IQ differences: An unstatesmanlike view. *Behavioral and Brain Sciences* 3:343–44. [LSG]

(1985) The black–white factor is g. *Behavioral and Brain Sciences* 8:229–31. [rARJ, LSG]

Gorsuch, R. L. (1974) *Factor analysis*. W. B. Saunders. [rARJ]

Gottfredson, L. S. (1984) *The role of intelligence and education in the division of labor*. Report No. 355. Johns Hopkins University, Center for Social Organization of Schools. [LSG]

(1985) Education as a valid but fallible signal of worker quality: Reorienting an old debate about the functional basis of the occupational hierarchy. In: *Research in sociology of education and socialization*, vol. 5, ed. A. C. Kerckhoff. JAI Press. [LSG]

(1986) Societal consequences of the g factor in employment. *Journal of Vocational Behavior* 29:379–410. [rARJ, LSG]

(in press) Intelligence versus training: Job performance and black–white occupational inequality. *Personality and Individual Differences*. [LSG]

Gottfredson, L. S. & Brown, V. C. (1978) Holland codes for the 1960 and 1970 censuses: Detailed occupational titles. JSAS *Catalog of Selected Documents in Psychology* (No. 1660) 8:22. [LSG]

Gustafsson, J.-E. (1985) Measuring and interpreting g. *Behavioral and Brain Sciences* 8:231–32. [MCC]

Hawk, J. A. (1970) Linearity of criterion-GATB aptitude relationships. *Measurement and Evaluation in Guidance* 2:249–51. [LSG]

Hilton, T. L. & Myers, A. E. (1967) Personal background, experience and school achievement: An investigation of the contribution of questionnaire data to academic prediction. *Journal of Educational Measurement* 4:69–80. [MCC]

Hitchcock, D. C. (1976) *Intellectual development and school achievement of youths 12–17 years: Demographic and socioeconomic factors*. (Vital and Health Statistics: Series 11, No. 158; DHEW Publication No. 77–1640) U.S. Government Printing Office. [LSG]

Hunter, J. E. (1983a) A causal analysis of cognitive ability, job knowledge, job performance, and supervisor ratings. in: *Performance measurement and theory*, ed. F. Landy, S. Zedeck & J. Cleveland. Erlbaum. [LSG]

(1983b) *Test validation for 12,000 jobs: An application of job classification and validity generalization analysis to the General Aptitude Test Battery (GATB)*. Division of Counseling and Test Development, Employment and Training Administration, U.S. Department of Labor. [LSG]

Hunter, J. E. & Hunter, R. F. (1984) Validity and utility of alternative predictors of job performance. *Psychological Bulletin* 96:72–98. [LSG]

Hunter, J. E., Schmidt, F. L. & Rauschenberger, J. (1984) Methodological, statistical, and ethical issues in the study of bias in psychological tests. In: *Perspectives on bias in mental testing*, ed. C. R. Reynolds & R. T. Brown. Plenum. [LSG]

Jencks, C. (1972) *Inequality: A reassessment of the effect of family and schooling in America*. Basic Books. [NB]

Jensen, A. R. (1977) An examination of culture bias in the Wonderlic Personnel Test. *Intelligence* 1:51–64. [LSG]

(1980) *Bias in mental tests*. Free Press. [NB]

(1985) The nature of the black–white difference on various psychometric tests: Spearman's hypothesis. *Behavioral and Brain Sciences* 8:193–263.

(1985r) The black–white difference in g: A phenomenon in search of a theory. *Behavioral and Brain Sciences* 8:246–58. [rARJ]

(1986) g: Artifact or reality? *Journal of Vocational Behavior* 29:301–31. [rARJ]

(1987a) The g beyond factor analysis. In: *The influence of cognitive psychology on testing and measurement*, ed. J. C. Conoley, J. A. Glover & R. R. Ronning. Erlbaum. [rARJ]

(1987b) Individual differences in the Hick paradigm. In: *Speed of information processing and intelligence*, ed. P. A. Vernon. Ablex. [rARJ]

Jensen, A. R. & Reynolds, C. R. (1982) Race, social class and ability differences on the WISC-R. *Personality and Individual Differences* 3:423–38. [NB]

Jensen, A. R. & Vernon, P. A. (1986) Jensen's reaction-time studies: A reply to Longstreth. *Intelligence* 10:153–79. [rARJ]

Matarazzo, J. D. (1972) *Wechsler's measurement and appraisal of adult intelligence*, 5th ed. Williams & Wilkins. [LSG]

McDaniel, M. A. (1986) The evaluation of a causal model of job performance: The interrelationships of general mental ability, job experience, and job performance. Unpublished doctoral dissertation, George Washington University. [LSG]

Mercer, J. R. (1948) What is a racially and culturally non-discriminatory test? In: *Perspectives on bias in mental testing*, ed. C. R. Reynolds & R. T. Brown. Plenum. [NB]

Meyer, W. J. & Bendig, A. W. (1961) A longitudinal study of the Primary Mental Abilities Test. *Journal of Educational Psychology* 52:97–103. [MCC]

Naglieri, J. A. (1986) WISC-R and K-ABC comparison for matched samples of black and white children. *Journal of School Psychology* 24:81–88. [rARJ]

Naglieri, J. A. & Jensen, A. R. (1987) Comparison of black–white differences on the WISC-R and the K-ABC: Spearman's hypothesis. *Intelligence* 11:21–43. [rARJ]

Rabbitt, P. M. A. (1985) Oh g Dr. Jensen! or, g-ing up cognitive psychology? *Behavioral and Brain Sciences* 8:238–39. [LSG]

Reynolds, C. R. & Gutkin, T. B. (1981) Multivariate comparison of the intellectual performance of blacks and whites matched on four demographic variables. *Personality and Individual Differences* 2:175–80. [NB]

Rushton, J. P. (1985) Differential K theory and group differences in intelligence. *Behavioral and Brain Sciences* 8:239–40. [LSG]

Sandoval, J. (1982) The WISC-R factor validity for minority groups and Spearman's hypothesis. *Journal of School Psychology* 20:198–204. [NB]

Scarr, S. (1981) *Race, social class, and individual differences in IQ*. Erlbaum. [NB]

Schmid, J. & Leiman, J. M. (1957) The development of hierarchical factor solutions. *Psychometrika* 22:53–61. [rARJ]

Schmidt, F. L. & Hunter, J. E. (1981) Employment testing: Old theories and new research findings. *American Psychologist* 36:1128–37. [LSG]

Schönemann, P. H. (1985) On artificial intelligence. *Behavioral and Brain Sciences* 8:241–42. [rARJ, WS]

Stanovich, K. E. (1985) The black–white differences are real: Where do we go from here? *Behavioral and Brain Sciences* 8:242–43. [LSG]

Sternberg, R. J. (1985) The black–white differences and Spearman's *g*: Old wine in new bottles that still doesn't taste good. *Behavioral and Brain Sciences* 8:244. [MCC, LSG]

Stewart, N. (1947) Relationship between military occupational specialty and Army General Classification Test standard score. *Educational and Psychological Measurement* 7:677–93. [LSG]

Thorndike, R. L. (1985) The central role of general ability in prediction. *Multivariate Behavioral Research* 20:241–54. [rARJ, LSG]

(1986) The role of general ability in prediction. *Journal of Vocational Behavior* 29:332–39. [rARJ]

(in press) Stability of factor loadings. *Personality and Individual Differences*. [rARJ]

U.S. Bureau of the Census (1971) *1970 census of population: Classified index of industries and occupations*. U.S. Government Printing Office. [LSG]

(1973) *Census of population: 1970 subject reports, Final Report PC(2)-7A Occupational characteristics*. U.S. Government Printing Office. [LSG]

(1982) *1980 census of population: Classified index of industries and occupations*. U.S. Government Printing Office. [LSG]

(1983) *Detailed occupation and years of school completed by age, for the civilian labor force by sex, race, and Spanish origin: 1980*. Supplementary Report PC80-S1-8. U.S. Government Printing Office. [LSG]

U.S. Department of Labor (1970) *Manual for the USTES General Aptitude Test Battery*. U.S. Government Printing Office. [LSG]

Wilson, J. R. (1985) Jensen's support for Spearman's hypothesis is support for a circular argument. *Behavioral and Brain Sciences* 8:246. [MCC]

Commentary on **Nicholas P. Spanos (1986) Hypnotic behavior: A social-psychological interpretation of amnesia, analgesia, and "trance logic." BBS 9:449–502.**

Abstract of the original article: This paper examines research on three hypnotic phenomena: suggested amnesia, suggested analgesia, and "trance logic." For each case a social-psychological interpretation of hypnotic behavior as a voluntary response strategy is compared with the traditional special-process view that "good" hypnotic subjects have lost conscious control over suggestion-induced behavior. I conclude that it is inaccurate to describe hypnotically amnesic subjects as unable to recall the material they have been instructed to forget. Although amnesics present themselves as unable to remember, they in fact retain control over retrieval processes and accommodate their recall (or lack of it) to the social demands of the test situation. Hypnotic suggestions of analgesia do not produce a dissociation of pain from phenomenal awareness. Nonhypnotic suggestions of analgesia and distractor tasks that deflect attention from the noxious stimuli are as effective as hypnotic suggestions in producing reductions in reported pain. Moreover, when appropriately motivated, subjects low in hypnotic suggestibility report pain reductions as large as those reported by highly suggestible hypnotically analgesic subjects. Finally, the data fail to support the view that a tolerance for logical incongruity (i.e., trance logic) uniquely characterizes hypnotic responding. So-called trance-logic-governed responding appears to reflect the attempts of "good" subjects to meet implicit demands to report accurately what they experience.

Phenomenal awareness and self-presentation

Donald R. Gorassini

Department of Psychology, King's College, University of Western Ontario, London, Ontario, Canada N6A 2M3

The recent target article by Spanos (1986) underscores a central controversy in the field of hypnosis research. The disagreement is between the special process position, which holds that traditionally studied hypnotic phenomena are indicators of hypnotically induced cognitive changes; and a social-psychological position, which holds that the phenomena reflect deliberate attempts by subjects to appear hypnotized. To inform the issue, the target article contains a review of the substantial literature on three of the most studied hypnotic responses: amnesia, analgesia, and trance logic. Judging from the evidence, rather strong support for the social-psychological position is available, whereas there is rarely support for and often counterevidence against the special-process view. Yet, as a reading of the several commentaries that follow the target article quickly demonstrates, not everyone is in fundamental agreement with this conclusion. Several of the commentaries indicate skepticism that a social-psychological account can explain the kinds of subjective reports commonly made by hypnotic subjects (see the commentaries of P. Bowers; Evans; Kihlstrom; Kirsh; Orne, Dinges & Orne; Spiegel). It is to this important aspect of the controversy that I will direct my own comments.

When asked, some hypnotic subjects indicate that they experienced the cognitive changes that were suggested by the experimenter. Subjective accounts of this kind are often made