

Applying Double Standards to “Divisive” Ideas

Commentary on Hunt and Carlson (2007)

Linda S. Gottfredson

School of Education, University of Delaware

“Research on group differences in intelligence is scientifically valid and socially important” (Hunt & Carlson, 2007, this issue, p. 195). Hunt and Carlson are to be commended not only for standing up for the legitimacy of such research but also for reporting basic facts on the science that others outside the field often deny or distort.

Among the much-replicated empirical findings that Hunt and Carlson mention in their article are that IQ tests measure a general learning ability, predict many kinds of life success to some degree, measure cognitive ability equally well among American Blacks and Whites (i.e., there is no measurement bias), and predict academic achievement equally well in both groups (i.e., there is no prediction bias for that outcome). In addition, large racial gaps in cognitive abilities and achievements continue to create trade-offs among goals for schools and employers. The important scientific question is not whether races differ in (average) phenotypic intelligence but why they differ. Hunt and Carlson also mention that social environments are not just external but (like IQ differences among individuals) have both genetic and environmental components and that race exists as a biological entity or continuum. Such conclusions are mainstream among specialists on human variation in intelligence (Gottfredson, 1997; Neisser et al., 1996).

In regard to political intimidation against reporting such findings, Hunt and Carlson agree that such attacks occur, that they are deplorable, and that they have driven some investigators into professionally safer pursuits. Hunt and Carlson show how major social policies, such as the No Child Left Behind Act and U.S. employment discrimination law, can fail their aims and impose serious social and economic costs when they disregard such knowledge and presume a contrary reality. Scientific knowledge requires good evidence and inference; thus, the authors review various standards—their 10 principles of design, analysis, and reporting—for evaluating research articles (e.g., representativeness of samples, construct validity of measures,

alternative explanations of results). It is important to note that Hunt and Carlson also point to questions that science lacks the answers to but must keep asking (e.g., Are individual and group differences in intelligence malleable by either biological or socioeducational means?).

Hunt and Carlson support inquiry on group differences, in principle, but the thrust of their article is to hobble it in practice by holding it to stricter standards than other work. Thus, they inadvertently illustrate how politically unwelcome questions and answers are commonly suppressed in effect, if not by intent. I understand that they intend no such thing. Rather, as I shall illustrate, I believe they argue from faulty logical and empirical premises so ingrained in public discourse that even the most knowledgeable people often perpetuate them.

The authors’ line of argument shows a struggle to reconcile principle with contrary practice. Hunt and Carlson present previous findings on race and intelligence as scientifically valid and acknowledge that researchers who report them risk nonintellectual attack, but they then suggest that the research community has invited such attacks by tolerating substandard quality. Moreover, because research on group differences is akin to “working with dynamite,” it “is the duty of scientists to exercise a higher standard of scientific rigor in their research” (p. 195). But the voluntary self-monitoring system they offer results in more stringent standards for external review (“rules against dangerous play,” enforced by “referees,” p. 210).

Hunt and Carlson say their aim in recommending double standards is not to suppress research showing group differences but to “[reduce] the chances that an attack will have intellectual merit” (p. 210). To those who fear that their recommendation will invite yet more attacks without intellectual merit, “[o]ur reply is simple: you cannot do anything about an attack that is without intellectual merit” because such attacks are “political rather than scientific phenomena” (pp. 210). Hunt and Carlson agree that their “guidelines . . . may provide ammunition for those who wish to suppress studies of racial differences . . . [that do not report] equality of groups,” but “we see this once again as a political problem rather than a scientific one” (p. 210). That, however, is precisely the problem. In the name of science, they

Address correspondence to Linda S. Gottfredson, School of Education, University of Delaware, Newark, DE 19716, e-mail: gottfred@udel.edu.

invite selective political suppression against which there is no effective scientific defense.

Hunt and Carlson's proposal would legitimize what many scholarly journals have been doing surreptitiously for decades. Collectively, they have levied a stiff professional "tax" on scholars whose work on race or intelligence discomfits reviewers for nonscientific reasons. Editors have occasionally applied the "dangerous-idea criterion" openly. For example, Charles Kiesler, then editor of *American Psychologist* (and for many years the American Psychological Association's chief executive officer), explained that he rejected an article from Arthur Jensen that tested Spearman's Hypothesis¹ because, as editor, he should not accept anything "less than absolutely impeccable" when "this area is so controversial and important to our society." One problem, he said, was that the article left "a hanging implication" that the phenotypic differences Jensen had analyzed are genetic (C. Kiesler, personal communication, January 17, 1980). During his time as editor of *Psychological Science*, William Estes openly encouraged such editorial suppression when, in a postscript to a special section on ability testing (which included an article from Jensen), he advocated "developing an ethical code regarding the publication of research findings on group differences." According to Estes, the need for free and unfettered scientific exchange must be balanced against the need that no group in society feels threatened by such exchange (Estes, 1992).

But being of high quality does not protect research against media controversy. Indeed, it can provoke more fury by making unwelcome findings harder to dismiss on scientific grounds. The paradigmatic case is Arthur Jensen, the patriarch of modern research on intelligence as well as of its relation to race. Jensen is both one of the 50 "most eminent psychologists of the twentieth century" (Dittman, 2002, p. 29) and one of the most publicly vilified (hence the epithet, "Jensenism") because of his relentlessly objective, methodologically impeccable, and experimentally incisive investigations into human intelligence (Detterman, 1998).

The public vitriol and the impossible criteria heaped on such work are always excused by allegations that the work is potentially harmful (cf. M. Hunt, 1999, for other fields besieged by either the political left or the political right.) The threat to the social order posed by documenting phenotypic differences between races and by hypothesizing genotypic ones is treated as self-evident, often with dark references to slavery, intolerance

(e.g., Sternberg, 2003, pp. 386–387), and eugenics (Gardner, 1998, p. 23). I have yet to see anyone explicate exactly what hazards they pose and by what specific mechanisms they would cause injury or why letting "untruths" rule social policy (e.g., Glazer, 1994) is less destructive to the common good. Hunt and Carlson offer only a false analogy to justify burdening the unspecified views they label dangerous. Such ideas are like physical hazards—like working with dynamite or dangerous play in sports. It is therefore "simply being sensible" to impose special constraints upon them (p. 210).²

The authors insist that no scientific perspective should be silenced on nonscientific grounds, but the notion that certain scientific ideas are harmful encourages just that. First, partisans use the notion to impose an effectively insurmountable, nonscientific standard of proof on views they label dangerous—namely, that such ideas should not be seriously entertained until conclusively proved. Hunt and Carlson implicitly endorse this one-sided, beyond-all-possible-doubt standard: "We do not see any need for a potentially divisive 'default hypothesis'³ . . . in the absence of convincing evidence that rules out other hypotheses" (p. 210). This standard, often reflexively applied today, automatically renders a disfavored conclusion scientifically inferior to all competing ones and then burdens it further by giving critics license to generate an endless regress of doubts about it—regardless of where the preponderance of scientific evidence lies.

To illustrate, Nisbett (1998, 2005) is sometimes cited (e.g., Jencks & Phillips, 1998a, 1998b) as having discredited Rushton and Jensen's (2005) hereditarian hypothesis, which is that the Black–White phenotypic gap in IQ is in large part genetic (50–80%). Nisbett creates his illusory disproof when he first sweeps away 8 of their 10 independent bodies of evidence by labeling them "indirect." He then asserts that Rushton and Jensen "rode roughshod" over the "direct" evidence that he believes contradicted their hypotheses (Nisbett, 2005, p. 309). Hunt and Carlson (pp. 207–208) rightly note that the putatively damning set of studies that Nisbett cites to cast doubt on their hypothesis actually lack the ability to rule out any hypothesis at all, genetic or not, as both Jensen (1998) and Loehlin, Lindzey, and Spuhler (1975) had detailed years earlier. But Hunt and Carlson inadvertently sustain Nisbett's illusion by retracing his

¹Spearman's Hypothesis is that "variation in the size of the mean W-B [White–Black] difference across various mental tests is a positive function of variation in the tests' *g* loadings" (Jensen, 1998, p. 372, who cautions that the tests in the set must vary in *g* loading and be corrected for differences in reliability). Elsewhere in their article, Hunt and Carlson misstate Spearman's Hypothesis by expanding it into a genetic hypothesis, when it actually concerns only whether the White–Black gap in IQ represents a difference in phenotypic *g*. The support for the hypothesis does point to a genetic hypothesis, however, because other phenomena line up in the expected way along the same axis of test *g* loadings as do the mean White–Black score differences. These include the tests' heritabilities, degree of inbreeding depression they exhibit, their correlation with mean White–Black differences on elementary reaction time tasks, and the like (Jensen, 1998).

²The authors say that such constraints would apply "regardless of whether or not [studies] find for or against [group] differences" (p. 210), but only the former ever get branded as being dangerous. The latter, in contrast, tend to garner public accolades and professional honors (e.g., stereotype threat, which Hunt & Carlson used to exemplify violations of Principle 3; "Such changes cannot be used as evidence against group differences in intelligence," p. 205).

³They are referring here to Jensen's default hypothesis. "In brief, the default hypothesis states that the proximal causes of both individual differences and population differences in heritable psychological traits are essentially the same, and are continuous variables" (Jensen, 1998, p. 444). One corollary, the "hereditarian hypothesis" (Rushton & Jensen, 2005), is that because 50–80% of the IQ variance among whites is known to be genetic, 50–80% of the average White–Black difference will also be found to be genetic in origin. Hunt and Carlson mistakenly shrink the hypothesis to just this one corollary (pp. 200, 201).

missteps: They fail to mention that he acknowledges only one small sector of Rushton and Jensen's large network of evidence and then quote his false accusation that Rushton and Jensen "rode roughshod over [that part of] the evidence" (Nisbett, 2005, p. 309). The dangerous-idea criterion also allows partisans to demand new, more "direct" forms of evidence (e.g., molecular genetic differences by race involving the brain) that others will then quash for being "racist" or "dangerous" (e.g., the case of Bruce Lahn; Regalado, 2006).

The appropriate rule for scientifically adjudicating competing explanations is Carnap's Total Evidence Rule (Lubinski & Humphreys, 1997): Which theory accounts best for the totality of evidence and which is most consistent with the full pattern of the evidence to date? Scientifically successful explanations rest not on single studies (all of which have limitations) but on a dense nomological network of empirical evidence, ideally generated by diverse disciplines, methods, and theoretical perspectives—as has been the case for knowledge on abilities and achievements (e.g., Nyborg, 2003; Sternberg & Grigorenko, 1997). Some strands of evidence and inference will be stronger than others, but none are dissolved by being labeled as dangerous, divisive, or indirect. Nor is the whole body of evidence nullified by directing attention only to its weakest parts or to unfinished business, by emphasizing the "complications and ambiguities" of individual strands while neglecting the patterns they collectively weave, or by suggesting that it is unwise to draw causal inferences from existing patterns of evidence ("leaping . . . to . . . assertions" p. 201) until various new frontiers have been mapped (the "direct causal chains" for how genes affect brains, brains affect intelligence, etc.). These are, however, common ways of minimizing or generating doubt about a body of evidence without actually having to engage it and claiming victory without direct contest.

Science thrives on trenchant criticism that engages the evidence. Inevitably, some participants will introduce factual errors into debate or uncritically repeat others' false claims. However, the beyond-all-possible-doubt standard encourages and capitalizes on false claims about the burdened view. Even when they contradict each other, false claims cumulate into cascades of doubt about a view because they create doubt and confusion. The supposed trouncing of *The Bell Curve* (e.g., Fraser, 1995) is a case in point.

The following sample of Hunt and Carlson's own misstatements illustrate how small but consistent error can reinforce the prevailing public misperceptions that the intelligence-research community is still in a muddle about what intelligence is and that it lacks due diligence in looking for test bias and shrinking IQ gaps. Three of these examples are mistaken criticisms of core intelligence research, the fourth reflects uncritical acceptance of others' false claims against it, and the fifth is a conceptual error resulting in false praise for an unwarranted claim against IQ tests.

- "Evidently intelligence is at least as fuzzy a concept as race is!" (p. 199). This is false. The same word—fuzzy—is used to

stand in for two different ideas (i.e., races are "fuzzy sets" in the mathematical sense, and the public has jumbled views on the facts on intelligence). The effect is to connote a paucity of scientific knowledge on intelligence when, in fact, the intelligence community has intensively investigated and debated the latent constructs that IQ tests measure, especially their "massive central axis" (Loehlin et al., 1975, p. 258), called *g* (e.g., Carroll, 1993; Snow & Lohman, 1989; Sternberg & Grigorenko, 2002).

- "The lack of data on the [differential?] prediction of job performance presents a serious problem in the use of tests" (p. 203). This is false. There is no such gaping hole in the evidence. The question was for decades a major focus of personnel selection research in industry, military, and government settings (e.g., Campbell & Knapp, 2001; Schmidt, 1988). As noted by the latest edition of the *Principles for the Validation and Use of Personnel Selection Procedures* (Society for Industrial and Organizational Psychology, 2003, pp. 32–33), cognitive tests predict job performance equally well for different races in the U.S.
- Roth et al. (2001) "proceed to ignore [group] differences over time" (p. 206) in their large meta-analysis of Black–White gaps on selection tests. This is false. Roth, Bevier, Bobko, Switzer, and Tyler (2001, p. 323) had, in fact, detailed their attempt to investigate whether racial gaps had changed and why data limitations foiled the effort. Nor is it true that "[t]heir estimate for industrial applicants was based on data reported by Wonderlic & Wonderlic (1972) almost 30 years before Roth et al.'s date of publication" (p. 206). In fact, their Wonderlic data included normative samples from 1970, 1983, and 1992.
- "In both cultures, knowledge of important culture-specific activities—traditional medicine [in rural Kenya] and hunting practices [among the Yu'pik Inuit of Alaska]—were not correlated with measures of *g*" (p. 204). This is false. Here, Hunt and Carlson have uncritically accepted the original authors' rendition of results (Grigorenko et al., 2004; Sternberg et al., 2001). To take one example, measures of *g* did, in fact, correlate consistently and substantially with hunting knowledge when such knowledge was relevant to everyday life (in rural populations of Yu'pik but not in semiurban ones).⁴
- "What subset of abilities a test evaluates is determined by a social decision about what the society thinks is important and

⁴A second example is that the indigenous medicine test assessed the degree to which the Kenyan adolescents subscribed to traditional beliefs about the causes and cures of illness (e.g., the "evil eye" was a correct answer to the cause of some maladies). The test on indigenous practices actually correlated negatively (–.27) with Raven IQ, as Hunt and Carlson note, but it also correlated negatively with socioeconomic status and a vocabulary test. The scoring key and this pattern of results both suggest that the test reflects lack of knowledge about modern medicine. Hunt and Carlson use the two studies to imply what the original authors assert but never show (and cannot, given their study design)—that Western IQ tests are not equally predictive of equally important life skills ("practical intelligence") in different cultures.

what the purpose of the test is” (p. 199). This is false. An ability test’s construct validity (does the test really measure the construct claimed?) can never be presumed simply from its intended purpose or manifest content (American Educational Research Association, American Psychological Association, National Council on Measurement in Education, 1999). The authors thus wrongly praise Fagan and Holland (2002) for careful attention to construct validity by being “explicit about their view of intelligence as a concept” (p. 202). They also fail to note that the study’s claims gain credence only by violating their third principle (i.e., “Test scores may be changed by training, education, or, in some cases, by changing motivation; such changes cannot be used as evidence against group differences in intelligence unless the altered scores can be shown to be at least as predictive of criterion performance as the unaltered scores”, p. 205).

Finally, their exposition illustrates the distressingly common problem of failing to draw crucial distinctions which, if left muddled, cultivate confusion and suspicion.

- IQ (a measure) versus *g* (a construct, the primary latent trait that IQ tests actually measure) versus general intelligence (often used as a synonym for *g*) versus intelligence (a lay word with multiple meanings; an umbrella term in science for a wide range of cognitive abilities).
- Intelligence versus achievement: The “cognitive tests” on which racial gaps narrowed (p. 200) were tests of math and reading achievement, not of general intelligence (Gottfredson, 2005).
- Genotypic versus phenotypic differences: Spearman’s Hypothesis is not about genetic differences between groups (p. 200) but about phenotypic differences between groups.
- Biological versus genetic differences: The former can be environmental or genetic in origin (pp. 197, 208–209).

I appreciate Hunt and Carlson’s call for more intellectual rigor in research and debate, as well as their attempt to be even-handed. However, I offer different advice. First, if there is room to hold group-differences research to higher standards than those used in other research, then the latter lacks sufficient quality too. Journals should, therefore, hold all work to higher standards—with special attention to assessing reliability and validity of measures; reporting sample means, standard deviations, and zero-order correlations; testing competing hypotheses; and checking facts. Second, the major confusions, misperceptions, and fallacies that taint valid, socially important intelligence research are quite predictable, though not always obvious. We in the intelligence-research community, especially, should anticipate them when we write and speak. Attentive phrasing and preemptive clarifications serve this pedagogical aim. When we fail to recognize or address them, we tacitly capitulate to them.

REFERENCES

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: Author.
- Campbell, J.P., & Knapp, D. (Eds.). (2001). *Exploring the limits in personnel selection and classification*. Mahwah, NJ: Erlbaum.
- Carroll, J.B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge, United Kingdom: Cambridge University Press.
- Detterman, D.K. (Ed.). (1998). A king among men: Arthur Jensen [Special issue]. *Intelligence*, 26(3).
- Dittman, M. (2002, July/August). Study ranks top 20th century psychologists. *APA Monitor*, 33, 28–29.
- Estes, W. (1992). Ability testing: Postscript on ability tests, testing, and public policy. *Psychological Science*, 3, 278.
- Fagan, J.F., & Holland, C.R. (2002). Equal opportunity and racial differences in IQ. *Intelligence*, 30, 361–387.
- Fraser, S. (1995). *The Bell Curve wars*. New York: Basic Books.
- Gardner, H. (1998, Winter). A multiplicity of intelligences. *Scientific American Presents*, 9(4), 18–23.
- Glazer, N. (1994, October 31). The lying game. *The New Republic*, 211, 15–16.
- Gottfredson, L.S. (1997). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence*, 24, 13–23.
- Gottfredson, L.S. (2005). Implications of cognitive differences for schooling within diverse societies. In C.L. Frisby & C.R. Reynolds (Eds.), *Comprehensive handbook of multicultural school psychology* (pp. 517–554). New York: Wiley.
- Grigorenko, E.L., Meier, E., Lipka, J., Mohatt, G., Yanez, E., & Sternberg, R.J. (2004). Academic and practical intelligence: A case study of the Yup’ik in Alaska. *Learning and Individual Differences*, 14, 183–207.
- Hunt, E., & Carlson, J. (2007). Considerations relating to the study of group differences in intelligence. *Perspectives on Psychological Science*, 2, 194–213.
- Hunt, M. (1999). *The new know-nothings: The political foes of the scientific study of human nature*. New Brunswick, NJ: Transaction.
- Jencks, C., & Phillips, M. (1998a). The Black–White test score gap: An introduction. In C. Jencks & M. Phillips (Eds.), *The Black–White test score gap* (pp. 1–51). Washington, DC: Brookings.
- Jencks, C., & Phillips, M. (1998b, September/October). The heredity–environment controversy. *The American Prospect*, 19, 48.
- Jensen, A.R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.
- Loehlin, J.C., Lindzey, G., & Spuhler, J.N. (1975). *Race differences in intelligence*. San Francisco: W.H. Freeman.
- Lubinski, D., & Humphreys, L.G. (1997). Incorporating general intelligence into epidemiology and the social sciences. *Intelligence*, 24, 159–201.
- Neisser, U., Boodoo, G., Bouchard, T.J., Jr., Boykin, A.W., Brody, N., Ceci, S.J., et al. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, 51, 77–101.
- Nisbett, R.E. (1998). Race, genetics, and IQ. In C. Jencks & M. Phillips (Eds.), *The Black–White test score gap* (pp. 86–102). Washington, DC: Brookings.
- Nisbett, R.E. (2005). Heredity, environment, and race differences in IQ: A commentary on Rushton and Jensen (2005). *Psychology, Public Policy, and Law*, 11, 302–310.

- Nyborg, H. (Ed.). (2003). *The scientific study of general intelligence: Tribute to Arthur R. Jensen*. New York: Pergamon.
- Regalado, A. (2006, June 16). Head examined: Scientist's study of brain genes sparks racial backlash. *Wall Street Journal*, pp. A1, A12.
- Roth, P.L., Bevier, C.A., Bobko, P., Switzer, F.S., & Tyler, P. (2001). Ethnic group differences in cognitive ability in employment and educational settings: A meta-analysis. *Personnel Psychology*, *54*, 297–330.
- Rushton, J.P., & Jensen, A.R. (2005). Thirty years of research on race differences in cognitive ability. *Psychology, Public Policy, and Law*, *11*, 235–294.
- Schmidt, F.L. (1988). The problem of group differences in ability test scores in employment selection. *Journal of Vocational Behavior*, *33*, 272–292.
- Snow, R.E., & Lohman, D.F. (1989). Implications of cognitive psychology for educational measurement. In R.L. Linn (Eds.), *Educational measurement* (3rd ed., pp. 263–331). New York: Collier.
- Society for Industrial and Organizational Psychology. (2003). *Principles for the validation and use of personnel selection procedures* (4th ed.), Bowling Green, OH: Author.
- Sternberg, R.J. (2003). My house is a very very very fine house—but it is not the only house. In H. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen* (pp. 373–395). New York: Pergamon.
- Sternberg, R.J., & Grigorenko, E. (Eds.). (1997). *Intelligence, heredity, and environment*. Cambridge, United Kingdom: Cambridge University Press.
- Sternberg, R.J., & Grigorenko, E. (Eds.). (2002). *The general factor of intelligence: How general is it?* Mahwah, NJ: Erlbaum.
- Sternberg, R.J., Nokes, P., Geissler, P.W., Prince, R., Okatcha, F., Bundy, D.A., Grigorenko, E.L. (2001). The relationship between academic and practical intelligence: A case study in Kenya. *Intelligence*, *29*, 401–418.