Guest Comment: Are there innate cognitive gender differences? Some comments on the evidence in response to a letter from M. Levin

In a recent letter, Professor Michael Levin[1] states that "It is *not disputed* that males outperform females on tests of mathematical ability" [my emphasis] and further attributes this phenomenon to innate (i.e., biological) gender differences. In fact, the only element in his assertions that is not strongly disputed is that in *North America* the *average* of boys' scores exceeds that of girls' on the math SAT and similar standardized tests, such as the ACT. However, as discussed below, this disparity does not necessarily persist when other types of math tests are used, nor is it uniform across cultures. Moreover, no less prestigious a scientific body than the British Royal Society[2] (hardly a bastion of radical feminist theory) concluded after thorough study that there was *no convincing evidence* for innate gender difference in mathematical ability.

Because of the widespread belief that boys do outperform girls on most mathematics tests, it is worth emphasizing that the actual picture is far more complex. A number of studies[3–7] have shown that girls perform as well as boys on some types of mathematical tests. Two studies, one using an advanced high-school mathematics class[6] and another using entering freshmen at American University,[7] showed no significant gender differences on achievement tests despite lower math SAT scores for females in both groups. The results of over 100 studies were analyzed by Hyde *et al.*[5,8] who found that the math SAT produced a substantially larger gender difference than their "meta-analysis" of other mathematics tests. Moreover, the size of the gender gap varies with ethnic subgroup.[8,9] One study[8] reported that the gap is largest for Hispanics and smallest for Afro-Americans. Finally, it also worth emphasizing that, as Levin admits, gender differences in math scores do not generally emerge until the early teens.[5]

Levin buttresses his assertions by reference to the widely publicized work of Benbow and Stanley[10] based upon studies of "mathematically precocious youth" (SMPY) emanating from Johns Hopkins University. The journal *Behavioral and Brain Sciences* published a lengthy review[9] (by Benbow) of this work, accompanied by over 40 critiques and a rebuttal by Benbow. The assertion that their work is "... especially notable for controlling for the socialization variables...," which Levin makes in his letter, has been so widely debated[9,11,12] that it hardly merits further discussion. Let me only suggest that anyone inclined to take this claim seriously read the critiques accompanying Benbow's review. There are, however, some less publicized aspects of the Benbow–Stanley work worthy of comment.

A noteworthy omission from Benbow's report[9] is the fact that, *before* the young (typically, seventh grade) children take the exam, Stanley's center at Hopkins sends them a brochure[13,14] containing the information that boys outperform girls on the math SAT. In addition to the very real possibility that it could bias their results, this transgression represents such a serious and fundamental error in experimental design as to cast doubt upon the validity of their entire enterprise. In a related vein, Eccles and Jacobs[15] observed that widespread news reports of Benbow and Stanley's assertions have lowered parents' perceptions of their daughters' mathematics ability. Thus it appears likely that Benbow and Stanley may now be contributing to the very effect they purport to measure.

It should also be pointed out that Benbow and Stanley attach considerable importance to a rather nonstandard statistic, the male:female ratio for high scorers (who constitute much less than 1% of their sample). In discussing the question of whether the gender gap is decreasing in time, Benbow[9] asserted that this ratio "has remained relatively constant over...15 years [1972–86]" at about 12:1 for scores $\geqslant 700$. Something unusual must have happened in 1988; Hopkins' own data[14,16] give ratios[17] of about 4:1 for 1988, and 8:1 for 1989. Finally, it should be emphasized that, lower math SAT scores notwithstanding, the girls Benbow and Stanley tested in junior high subsequently outperformed boys by receiving *higher* grades in high-school math courses![9]

Although the SAT does not appear to have been systematically studied across cultures in other countries, the International Association for the Evaluation of Educational Achievement gave a different battery of tests to eighth-grade children in 20 different countries in 5 different subject areas of mathematics. These data were subsequently analyzed for gender differences by Hanna[3,4] who found that in some countries and subjects girls outperformed boys; in others, the reverse. In all cases, the *differences between countries were much larger* than the differences between the sexes. In three subjects (algebra, arithmetic, and statistics) there were *no significant overall sex differences*; however, there were small differences favoring boys in measurement and geometry. In view of the latter's possible connection to alleged differences in spatial ability, the results in this area were analyzed further. With a few exceptions, the difference between the sexes was *not* statistically significant in those countries with high geometry scores, whereas it was significant in those countries (including the US) with low scores. In fact, the average scores for American boys and girls were 39.7 and 37.9, respectively; the "gender gap" of 1.8 pales in comparison to the abyss between American students and those of top-scoring Hungary and Japan, in which all subgroups had averages in the 55–60 range. While the nature of the study does not allow any definitive rankings between countries, the size of the gaps clearly demonstrates that cultural factors and educational systems are far more important than gender.

Levin's assertion of gender differences in spatial ability might appear to be on firmer ground. However, careful examination of the evidence again gives a less convincing picture. In 1985, Caplan, *et al.*[18] surveyed the literature on the subject and found serious inconsistencies, a lack of conclusive evidence, and difficulties with the construct itself. Their description of one widely cited set of experiments is worth repeating.

> Porteus (1965) has reported that in 99 out of 105 studies males obtained higher test scores than females on his test in which examinees are asked to draw their way through line mazes. These figures are impressive, but a closer examination...reveals serious flaws.
>
> Porteus himself reported that in only 18 of the 105 studies were t tests done, and in only 4 of those did they reach statistical significance. Accordingly, he then questioned the reliability of the t test as a tool to detect significant differences. He reported no other statistical tests

from any of the 105 studies.

Caplan *et al.*[18] further report that, although a significant sex difference was thus demonstrated in only 4 out of 105 studies, these results have been widely cited as attesting to male superiority in spatial ability with some reviewers implying that significant differences were actually found in 99 of the studies.

It is also worth noting that the size of the gap varies with the testing procedures used.[18] Indeed, the types of tests used to measure spatial and verbal ability raise some questions. For example, *The New York Times* gave front-page publicity to a study which asserted that women's verbal and spatial abilities fluctuated with the level of estrogen in their bodies.[19] Verbal ability was reportedly tested by timing how fast the women could say "A box of mixed biscuits in a biscuit mixer" five times in succession.

Levin's suggestion that MIT's admissions policy reflects lower standards for women should not go unchallenged. I discussed his assertions with representatives of the admissions office and one member of the physics faculty. While it is true that the average math SAT scores of the women admitted to MIT are slightly lower than those of the men, all those I contacted emphasized that this did *not* result in a lowering of standards.[20] On the contrary, their policy results in a student body in which women perform, as measured by college grades, as well as men at MIT. Moreover, on average, women do as well as men across departments (including physics) so that their success cannot be attributed to any preference for "softer" courses as alleged by Levin.

In addition, almost all of MIT's students score above 700 on the math SAT; and the College Board's own guidelines imply that one cannot draw conclusions about the relative ability of students whose scores differ by as much as 70 points.[21] MIT's policy is also consistent with both the College Board's strong recommendation that SAT scores not be used as the *sole* basis for admission[22] and with a number of studies showing that other factors, such as high-school grades, may correlate better with college math performance than the SAT.[21,23] As noted above, Benbow and Stanley themselves report[9] that, lower math SAT scores notwithstanding, the girls in their SMPY study *subsequently* outperformed the boys in mathematics courses.

Levin distorts Professor Janice Button-Shafer's argument[24] when he suggests that she necessarily finds a 50/50 ratio for male and female physicists intrinsically optimal. Her thesis (amply supported by data from government agencies and professional societies)[25–28] is that the current percentage of women in physics is much lower than the percentages in mathematics and other areas of science and engineering which require the same type of skills and abilities as physics. Furthermore, a number of other countries, including Belgium, France, Israel, Spain, Poland, and China, have far more women physicists in high-level positions than the US.[29] In the US, even girls in advanced math classes are much less likely to study high-school physics than boys at the same level of mathematical ability; according to American Institute of Physics data,[25] 80% of such boys study physics, but only 60% of girls. The point is not that, in a perfect world, a 50/50 ratio would be either inevitable or desirable, but that many capable women do not pursue careers in the physical sciences, often making critical decisions at a rather young age.

I do agree with Levin that a complete absence of innate gender differences has also not been established, but find that of little importance. None of the tests measure large gender differences, and there is substantial evidence that at least some of those differences can be attributed to culture, education, and social factors. Therefore, any reasonable interpretation of the data gives evidence for, at most, a very small *average* gender difference in ability, yielding a substantial cohort of women quite capable of successful careers in science and engineering. Unfortunately, even these women may be discouraged by the publicity and distortions. Indeed, it is not uncommon for girls at the very top of the distribution (e.g., first in a math class) to be told that *they* cannot be scientists because girls (in general) aren't as good as boys. (My own experience was that such comments are most likely to come from nonscientists; however, other women scientists have reported differently.) Nor is there much evidence for Levin's assertion that parents who encourage sons in math more than daughters are reflecting their children's performance. On the contrary, Eccles *et al.*[30] have found that parents often deny their daughters' math ability even when they perform well, and that this phenomenon has been exacerbated by the publicity given to assertions of innate gender differences.[15]

It is worth observing that the same forces that discourage capable women from scientific careers often simultaneously encourage boys with mediocre talent. There is evidence that female students generally receive higher grades in calculus than male students. For example, Hughes[31] reported the results of a survey in which 31% of women vs 20% of men received *A*; 34% of women vs 27% of men received *B*; but only 15% of women vs 25% of men received *D* or *F*. Because fewer women choose to study calculus, such differences say little about relative ability, but they do say something sobering about the caliber of students pursuing various career paths. My own experience, which I suspect is typical, suggests that the picture is even more distorted than the data indicate. Women who receive calculus grades of C, or even low B, rarely pursue careers in science or engineering; however, I have frequently encountered men who intend to become engineers despite repeated calculus grades at the D or low C level. Thus it appears our society's propensity to encourage children on the basis of gender stereotypes rather than achievement may actually serve to *lower* standards.

Furthermore, there are several examples of high-quality educational environments in which males and females perform equally well.[32] This leads me to speculate that some of the gender gap observed in North America may be the result of deficiencies in our educational system, and to hope that improved math and science education would diminish the sex differential. Real reform will require an enormous investment in both personnel and resources, as well as changes in attitude. I believe such efforts are worth the price, and will reduce the gender gap; in any case, the worst that could happen is that we would have better male scientists.

As a physical scientist, accustomed to quantitative reasoning and objective, reproducible experiments, I found reading some of the literature on this subject, particularly the reviews by Benbow[9] and Caplan *et al.*,[18] to be almost surreal at times. Most of the respondents to Benbow's review[9] were identified as affiliated with psychology departments; a few with education or biosciences (e.g., neurophysiology); but not one from a department of mathematics, statistics, or physical science. Although many of these respondents gave very cogent critiques, I missed the voice of a statistician and regret the consequent lack of a serious critique about the reliability of inferences obtained from data in the tail of a curve.It is understandable that experiments in psychology and education do not meet the same standards of rigor and objectivity as those in a physics laboratory. However, that is no excuse for presenting speculation based upon dubious data to the news media and general public as if it were scientific fact.

It may be useful for physicists to compare the gender difference controversy with the recent suggestion of a "fifth force" or other modification to Newton's law of gravity.[33] In both cases, individual experiments, some of them carefully done, seem to provide strong support for a particular hypothesis. However, other experiments suggest the opposite. While physicists may not have been entirely satisfied with the coverage of the "fifth force" controversy in the news media, they did at least report the existence of contradictory data. By contrast, the news media frequently report speculative work alleging a gender difference as if it were scientific fact, but give scant attention to those who find otherwise.

It is unfortunate that the continued need to rebut assertions of sex-based differences in mathematical ability diverts attention away from related serious issues—namely, the need to find ways to counter the cultural factors that still deter women from studying the physical sciences, the need to substantially improve mathematics and science education for children of *both* sexes in the United States, and the need to find ways to encourage children of both sexes and all races to aspire to excellence and choose careers on the basis of interests and ability rather than sexual, ethnic, and racial stereotypes.

My own views on some of these matters have been expressed elsewhere.[34] I hope that other readers will accept Editor Romer's invitation to use this Journal as a forum for further discussion of how to meet this challenge.

I am grateful to many people, including Professor Richard Dudley, Professor Gila Hanna, Allyn Jackson, and Dr. Barbara Peskin, for helpful information, discussions, and comments on an early draft of this manuscript. Needless to say, both the opinions expressed here and the responsibility for the accuracy of the citations are entirely my own.

Mary Beth Ruskai
Department of Mathematics
University of Lowell
Lowell, MA 01854

[1]M. Levin, "Women—why so few?" Am. J. Phys. **58**, 905–906 (1990).
[2]D. Dickson, "Britain's Royal Society condemns sex bias in math teaching," Science **233**, 618–619 (1986).
[3]G. Hanna, "Mathematics achievement of girls and boys in grade eight: Results from 20 countries," Educ. Stud. Math. **20**, 225–232 (1989).
[4]G. Hanna, "Sex differences in geometry: Results from 20 countries," paper presented at the Conference on Learning and Teaching Geometry; reprinted in Assoc. Women Math. Newslett. **19**(4), 10–17 (1989) and **19**(5), 14–17 (1989).
[5]J. S. Hyde, E. Fennema, and S. J. Lamon, "Gender differences in mathematics performance: A meta-analysis," Psychol. Bull. **107**, 139–155 (1990); see also M. C. Linn and J. S. Hyde, "Gender, mathematics, and science," Educ. Res. **18**, 17–27 (1989).
[6]C. Norman, "Math education: A mixed picture," Science **241**, 408–409 (1988).
[7]K. R. Sheehan and M. W. Gray, "Gender bias in standardized mathematics tests," preprint.
[8]E. G. J. Moore and A. Wade Smith, "Sex and ethnic group differences in mathematics achievement," J. Res. Math. Educ. **18**, 25–36 (1987).
[9]C. P. Benbow, "Sex differences in mathematical reasoning abillty in intellectually talented preadolescents: Their nature, effects, and possible causes," Behav. Brain Sci. **11**, 169–232 (1988).
[10]C. P. Benbow and J. C. Stanley, "Sex differences in mathematical ability: Fact or artifact," Science **210**, 1262–1264 (1980).
[11]A. Bellisari, "Male superiority in mathematical aptitude: An artifact," Hum. Organ. **48**, 273–278 (1989).
[12]E. Fennema, "Explaining sex-related differences in mathematics: Theoretical models," Educ. Stud. Math. **16**, 303–320 (1985).
[13]P. Campbell, T. Kibler, and Kathryn Campbell-Kibler, "The SAT at twelve: A family's view of the Johns Hopkins talent search," Coll. Prep. (in press). Campbell reports that she was unable to find a reference to this aspect of the program in any of Benbow and Stanley's numerous publications describing their research.
[14]Center for Advancement of Academically Talented Youth, "Educational Planning Guide," Johns Hopkins University (1989). [This is the brochure sent to students before they take the SAT through the SMPY program. My copy was received courtesy of P. Campbell.]
[15]J. E. Jacobs and J. S. Eccles, "Gender differences in math ability: The impact of media reports on parents," Educ. Res. **14**, 20–25 (1985); J. S. Eccles and J. E. Jacobs, "Social forces shape math attitudes and performance," Signs **11**(2), 367–80 (1986).
[16]Center for the Advancement of Academically Talented Youth, "The 1989 Talent Search Report," Johns Hopkins University (1989). [This brochure was sent to students after they took the SAT through the SMPY program. My copy was received courtesy of P. Campbell.]
[17]I quote these numbers only to refute Benbow's assertion that this ratio is constant, and to demonstrate that it is simply not a reliable measure. It is hardly credible that these ratios accurately describe a gender difference which declined dramatically in 1988 and then doubled in 1989. In any case, the ratio was substantially less than 12:1 in both years. (Benbow also reported that this ratio is only 4:1 for Asian–American children.)
[18]P. J. Caplan, G. M. MacPherson, and P. Tobin, "Do sex-related differences in spatial abilities exist? A multilevel critique with new data," Am. Psychol. **40**, 786–799 (1985).
[19]S. Blakeslee, "Female sex hormone is tied to ability to perform tasks," *The New York Times* (18 November 1988), pp. A1 and D20; see also, "Sex hormones linked to task performance," Science **242**, 1509 (1988). As with many such reports in the news media, the report was based upon an oral presentation at a professional conference and no citation was given to a published paper in a refereed Journal; nor was sufficient other information presented to permit serious scientific scrutiny.
[20]Although MIT has not released any of their extensive data on these matters in a public report, the accuracy of this information was verified, before publication, by two of the representatives with whom I spoke.
[21]College Board, *ATP Guide for High Schools and Colleges 1990–91* (CEEB, New York, 1990).
[22]College Board, *Guidelines on the Uses of College Board Test Scores and Related Data* (CEEB, New York, 1988).
[23]*The College Board Technical Handbook for the Scholastic Aptitude Test and Achievement Tests,* edited by T. F. Dolan (CEEB, New York, 1984).
[24]J. Button-Shafer, "Guest Comment: Why so few women?" Am. J. Phys. **58**, 13–14 (1990).
[25]B. F. Porter, "Scientific resources for the 1990's: Women, the untapped pool," invited paper presented in the panel *Women in Physics: Why so Few?* organized by CSWP at the January 1989 joint APS/AAPT/AAAS meeting in San Francisco. (Dr.Porter is manager of the Education and Employment Statistics Division of the American Institute of Physics.)
[26]"Statistics on women mathematicians compiled by the AMS" Not. Am. Math. Soc. **37**, 946–947 (1990).
[27]National Science Foundation, "Achieving full participation of women in science and engineering" (1989 NSF report); "Women and minori-

ties in science and engineering" (1988 NSF report).

[28]National Research council, "Women: Their under-representation and career differentials in science and engineering" (1987).

[29]B. Wilson, "Women in physics: An international perspective," CSWP Gaz. **7**(2), 1–3 (1987).

[30]J. Eccles-Parsons, T. F. Adler, and C. M. Kaczala, "Socialization of achievement attitudes and beliefs: Parental influence," Child Dev. **53**, 310–321 (1982); D. Y. Yee and J. Eccles, "Parent perceptions and attributions for children's math achievement," Sex Roles **19**, 317–333 (1988).

[31]R. J. Hughes, "Calculus reform and women undergraduates," in *Calculus for a New Century*, edited by L. A. Steen (MAA, Washington, DC, 1987), pp. 125–129.

[32]L. Gilman, "Teaching programs that work," Focus **10**(1), 7–10 (1980); P. Rogers, "Thoughts on power and pedagogy," in *Gender and Mathematics: An International Perspective*, edited by L. Burton (Unesco, 1990); reprinted in Assoc. Women in Math. Newslett. **19**(4), 6–10 (1989).

[33]B. Schwarzschild, "From mine shafts to cliffs—the 'fifth force' remains elusive," Phys. Today **41**(7), 21–24 (1988).

[34]M. B. Ruskai, "How stereotypes about science affect the participation of women," invited paper presented in the panel *Women in Physics: Why so Few?* organized by CSWP at the January 1989 joint APS/AAPT/AAAS meeting in San Francisco. M. B. Ruskai, "Why women are discouraged from studying science," The Scientist (5 March 1990); reprinted in the June 1990 issue of the CSWP Gaz.

THE SCIENTIFIC METHOD: DOING ONE'S DAMNEDEST WITH ONE'S MIND, NO HOLDS BARRED

It seems to me that there is a good deal of ballyhoo about scientific method. I venture to think that the people who talk most about it are the people who do least about it. Scientific method is what working scientists do, not what other people or even they themselves may say about it.... I think that the objectives of all scientists have this in common—that they are all trying to get the correct answer to the particular problem in hand. This may be expressed in more pretentious language as the pursuit of truth.... All these things together give that "objectivity" to science which is often thought to be the essence of the scientific method. But to the working scientist himself all this appears obvious and trite. What appears to him as the essence of the situation is that he is not consciously following any prescribed course of action, but feels complete freedom to utilize any method or device which in the particular situation before him seems likely to yield the correct answer. In his attack on his specific problem he suffers no inhibitions of precedent or authority, but is completely free to adopt any course that his ingenuity is capable of suggesting to him. No one standing on the outside can predict what the individual scientist will do or what method he will follow. In short, science is what scientists do, and there are as many scientific methods as there are individual scientists.

Percy W. Bridgman, *Reflections of a Physicist* (Philosophical Library, New York, 1955), 2nd ed., pp. 81–83.