J. Social Biol. Struct. 1984 7, 125–130

Jensen oversimplified: a reply to Sternberg

Arthur R. Jensen
Institute of Human Learning, University of California, Berkely, California 94720, USA

Introduction

The late Theodosius Dobzhansky, the famous geneticist, about a year before his death, told me something that struck me as a most interesting point. He said that it was his policy never to debate or argue about a scientific subject with anyone with whom he felt he was not in at least 90 per cent agreement about the relevant issues. This seemed to him to be an important precondition for an interesting and fruitful discussion of the points of disagreement. (I might add that Dobzhansky was eager to argue with me, then, about my book *Educability and Group Differences.*)

Hence, recalling Dobzhansky's words of wisdom, I was happy to learn from the editors of this journal that they had solicited a critical commentary on my article from Robert Sternberg. With some effort, I have kept up, over the years, with all of his prodigious publications, and I always felt we were mainly in agreement on almost all of the essential issues in our field of research on intelligence, allowing, of course, for a few personal differences in special interests and emphases in various topics. The nature of Sternberg's present assignment naturally imposes an emphasis on what he considers to be his points of disagreement with me. However, I believe that a good part of this apparent disagreement would tend to vanish if my own position were represented in somewhat less oversimplified terms. I feel that Sternberg has strained a bit to make it rather too easy for himself and almost anyone else, even me, to disagree with 'Jensen'. In general, cavalier dissent, without acknowledgement of the complex details in the points at issue (which admittedly is severely constrained by the allotted space) perhaps only best serves the understandable inclination of many critics and their audiences to distance themselves from the troublesome social implications of some of the findings I have presented. Therefore, I will take this opportunity to reiterate some of the fine-grain of my views, which, if taken into account, will be seen to moderate at least some of the points of seeming disagreement, and I may perhaps even sharpen some other points. Then we can have a clearer notion of the kinds of empirical evidence that we still need to gain, for the only legitimate basis for consensus in science is appeal to evidence and to reasoning based upon it.

Requests for reprints should be sent to Arthur R. Jensen, Institute of Human Learning, University of California, Berkeley, California 94720.

0140–1750/84/020125 + 06 $03.00/0

Simplification and oversimplification

Before taking up the substantive issues, I first must comment on Sternberg's complaints of simplicity and oversimplicity. According to my philosophy of science, one of our chief aims should be the analytic and conceptual simplification of the welter of raw observations of natural phenomena. Thus scientists have arrived at such simplifications as, for example, $S = 1/2gt^2$, $E = mc^2$, and $R = k \log S$. The fact that practically all such simplifications have boundary conditions, or limiting qualifications, does not make them scientifically invalid or useless, or warrant their being pejoratively labeled '*over*simplifications'. The simplicity I seek in my own research represents an attempt to discern a number of the most elemental and lawful, or dependable, relationships within the realm of such complex behavioral phenomena as individual differences in cognitive abilities. It seems to me to be a reasonable strategy for eventually understanding a highly complex phenomenon to begin by seeking functional relationships between relatively simple variables (such as reaction times) and the relatively complex (such as Spearman's *g*). I have never simply identified the one with the other. I doubt that scientific understanding is advanced by one's trying to consider every facet of a phenomenon simultaneously and wallowing in the subjective complexity of raw observations. A highly analytical and abstractive approach (which, in fact, Sternberg has energetically pursued in his own research) will lead to a knowledge of the elements or most fundamental relationships from which we can then reconstruct the full complexity of the phenomenon we had originally set out to study, but now in a way that permits us to understand it in a scientific sense. I find considerable appeal in Newton's famous motto, 'Nature is simple', which I take to mean that relatively simple conceptual relationships can be found within Nature's phenomenal complexity, and it is the scientist's job to discover these. The main thrust of the new experimental research on human intelligence, in fact, is essentially a search for simpler data with more elemental and analyzable facets and clear functional relationships than can ever be provided by the omnibus psychometric tests and factor analytic methods that have dominated the first 75 years of scientific interest in intelligence.

Mental speed and general intelligence, or *g*

(1) The general factor, *g*, which emerges from the factor analysis of virtually all complex cognitive tests, is a very commonly accepted working definition of intelligence. It is usually measured by reference tests which have especially large and clear-cut correlations (factor loadings) with this general factor. Just exactly what *g* consists of beyond this is currently the subject of much research, by myself, by Sternberg, and many others. As yet, psychologists have not arrived at either a comprehensively formulated or generally accepted theory of the nature of *g*. That possibility lies somewhere in the future. But it seems safe to say that the rate of progress of research toward this goal has markedly speeded up within the last few years.

(2) My own multiple-choice reaction time–movement time (RT–MT) task was never intended as a 'measure of intelligence', in the sense that a standardized test like the Standford–Binet or the Wechsler Scales are measures of intelligence, but was intended merely as a means for analytically investigating the nature of the *g* factor as measured by complex standardized tests, which themselves do not afford sufficiently simple facets for analysis in terms of elemental cognitive processes. The RT–MT task permits precision measurement of certain clear-cut (but not uncorrelated) facets common to a variety of elementary cognitive tasks which have been shown to be correlated with *g* and can be

regarded as reflecting certain elemental components of *g*. Speed of simple reaction (simple RT) and response execution (MT), the rate of increase in RT with the increase in the number or complexity of choice decisions, and the subject's variability or oscillation of RT from trial to trial under identical testing conditions are some of the main analyzable facets which have been found to be correlated with the *g* factor as measured by complex standard IQ tests. The speed factor in RT is important in all these facets, but, as I have clearly pointed out elsewhere (Jensen, 1979, 1980b, 1982a, b), the average differences in speed as a function of number of choice decisions, or complexity of the reaction stimulus in general, are also crucial variables in relation to *g*, as is intraindividual trial-to-trial variability in RT, which there is reason to believe is an even more fundamental variable than decision speed per se.

(3) The degree of correlation between RT (or speed of decision) and *g* represents a curvilinear or inverted U-shape function of task complexity, increasing with degree of complexity, or the amount of information to be processed—but only up to a point. Increases in task complexity beyond that point result in inconsistent or lower correlations between overall response speed and *g*. I have termed this the 'speed-complexity paradox' (Jensen, 1980b). This is most strikingly illustrated by the fact that individual differences in the average response latencies on relatively simple choice RT tasks show correlations with *g* ranging from about 0.30 to 0.60, whereas individual differences in the average latencies for complex test items, such as Raven's Progressive Matrices, show close to zero correlations with *g*, even though the number of items answered correctly on the Raven constitutes a very good measure of *g*. A reasonable explanation of this 'paradox' is that when the amount of information to be processed exceeds a certain limit, subjects are forced to process the information sequentially in parts, while retaining the already processed elements in working memory, to be retrieved later, and so on. The order or strategy for dealing with these various components of an informationally complex task, and the distribution of the times taken for each step, allow varying processing possibilities, or different cognitive strategies, in which there are found reliable individual differences, which may also be correlated with *g*, as Sternberg has noted. Such *g*-related differences in componential time distribution are evident even in such a simple form of information processing as my RT–MT task—a point I have discussed in considerable detail (Jensen, 1982a) in a book edited by Sternberg. Brighter subjects, in general, spend *relatively* more time on the stimulus encoding and choice-decision stage (i.e. RT) than on the response execution stage (i.e. MT), as compared with duller subjects.

Yet the overall importance of a general speed factor, even in very complex test items, becomes clearly visible when we look at the overall average solution latencies for items, in which the individual differences in processing strategies and componential time distributions are, in effect, averaged out. (That is, we are looking here at differences between the average response latencies of different items, not at individual differences between subjects.) If we rank order Raven matrices items in terms of difficulty, as indexed by the percentage of subjects who 'fail' the item when the test is taken without time limit, and if we also rank order the items in terms of the *average* response latencies (i.e. solution times) of each (correctly answered) item, we find a nearly perfect correlation between these two indices of item difficulty (percentage failing and average response latency).

So the speed factor does seem to be of basic importance, although its role in individual differences is often obscured by other factors in complex tasks. Brighter subjects process more information in solving a complex problem, like a difficult matrix item, than do duller subjects, who, facing the same problem, process less information in about the same amount of time. But the brighter subjects, because of their deeper, more comprehensive (and therefore more time-consuming) processing, achieve more correct solutions. The fundamental primacy of mental speed has been shown even in Sternberg's own relatively complex analogies tasks. When the time taken for each of the component processes in the analogies task are factor analyzed with psychometric reference tests of *g*, individual differences in the average time for all of them (i.e. the intercept or regression constant in Sternberg's analysis) shows a higher correlation with *g* than any of the single component latencies. To quote Sternberg's (1979a) own words:

> Information-processing analyses of a variety of tasks have revealed that the 'regression constant' is often the individual differences parameter most highly correlated with scores on general intelligence tests. This constant measures variation that is constant across all of the item or task manipulations that are analyzed via multiple regression. The regression constant seems to bear at least some parallels to the general factor. (p. 24)

Referring to the same point elsewhere, Sternberg (1979b) says this about the 'regression constant': '... we can feel pleased to be rediscovering Spearman's *g* in information processing terms'.

Importance of *g* in the real world

I have never claimed that ability and personality factors other than *g* are not correlated with scholastic achievement and job performance, or that predictive validity could not be enhanced by statistically significant increments by including other predictor variables in addition to *g*. I do claim, however, that, among all measurable psychological variables, *g* is the major predictor, accounting for much more of the criterion variance than any other single predictor variable independent of *g*. The predictive validity of *g*, of course, depends on the *g* demands of the criterion performance. In addition to scholastic achievement, there is significant *g* validity for the prediction of success in literally hundreds of different occupations, as shown by numerous studies conducted by the US Employment Service (United States Department of Labor, 1970).

Test bias and cultural differences

(1) My own review (Jensen, 1980a) of the evidence and conclusions on test bias in the United States are not in the least eccentric or out of tune with the consensus of the many psychometricians who have studied the matter. An investigation by a 19-member panel of experts sponsored by the National Research Council and the National Academy of Sciences has announced essentially the same conclusions (Wigdor & Garner, 1982) as those found in my book, *Bias in Mental Testing.*

(2) Predictive validity is only one of many psychometric methods for detecting test bias. Other methods, such as the congruence of factor structures of tests in different subpopulations and a whole class of various methods essentially based on groups × items interaction, have much stronger implications for the construct validity of tests as measures of *g* across different cultural and racial groups. A cultural-difference hypothesis of the test performance differences between whites and blacks, in particular, is rendered highly implausible by the virtual nonexistence of a group × item interaction, even in intelligence tests comprising extremely heterogeneous items, such as the Stanford-Binet and Wechsler batteries. No theory of cultural diffusion from the majority culture to the minority population can begin to account for why there should be such negligibly small group × item interaction while at the same time there is such a large 'main effect' for the race difference (at least one

standard deviation, equivalent to about 15 IQ points). A highly detailed and tightly argued analysis of this phenomenon–probably the single most telling of all the phenomena in the research on test bias–has been made by sociologist Robert Gordon (in press). His article merits careful study, as it has the most profound implications concerning the nature of the black deficit on intelligence tests.

(3) I have never ventured into what I would regard as true cross-cultural testing. The minute group × item interactions found on standard tests given to whites and blacks in the United States today are highly inconsistent with a theory of large cultural differences affecting test performance (see Gordon, in press). Yet large group × item (or subtest) interactions are quite easy to demonstrate in cultural groups that are patently different, such as Chinese immigrants or, in the case of vocabulary tests, even English children compared with the American standardization samples. Factor analytic studies show also that the *g* factor of such tests as the Wechsler is the identically same *g* for whites and blacks, although whites and blacks differ, on average, by more than one standard deviation on the *g* factor scores (Jensen & Reynolds, 1982). I find no hint of any evidence that American blacks possess a qualitatively different kind of intelligence, or *g*, than that of whites or Asians; the observed differences simply appear to be quantitative. I do not believe the difference can be adequately explained merely by exaggerating the supposed cultural differences between American blacks and whites. Blacks and whites share the same language, attend the same schools, watch the same TV programs, play the same games, go to the same movies, shop in the same supermarkets, work in the same industries, aspire to the same careers, and want the same things for themselves and their children. One strains in vain to find the great cultural differences that would be required to account for such a large disparity in test performance. Moreover, shouldn't we expect blacks of the 1980s to be more acculturated to the majority culture than were those in 1918, when the first large-scale testing was done? But the same 1 standard deviation difference on IQ tests still exists now as it did then. It hasn't changed in 65 years. I wonder if Sternberg could propose any type of test or task which would reflect his own conception of the construct of intelligence, yet would not show an average difference between typical samples of blacks and whites? But it should be realized that, as Lloyd Humphreys has repeatedly emphasized, the test differences themselves do not cause, but merely reflect, actual performance differences in the 'real world': 'The extent to which minorities are excluded from proportionate participation at all levels in our society is not the result of their lower average test performance. The basic deficit is their performance, on average, in education, industry, and the military' (Humphreys, 1983: 303). Blacks are not asking for recognition of a different *kind* of intelligence, with different consequences for educability and employability of this society. What they want is the same distribution of success rates as whites, in school, college, and the job market–success rates which are importantly related to the psychologists' construct of *g*, and this relationship is the same for blacks as for whites. That indeed is the real problem.

(4) I am interested mainly in that aspect of human intelligence which is not determined by what is peculiarly valued by any particular cultural group. I believe that human evolution has endowed all biologically normal *Homo sapiens* with essentially the same fundamental cognitive processes involved in *g*, although because culture obviously influences the contents and expression of intelligence, the tests or techniques for measuring individual differences in *g* must be culturally appropriate. From a biological, evolutionary standpoint, I do not believe that the essential construct of intelligence can be properly defined merely in terms of cultural values. The fact that running speed, visual acuity, hand–eye co-ordination, and spear-throwing skills are the most highly valued traits in a hunting culture, and are perhaps the most important for the survival of its members, does not qualify these characteristics

as a definition of intelligence. My working hypothesis is that one and the same set of *g*-processes can be found in members of a hunting culture as in any other human culture, although *g* may not be a very salient or valued trait in some cultures. Even Kalihari Bushmen, when shown a number of highly *g*-loaded performance tests of our Western variety, were able to pick out successfully those of their fellow tribesmen who would perform best on the tests, regardless of the Bushmen's opinion of the importance (or triviality) of the kind of mental ability reflected in the tests (Reuning, 1972). If we find that a test of English vocabulary in Great Britain and America is highly *g*-loaded, I suspect vocabulary will be similarly *g*-loaded in Tamil (in South India), in Urdu (in Pakistan), in Bantu (in South Africa), in Mandarin (in China), etc. But, of course, vocabulary tests in each of these languages would not by themselves permit any direct inference about the average differences in *g* capacity of the respective peoples. That problem has not yet been solved.

Dispensability of intelligence

(1) I have found no evidence that other traits, independent of *g*, can *substitute* for intelligence, when the intelligence is below some minimal threshold required for successful performance, a threshold which varies (in a probabilistic fashion) for different levels of education and occupation. Provided that an individual's intelligence exceeds this prerequisite threshold, other traits–special talents, drive, persistence, dependability, character, etc.–may, of course, significantly enhance the individual's chances of success. But *g* has some significant degree of predictive validity for efficiency of performance in virtually every type of job in our society above the level of the most unskilled labor.

(2) I doubt that anyone would argue with Sternberg's valid observation that outstandingly successful people in every field owe much of their 'outstandingness' to a nearly optimal deployment of some combination of a number of different traits–including at least average or superior intelligence. Neither I nor anyone else that I know of has claimed that *g* is the *only* source of variance in people's achievements. But who would claim that it is not one of the very important factors?

References

Gordon, R. A. (in press). In (C. R. Reynolds & R. T. Brown, Ed.) *Perspectives on Bias in Mental Testing*. New York: Plenum.

Humphreys, L. G. (1983). *Am. Sci.* **71**, 302–303.

Jensen, A. R. (1979). *Creative Sci. Tech.* **2**, 16–29.

Jensen, A. R. (1980a). *Bias in Mental Testing*. New York: Free Press.

Jensen, A. R. (1980b). Chronometric analysis of intelligence. *J. soc. biol. Struct.* **3**, 103–122.

Jensen, A. R. (1982a). In (R. J. Sternberg, Ed.) *Advances in the Psychology of Human Intelligence* (Vol. 1). Hillsdale, NJ: Erlbaum.

Jensen, A. R. (1982b). In (H. J. Eysenck, Ed.) *A Model for Intelligence*. New York: Springer-Verlag.

Jensen, A. R. & Reynolds, C. R. (1982). *Personal. Indiv. Diff.* **3**, 423–438.

Reuning, H. (1972). In (L. G. Cronbach & P. J. D. Dreuth, Eds.) *Mental Tests and Cultural Adaptation*. 171–181. The Hague: Mouton.

Sternberg, R. J. (1979a). *Components of Human Intelligence* (Technical Report No. 19). Office of Naval Research.

Sternberg, R. J. (1979b). In (R. J. Sternberg & D. K. Detterman, Eds.) *Human Intelligence: Perspectives on its Theory and Measurement*. Norwood, NJ: Ablex.

United States Department of Labor: Manpower Administration (1970). *Manual for the USES General Aptitude Test Battery*. Washington, D.C.: U.S. Employment Service.

Wigdor, A. K. & Garner, W. R. (Eds.) (1982). *Ability testing: Uses, consequences, and controversies. Part I: Report of the committee; Part 2: Documentation section*. Washington, D.C.: National Academy Press.