An Alternative, Thurstonian View of Intelligence

John B. Carroll
University of North Carolina at Chapel Hill

Any reader of Humphreys's target article must entertain great respect for Humphreys's views, matured over years of research and thoughtful inquiry. For the most part, I find myself in agreement with his general approach. I feel the need, however, to raise certain questions about some of his assumptions and even about some of his conclusions. I have the impression that the concerns that I am about to express derive from inclinations that I have inherited from Thurstone, my chief role model and erstwhile mentor in the factor-analytic enterprise.

Humphreys starts by asserting that he approaches the study of intelligence as a "pragmatic behaviorist." Certainly intelligence, whatever it is, is manifested in behavior, but I like to consider it in terms of what in the individual is manifested in behavior. That is, it seems that there must be something in the individual that causes his or her behavior to be more intelligent or less intelligent than the behavior of others, just as we might postulate entities in the individual that cause his or her behavior to be more active, more motivated, more emotional, more introverted, or whatever, than the behavior of some others. For this reason, I question Humphreys's definition of *intelligence* as a "phenotypic behavioral trait." Actually, I'm not sure what Humphreys means by saying that intelligence is a phenotypic trait. With what is he contrasting this conception? I agree that intelligence is not a genotypic trait, because, even as Humphreys discusses it, it has both genetic and environmental substrates. Nevertheless, I am inclined to believe that intelligence is more than merely phenotypic—that, for example, one could symbolize it by one or more parameters that characterize whatever states of the individual there are that may be manifested in more intelligent or less intelligent behavior. In my view, intelligence measurement methodology seeks to establish such parameters, taking the form of the various "factors" of congitive ability that have been or can be established and the scores or measurements of such factors. It is not necessary to be concerned with the precise nature of the states in the individual that are represented by ability parameters.

Humphreys proceeds to amplify his definition of intelligence by asserting that it is "the acquired repertoire of all intellectual (cognitive) skills and knowledge available to the person at a particular point of time." Note that this definition implies that the actual repertoire of skills and knowledge available to the individual is in fact covert and therefore unobservable. It can only be sampled, by psychological tests or perhaps by raters, to provide estimates of its size. From this standpoint, intelligence is only indirectly observable, making it (I would say) less than completely phenotypic. I prefer the notion that one can postulate and estimate parameters associated with individuals that account for varied manifestations of intellectual behavior.

The very notion of intelligence as represented by a repertoire of skills and knowledge requires detailed specification and analysis of the nature and scope of that repertoire. Surprisingly, Humphreys suggests that this repertoire may be defined by "standard" tests such as the Stanford–Binet intelligence test or a Wechsler scale. But such tests only begin to cover the total repertoire of cognitive abilities (Carroll, 1993). For example, they fail to provide adequate measures of abilities in the domains of idea production, auditory perception, and cognitive speed—domains that many would agree are part of the intellectual repertoire. Beyond this, Humphreys admits that multiple factors (perhaps thousands; Humphreys, 1981, p. 88) could be included in the repertoire on the basis of his supposition that different tests (and factors) might be generated almost without limit by the combination of facets such as content, operation, product, difficulty, item format, and scoring method. I am unaware of any clear evidence that this might be the case. I prefer to believe—and the evidence we have thus far supports it—that there is a relatively small number of cognitive factors; the factorial enterprise is addressed to finding out exactly what these factors are. (I would explicitly disavow any evidence that variations in difficulty could generate factors; see Carroll, 1983.)

Certainly Humphreys is aware of and, in fact, has been one of the main proponents of (Humphreys, 1982) the hierarchical analysis of cognitive abilities. Such analysis has yielded clear evidence for a "general" factor at the top of the hierarchy, a small number of "broad" factors below it, and perhaps 60 or 70 "narrow" factors at the base (Carroll, 1993). (Note that the identification of just one general factor is an empirical result, as opposed to the supposition that there might be more than one such factor.) Would Humphreys grant that the "intellectual repertoire" includes all the factors revealed by hierarchical analysis? The article under scrutiny here is not completely clear on this point—Humphreys seems to disparage "the definition of large numbers of intellectual factors" by saying that it is "useful only as a way of characterizing the extent of the intellectual repertoire." Instead, Humphreys draws attention mainly to the general factor, insisting that it

contributes the major proportion of the variance of cognitive tests. In contrast, I have pointed out (Carroll, 1993, p. 57) that the general factor on the average contributes only a little more than half the common factor variance of a given test; thus, lower order factors can have almost as much importance as the general factor.

At many points, Humphreys appeals to classical test theory to justify his concept of intelligence. For example, he calls for homogeneity of the test items and for sufficient numbers of test items to achieve high reliability in the measurement of intelligence. The concept of test theory that he relies on is somewhat outdated or at least incomplete. What is needed is a thorough examination of cognitive ability measurement in terms of item response theory (IRT; Lord & Novick, 1968), expanded to include a theory of multiple factors underlying items, as required by hierarchical factor analysis, or, for that matter, by any multiple-factor model used in confirmatory factor analysis. IRT posits a distinct form of relation between abilities (factors) and test scores, embodied in the familiar three-parameter logistic equation for the probability of correct item response as a function of item difficulty, slope of the item discrimination function, and a "guessing" parameter. For the most part, IRT has been developed only on the assumption that a test measures one ability (or cluster of abilities); given the parameters for a set of items, IRT can predict, for example, the reliability of the test. In principle, an expanded IRT should be able to predict reliabilities, factor loadings, and other psychometric characteristics of tests that are assumed to measure more than one factor, such as a general factor, a second-stratum factor, and a first-order "primary" factor (as indicated by the factor loadings of a test on different levels of factors). Such an expanded IRT should make unnecessary much of what Humphreys says about the measurement of factors. More important, it should make clear how tests measure underlying abilities rather than merely indicating the size of particular repertoires. I prefer to think that tests do measure underlying abilities or "latent traits" rather than "repertoires."

In fact, it is surprising to me that Humphreys puts so much stock in tests as measures of "repertoires." Take the case of pitch discrimination ability, which I have discussed extensively (Carroll, 1983, 1988) and which might even be regarded as a cognitive ability. Pitch discrimination ability can be considered to be a repertoire only if a repertoire is considered to be all possible occasions on which an individual might have to discriminate between musical pitches. But what is critical—and what a pitch discrimination test attempts to assess—is the individual's threshold in terms of the size of pitch differences that the individual can discriminate. Similarly, many cognitive ability tests are designed to assess thresholds, in the sense of a level of difficulty that an individual can master with a desired probability. This would be true, for example, of a vocabulary test, which only indirectly assesses the size of vocabulary; rather, it attempts to estimate the level of word difficulty (measured in terms of word frequency or familiarity) the individual can deal with. The individual's vocabulary knowledge would be represented by the ability parameter (usually symbolized as b in IRT formulation) that would be estimated by the test.

There are other problems with the assertion that intelligence corresponds to the size of a repertoire. One is that the size of a repertoire must almost always be judged relative to age, experience, or opportunity to learn. Humphreys gives scant attention to this, seeming to downplay the use of such relative measures as IQ, which in my view comes close to being a parameter of general ability.

I believe, also, that Humphreys gives too much weight to the portrayal of the intellectual repertoire in terms of Guttman's (1954) radex model (Marshalek, Lohman, & Snow, 1983). In my view, the radex model is essentially a transformation of the hierarchical factor model and adds little, if anything, to the interpretation of factorial results beyond what can be ascertained from a hierarchical factor matrix. Thus, far, nobody has presented a radex model for a complete representation of all cognitive factors at all strata. I suspect that such a representation could not be made in terms of the type of two-dimensional plot thus far presented.

I would like also to protest Humphreys's rejection of the notion of intelligence as representing a capacity. He relies too heavily on the notion of intelligence as an accumulation of random gains, as suggested by Anderson (1939) and Roff (1941). Cronbach and Snow (1977) showed that the Anderson–Roff "overlap" model is untenable and that "true mental age does predict gain in subsequent years" (p. 146). From this result, mental age (or intelligence) can be regarded as indicating a capacity.

I do not dispute much of what Humphreys states in the latter portion of his target article, except to the extent that I believe that a more realistic, comprehensive, and scientific view of the operation and functioning of cognitive abilities can be had by assuming that abilities are real entities in the individual that can be represented by underlying parameters.

Note

John B. Carroll, 409 North Elliott Road, Chapel Hill, NC 27514-7628.

References

Anderson, J. E. (1939). The limitations of infant and preschool tests in the measurement of intelligence. *Journal of Psychology, 3,* 351–379.

Carroll, J. B. (1983). The difficulty of a test and its factor composition revisited. In H. Wainer & S. Messick (Eds.), *Principals of modern psychological measurement: A Festschrift in honor of Frederic M. Lord* (pp. 257–283). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Carroll, J. B. (1988). Individual differences in cognitive functioning. In R. C. Atkinson, R. J. Herrnstein, G. Lindzey, & R. D. Luce (Eds.), *Stevens' handbook of experimental psychology* (2nd ed., Vol. 2, pp. 813–862). New York: Wiley.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York: Cambridge University Press.

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions.* New York: Irvington.

Guttman, L. (1954). A new approach to factor analysis: The radex. In P. F. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences* (pp. 258–348, 430–433). Glencoe, IL: Free Press.

Humphreys, L. G. (1981). The primary mental ability. In M. F. Friedman, J. P. Das, & N. O'Connor (Eds.), *Intelligence and learning* (pp. 87–102). New York: Plenum.

Humphreys, L. G. (1982). The hierarchical factor model and general intelligence. In N. Hirschberg & L. G. Humphreys (Eds.), *Multivariate applications in the social sciences* (pp. 223–240). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.

Marshalek, B., Lohman, D. F., & Snow, R. E. (1983). The complexity continuum in the radex and hierarchical models of intelligence. *Intelligence, 7,* 107–127.

Roff, M. (1941). A statistical study of the development of intelligence test performance. *Journal of Psychology, 11,* 371–386.

Education, Achievement, and General Intelligence: What Ever Happened to the *Psycho* in *Psychometrics?*

Stephen J. Ceci
Department of Human Development and Family Studies
Cornell University

Lloyd Humphreys has long been acknowledged as a pioneer in the field of measurement theory and one of the most thoughtful (and thought-provoking) scholars in the study of individual differences in intelligence. Nothing in his target article or in my commentary will alter this assessment. He is once again blazing trails and offering cogent evidence in support of his arguments. I want to make these positive statements about Humphreys at the outset, lest they get lost amid the questions and quarrels that follow.

My quarrels are directed at the research tradition that Humphreys represents, rather than at him specifically. This tradition is, of course, the established psychometric approach. It is one that seems to have forsaken the *psycho* for the *metric.* I find less and less that is psychological about modern psychometrics. I say this while acknowledging the great strides that its proponents have made on thorny measurement issues. Although the *psycho* depends on the *metric* (i.e., reliability and standard errors of estimate are useful in establishing theoretical validities), the former is not the ineluctable result of the latter. A psychologically informed theory requires going beyond correlations; description and explanation can be fundamentally disjunctive enterprises. Just as correlations are useful to test a model, a model is useful in deciding which correlations to test. Increasingly, I find the models tested by psychometricians to be "wooden."

For Humphreys, intelligence is the repertoire of all intellectual skills and knowledge available to a person at a particular point in time. This is a reasonable start on the road to operationalization and explanation. The problem is that, much like the drunk who loses the keys while getting out of the car but searches for them under the street light because the illumination is better there, Humphreys looks in the wrong places for clues about the nature of the repertoire that he equates with general intelligence (*g*). Specifically, he looks at performance on well-fashioned test items from popular IQ tests and psychometric batteries. I argue that these instruments are impoverished measures of the contents of the repertoire and that there is reason for doubting that they truly assess what Humphreys presumes they collectively assess—namely, *g*. I also argue that this repertoire is indistinguishable from achievement, and there is no compelling evidence to prod one to accept the baggage that usually accompanies conceptions of *g*. Let me give some concrete examples of what I mean before arguing that *g* is more illusory than real, notwithstanding its robust psychometric foundations.

Heritability

Some of Humphreys's argument centers on the heritability of intelligence. He correctly comments that "increasing equality of opportunity ... increases heritability" and that, "if heritability ... is lower today than it was a generation ago, ... liberals have no cause for rejoicing." The basis for Humphreys's assertion, which may be counterintuitive to some, is that, by reducing environmental sources of variation in a population, one