

---

## AUTHOR'S RESPONSE

---

### An Unrepentant Author

Lloyd G. Humphreys

University of Illinois, Urbana–Champaign

There is a thread that runs through most of the commentaries. The commentators want something deeper, more basic, more causal. These somethings are latent traits, capacities, powers, or entities. Jensen's (and Spearman's) *g* is a causal entity even though Jensen believes that the determinants are independent of one another (see Kranzler & Jensen, 1993, but also Carroll, 1993). For Brody, this something-more-basic is information-processing abilities. Carroll states that he is a factor analyst in the Thurstone tradition. Thurstone described factors as "primary mental abilities," and Carroll follows Thurstone in wanting them to be capacities. Because there is no way to measure these deeper constructs directly, they must be inferred from observations. It is incumbent on these critics to demonstrate that a hypothetical construct does more than make them and their readers feel good. Specific, testable hypotheses must be derived and research must be designed that provides for the possibility of disconfirmation. It is also important that theories be pitted against one another rather than against a null outcome (MacCallum, Wegener, Ushino, & Fabrigar, 1993).

#### My Alternative

Detterman evaluates my approach as "Dust Bowl empiricism," which I accept. It is also compatible with my training, professional experience, and the title of my target article. A positive manifold among cognitive measures—with *cognitive* defined either by Jensen or by my consensus among experts that does not require 100% agreement—defines a general intelligence that fits the definition of what Meehl (1986) called a *surface quasi-trait*. It is a surface trait because it depends on covariation among measures. It is a quasi-trait because there are no obvious sources of similarity. This lack requires a latent source of determinants that can be the multitude of genes and environmental events that combine, in varying degrees, to shape the anatomical structures and chemical processes in the developing organism.

I also state that a standard intelligence test estimates a person's standing on a general factor of intelligence that may be no more than a mathematical dimension.

This dimension is not estimated in an unbiased fashion by the first principal component of an R-matrix. It requires operations that allow for the full hierarchical model. Meehl (1986), following MacCorquadale and Meehl (1948), would call the general factor an intervening variable. It is not a causal entity and is not likely to become one soon. This is why I prefer *general factor* to *g* because the latter carries so much excess baggage, dating back to Spearman's *mental energy*.

A quasi-trait is still a phenotypic trait. If one selects a heritability coefficient—Ceci should note that my confidence limits provide a wide range of choice—and obtains the square root, a simple regression equation provides an estimate of a genotype from an observed phenotype. However, this adds no new information because the predictor and the estimate are perfectly correlated. The important information for theory is the research basis for the heritability conclusion, but this knowledge does not automatically enter a prediction equation.

#### Jensen's Epiphenomena

When a printed test highly correlated with a standard test of intelligence was correlated with a reaction-time score and a measure of speed of nervous conduction (in Vernon & Mori, 1992), a one-factor solution to the  $3 \times 3$  R-matrix resulting from the averaging of the matrices in the two studies produced an excellent fit. The general intelligence test has a loading of 1.00, and the supposedly more basic measures have loadings of .45.

One could add standing height, parental socioeconomic status (SES), and the Humphreys, Davey, and Kashima (1986) Physical Wellness and Wholeness Scale without requiring an additional factor. The first of these additions would have a loading of about .25 and the second and third of about .40 in the full range of high-school talent. Measures of performance in roles considered important by our society would also have loadings of a variety of sizes on the same factor.

At least for the foreseeable future, it is wise to consider measures such as speed of nervous conduction, visual evoked potential, brain size, and simple reaction time as a few of the many correlates of scores on a test of general

intelligence. I predict that these supposedly more basic measures will show no more stability during development than a standard test of intelligence.

Jensen fails to note that Loevinger (1951) was critical of all interpretations of factors. Because the difference between a unitary-determinant and a multiple-determinants explanation is not subject to empirical test from behavioral evidence, it is time to look carefully at the observable correlates of the general factor. Jensen also misinterprets my discussion of the difference between the traits of intelligence and of relative intelligence. This difference is important just as it is for height and relative height. For example, a retarded young adult trained as a caregiver can provide an intellectually stimulating environment for a preschool child with an IQ of 150. Intelligence was described as monotonic with the size of the repertoire because I was uncertain about the units of measurement for the growth of intelligence.

### Carroll's Primary Mental Abilities

Carroll criticizes my thousands of factors (Humphreys, 1981) for which I had two reasonable arguments: First, correlation between different measures that is less than unity after correcting for attenuation means that they are to some extent measuring different factors; second, any substantive change in one of the facets along which tests can be placed will produce a correlation less than unity with its nearest "relatives." Some of the factors defined in this way will be small so that large *N*s and careful test construction will be required to define them in a replicable fashion. There is no criterion of size of contribution of a common factor to the total variance of a measure that allows the factor to be considered primary.

I am afraid that I agree with Kelley (1939) that replicability and sizable contribution to variance of a factor may still place that factor in the category of no or limited importance. My preference is to start at the top of the hierarchy with the general factor and move down the hierarchy as needed to describe (explain?) behavior outside the testing room. The variance accounted for in measures of performance in roles valued by our society should be primary. General intelligence describes much more than 50% of the criterion variance in education and occupations that can be attributed to predictor tests. Furthermore, an intelligence test can increase the proportion of criterion variance accounted for when criterion performance is measured in a fashion parallel to that of a good intelligence test. That is, introduce as many different measures of performance as possible within the limits of the definition of the variable to be measured.

Cronbach and Snow (1977) demonstrated that the overlap model is untenable if a zero correlation be-

tween true-score bases and gains is required. It is still true, however, that a great deal of the stability of IQs from year to year is due to overlap. Cronbach and Snow's example, based on Bayley's (1949) small longitudinal sample, did produce a correlation of .60 between gain in true scores from age 6 to age 17 and the base at age 6, but they also presented a correlation for the gain from age 12 to age 17, which was about .30 (read from a graph).

The observed intercorrelations for Grades 5, 7, 9, and 11 in Humphreys and Parsons (1979) also formed a quasi-simplex, and the adjacent occasion stabilities were so high that only very small positive correlations between true-score bases and gains can be tolerated. A capacity that contributes substantially to variance in their data is untenable, and, in Cronbach and Snow's (1977) analysis, a capacity contributes only 36% of the variance of gain between age 6 and age 17 and 9% between age 12 and age 17. Correlations among gains made between adjacent occasions must be smaller than the correlations between bases and gains.

### Brody's Disagreements

Brody and I do not disagree on the importance of distinguishing between *genotype* and *phenotype*, but we do disagree on the proper usage of *genotype*. It is highly undesirable to equate *genotype* and *latent trait*. Latent traits are estimated from phenotypic measures. When defined by LISREL, it is a statistical construct that measures what its estimators have in common with one another. The latent trait is like the general factor in intelligence in that it may represent nothing more than a mathematical dimension. The second half of Brody's second paragraph would be less confusing and would become acceptable if he were to say that he hopes to find phenotype traits that would add to the information furnished by a test of general intelligence.

Information processing in infancy may be a likely source of additional information, but adequate data are not yet available. Samples are too small, number of measures are too few, and number of occasions in longitudinal studies are also too few. I predict that the information-processing tasks in infancy will be less highly correlated with intelligence at age 5 or 6 than with intelligence at age 2 or 3. Rejection of that hypothesis would indeed be a major blow to my theory.

There are educational, training, and job-definition changes that reduce the importance of individual differences in intelligence. Redefine jobs to narrow substantially the range of skills required. Educate and train narrowly, utilizing more drill and less problem solving. These changes are counter to present trends in education and industry.

Both Brody and I need more and better data concerning the effects of strong forms of affirmative action

selection in education, hiring, and promotion. I offered a plausible causal sequence based on dependable data; the regression of performance on valid test scores in a mixed-race sample shows that Black performance is slightly overpredicted by the test; strong forms of affirmative action produce substantial mean differences on valid predictors for Blacks and Whites; on average, peers are able to evaluate one another accurately with respect to performance in education and on jobs, and they do this without encouragement or despite active discouragement. It seems reasonable that strong forms of affirmative action lay the groundwork in both races for the development of undesirable attitudes that increase divisiveness. I agree that there must be weighing of costs and benefits in order to evaluate the net utility of the policy.

The weighing of costs and benefits is determined by values, but their size should be determined empirically. Three of my values especially relevant to the costs of strong forms of affirmative action can be briefly stated for Brody and interested readers. One is viewing each human being as an individual without regard to group membership, no matter how the group may be defined. I consider this a truly basic democratic value. A second value is the importance of recognizing, rewarding, and capitalizing on merit for the greater good. I recognize that merit is found along several dimensions in addition to the important one of general intelligence. The third value is defined by the inscription *e pluribus unum* on the Great Seal of the United States. I do not want to overemphasize the many at the expense of the one, but many proposed and actual policies concerning multiculturalism and diversity do overemphasize the many and place the one in jeopardy. The extent to which these and other values are shared has important consequences but is immaterial with respect to an individual taking a position.

### Ceci's Concerns

Based on Stich's (1992) account, Ceci recounts the military experience with enlisted personnel when misnorming of the Armed Forces Qualification Test (AFQT) allowed large numbers of previously unqualified persons to enter military service. Stich minimized the effects of the infusion of personnel with substantially inflated test scores. In doing this, he also minimized the importance for military service of what the AFQT measured, which was the general factor of intelligence.

There is more to Stich's story that can be found in Maier (1993). The error in the upper half of the distribution of AFQT was discovered almost immediately by military trainers, and a correction was applied. It took longer for the service's psychometricians to acknowl-

edge the reality of the problem in the lower half of the distribution despite complaints from the field and to correct the norms for the lower half of the distribution. Interim adjustments were made in the field by changing qualifying scores for assignment to specialties. The Army, for example, increased the qualifying score for more than 50 assignments.

The military services needed personnel badly in the early days of the all-volunteer force, and commanders made do with the personnel available. That a relatively large proportion was rated satisfactory by a supervisor was overinterpreted by Stich. Ratings tend to be relative to the persons being rated. An A at a postsecondary institution with a relatively recent change in name from *teachers college* to *university* is not the equivalent of an A in most curricula of the state's land-grant university. A manager considers a marginal employee satisfactory in a tight labor market but releases the same person if there are many applicants for every vacancy.

Ceci also downgrades the importance of general intelligence in his interpretation of a finding that test scores do not add to the accuracy of prediction of adult income when parental SES is controlled. A related finding, although not mentioned by Ceci, is that parental SES plays a larger role in college entrance and presumably graduation than its value as a predictor of merit in performance warrants. I interpret such data as revealing faults in our democracy.

Ceci may prefer to view the cognitive domain situationally, but he is wrong if he believes that psychometricians are unaware of the difference situations make. Every single behavioral act, including responses both to the tasks used in cognitive research and to the items on an intelligence test, measures mainly unique variance (Humphreys, 1976). Uniqueness is sometimes dependent on a current situation, sometimes on past situations. It is instructive to realize that a test of 100 items with a Kuder-Richardson coefficient of .95 has a mean interitem correlation of about .16. Linear combinations of both predictor items and criterion performance acts are required for reasonable accuracy in prediction.

A broadly based academic achievement test is a more valid substitute for a Wechsler or Stanford-Binet test of general intelligence than the Raven Progressive Matrices aptitude test when the examinees have been adequately "exposed" to the curriculum content. Fewer than 100 years ago, it was possible to find large numbers of young persons in the United States who had had little contact with schools and whose Stanford-Binet score was much higher than their reading, writing, and arithmetic. Military experience in basic-skills training in English shows sizable gains for Puerto Ricans but not for Blacks from American schools. Hispanic samples also obtain higher mean scores on the Wechsler Performance scales than on the Verbal scales.

### Is Behaviorism a Sin?

Flynn concentrates his critique of behaviorism on Skinner's variety, but my professional and personal antecedents include Hilgard, Tolman, Guthrie, Hull, and Kantor. Admittedly, the title of my target article was a "chip on the shoulder" reaction to the cockiness of my cognitive colleagues, but I have other colleagues who will forgive me. Flynn says in one context that I have no case to argue, because "we are all behaviorists." His "we" is now a small group.

IQ gains over time do pose a problem for me, but Flynn's anecdotal correlates of gains are not convincing. Consider Raven Progressive Matrices. Its loading on the general factor obtained in the manner that I described is only moderate, perhaps .5 to .6. There is also a visuo-perceptual component, probably smaller than .5. Absence of a sex difference on total score can be a combination of plus-and-minus differences on the several parts. The total test score is quite reliable, with perhaps a loading of .30 for error. (All of my estimates assume a wide range of talent.) There is still room for "methods" components based on the administrative directions, item content, and the specific task set for the examinee. For example, analogies tasks differ from similarities and series tasks used so commonly in so-called fluid intelligence tests. Gains over time can take place on any component of variance or on a combination of components.

A test of general intelligence constructed in accordance with my specifications minimizes both methods and group factor variance. I expect gains on such a test over time to result in a higher level on a measure of performance in a valued role in our society. This testable hypothesis requires more than impressions concerning performance then and now to test it. Flynn praises Jensen for his hypotheses but does not mention the more numerous ones in my article that are not dependent on a causal *g*. If Flynn wishes to use pretheory for one that starts small and does not try to explain everything, I cannot quarrel with him. Psychology's philosophical heritage has impelled many psychologists to neglect acquisition of dependable data, formation of a limited theory to tie the data together, and expansion and modification of the theory as research is conducted on the hypotheses generated. I hope that Flynn's conception of a real theory is more akin to those in the physical and biological sciences than to the norm in psychology.

### Detterman's Praise

Detterman's contribution to the debate is accepted happily. He has grasped my intent, which was to define general intelligence in a fashion that is congruent with the standard tests and with the dependable correlates of

those tests. I also intended that users and critics of intelligence tests would consider my definition before misinterpreting scores on intelligence tests. Detterman agrees that this would be desirable. Flynn appears to be more critical than Detterman, but his evaluation may not be fundamentally different.

### Notes

This research was supported by the Pioneer Fund.

Lloyd G. Humphreys, Department of Psychology, University of Illinois, 603 East Daniel Street, Champaign, IL 61820.

### References

- Bayley, N. (1949). Consistency and variability in the growth of intelligence from birth to eighteen years. *Journal of Genetic Psychology, 75*, 165–196.
- Carroll, J. B. (1993). The unitary *g* problem once more: On Kranzler and Jensen. *Intelligence, 17*, 15–16.
- Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods*. New York: Irvington.
- Humphreys, L. G. (1976). A factor model for research on intelligence and problem solving. In L. B. Rensick (Ed.), *The nature of intelligence* (pp. 329–340). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Humphreys, L. G. (1981). The primary mental ability. In M. P. Friedman, J. P. Das, & N. O'Connor (Eds.), *Intelligence and learning* (pp. 87–102). New York: Plenum.
- Humphreys, L. G., Davey, T. C., & Kashima, E. (1986). Experimental measures of cognitive privilege/deprivation and some of their correlates. *Intelligence, 10*, 355–376.
- Humphreys, L. G., & Parsons, C. K. (1979). A simplex process model for describing differences between cross-lagged correlations. *Psychological Bulletin, 86*, 325–334.
- Kelley, T. L. (1939). Mental factors of no importance. *Journal of Educational Psychology, 30*, 139–142.
- Kranzler, J. H., & Jensen, A. R. (1993). Psychometric *g* is still not unitary after eliminating supposed "impurities." *Intelligence, 17*, 11–14.
- Loevinger, J. (1951). Intelligence. In H. Helson (Ed.) *Theoretical foundations of psychology* (pp. 357–603). New York: Van Nostrand.
- MacCallum, R. C., Wegener, D. T., Ushino, B. N., & Fabrigar, L. R. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin, 114*, 185–199.
- MacCorquadale, K., & Meehl, P. E. (1948). On a distinction between hypothetical constructs and intervening variables. *Psychological Review, 55*, 95–107.
- Maier, M. H. (1993). *Military aptitude testing: The past fifty years*. Monterey, CA: Defense Manpower Center.
- Meehl, P. E. (1986). Trait language and behavior. In T. Thompson & M. Zeiler (Eds.), *Analysis and integration of behavioral units* (pp. 315–334). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Stich, T. G. (1992). Military testing and public policy: Selected studies of lower aptitude personnel. In B. R. Gifford & L. C. Wing (Eds.), *Test policy in defense: Lessons from the military for education training, and employment* (pp. 1–78). Boston: Kluwer.
- Vernon, P. A., & Mori, M. (1992). Intelligence, reaction times, and peripheral nerve conduction velocity. *Intelligence, 16*, 273–288.