Why Aren't More Women in Science?

Top Researchers Debate the Evidence

Edited by
Stephen J. Ceci and
Wendy M. Williams

American Psychological Association • Washington, DC

1

WOMEN AT THE TOP IN SCIENCE— AND ELSEWHERE

VIRGINIA VALIAN

Why are there so few women in science, especially at the top? Hold on a minute. Is that the right question? That phrasing implies that science is different from other fields. Yet is it?

In one way, science does differ from other fields. A smaller percentage of women get advanced degrees in most of the natural sciences (although not biology) than in most of the social sciences, the humanities, medicine, law, business, or nursing. Yet in another way, science is the same as other professions: Women make less money and advance through the ranks more slowly not just in the natural sciences but in every field (e.g., see Valian, 1998, 2005a; for my gender tutorials, see Valian, 2005b), including nursing (Robinson & Mee, 2004).

The ubiquity of women's underrepresentation at the top provides important information about where to look to understand women's underrepresentation in science. There is a need both to look below the sur-

---

I thank Martin Chodorow, Howard Georgi, Cathy Kessel, and Mary Potter for comments, helpful suggestions, and references. This work was supported in part by an award from the National Science Foundation (SBE 0123609). Correspondence may be addressed to gender.tutorial@hunter.cuny.edu.

face of any particular field to understand how people are evaluated in professional settings and to understand which features of organizations give men more opportunities to be successful. There are two questions to answer. Why are there so few women at the top, even in fields like nursing and restaurant cooking, and why are there fewer women in most of the natural sciences than in other fields?

I provide the same explanation for both problems—a combination of gender schemas and the accumulation of advantage. Let us get two other possible reasons out of the way first.

ARE WOMEN LESS TALENTED THAN MEN?

One explanation for the lack of women at the top is that women are less talented than men, especially in the natural sciences and math. Three questionable assumptions appear to underlie that statement. One is that there is a single talent that determines success in the natural sciences and math; another is that existing standardized quantitative tests measure that single talent, and the third is that talents and abilities are fixed rather than malleable. All three assumptions—if accepted by young people or their teachers—are likely to reduce the number and range of people who will do creative and substantive work in the sciences. Only those with the highest standardized test scores and the most confidence would continue.

The physicist Howard Georgi, in a letter in January 2005 to Harvard undergraduates, noted that he had observed thousands of undergraduate physics students and went on to say the following:

> 1—Talent is not a unitary thing. It is multidimensional and difficult to measure or quantify precisely.
>
> 2—Many different kinds of talents are critical to the advancement of physics or any other science interesting enough to be worth doing.
>
> 3—The spread of talents within any group, sex, race, etc., is very large compared to any small average differences that may exist between such groups.
>
> 4—Talent can be developed and enhanced by education, encouragement, self-confidence, and hard work.
>
> For these reasons, I think that it is not particularly useful to talk about innate differences to explain the differences in representation of various groups in physics. Instead, I conclude that we need to try harder to teach our wonderful subject in a way that nourishes as many different skills as possible. (Georgi, 2005, ¶ 4–¶ 8)

Georgi's views mark him as an *incremental* theorist. Dweck and her colleagues (see Levy, Plaks, Hong, Chiu, & Dweck, 2001) have distinguished between people who see a given trait as fixed and unchanging—*entity* theorists—and those who see the trait as malleable and capable of increasing—

incremental theorists. Most of us are entity theorists about some traits and incremental theorists about others. For any given trait, about 85% of people fit neatly into the entity or incremental category, and that 85% is roughly equally divided between the two categories (Levy et al., 2001). If someone is an entity theorist about math and science skills and abilities, that person will treat those skills as stable and largely unresponsive to training and effort. In contrast, an incrementalist will see the skills as traits that can be developed and that can differ from one context to another. Thus, Georgi is an incremental theorist about physics.

There are consequences of holding an entity or incremental view (Levy et al., 2001). Entity theorists see groups as internally more homogeneous than incremental theorists do. For example, an entity theorist would see physicists as a more homogeneous group than an incremental theorist would. Enrico Fermi, who won the Nobel Prize in physics in 1938, was reportedly asked whether the prize winners in physics had any characteristics in common. After some thought, he replied that he could not think of a single one, including intelligence (Shucking, 1994).[1] Fermi, then, was not an entity theorist about physics. Entity theorists also see the differences between groups as larger than incremental theorists see them. Finally, entity theorists are more prone than incremental theorists to see the traits of a group, such as scientists, as being due to innate characteristics; incremental theorists allow more room for experiences and environment (Levy et al., 2001).

Which group is right about "talent" in math and physics? The U.S. educational system as a whole acts like an entity theorist about math, whereas Japan's educational system acts like an incrementalist. Japanese educators see math as a set of skills that can be taught well or badly and that require effort on a student's part. An international report of the 2002–2003 Trends in International Mathematics and Science Study compare how eighth graders in different countries perform on math tests. The overall average score (combining the domains of knowing, applying, and reasoning, with reasoning as the most sophisticated form of performance) of Japanese girls was 569 and that of Japanese boys was 571 (Mullis, Martin, & Foy, 2005). In contrast, the comparable figures for the United States were 502 for girls and 507 for boys. In both countries there is a small sex difference favoring boys—2–5 points. It is important to note, though, that the Japanese girls outperform U.S. boys by 62 points. Japan has been outperforming the United States since international testing began in the mid-1960s. Similarly, girls and boys from Singapore, with average scores of 611 and 601, respectively, performed even better—a full standard deviation better than American children. The cross-national differences dwarf the sex differences. If high test scores were the main determinant of mathematical discovery, Asians—male and female—

---

[1] I am grateful to Dudley Herschbach, cowinner in 1986 of the Nobel Prize in chemistry, for alerting me to this quotation (personal communication, July 8, 2005).

would dominate mathematics (to the best of my knowledge, however, they do not).

In the case of the most sophisticated performance—reasoning—there were, again, very small cross-sex differences and very large cross-national differences. Further, when sex differences existed, they favored girls (Mullis et al., 2005). That is contrary to earlier research, which had suggested that girls did as well as boys at standard problems but lagged behind boys with problems that required unconventional solutions. The cross-national differences were stunning. A score of 446 placed a Japanese student at only the 5th percentile in the domain of mathematical reasoning in his country; a similar score, 448, placed a U.S. student at the 25th percentile. A Singaporean student at the 75th percentile in her country, with a score of 645, would be slightly above a child at the 95th percentile in the United States, with a score of 638. Mathematical competence can be nurtured.

Let us look now at what test scores tell us about who gets advanced degrees in science. In the United States (and most but not all other nations where scores are available), girls consistently have less variable distributions than boys and thus are less likely than boys to score at the top end in quantitative tests (for summary data, see Hedges & Nowell, 1995; Lubinski & Benbow, 1992). It is not known why that is the case (for reviews, see Halpern & LaMay, 2000; Valian, 1998), but we do know that the difference at the top end is decreasing, supporting an incrementalist perspective. In 1983, for example, seventh- and eighth-graders in a national sample of gifted children showed a large sex difference between children who scored 700 or above on the math SAT. There were 13 boys for every girl (Benbow & Stanley, 1983). By 2005 that difference had plummeted to 3 to 1 (Brody & Mills, 2005) or 4:1 (Benbow, personal communication, July 6, 2006). Such a striking change is incompatible with the idea of a fixed difference between boys and girls.

In addition, the differences do not predict U.S. youngsters' intentions to major in science in college (Xie & Shauman, 2003). Neither average sex differences on standardized math and science achievement tests taken in the 8th, 10th, and 12th grades nor sex differences among the top 5% of test takers account for the sex differences in students' intended college majors (Xie & Shauman, 2003). Twice as many boys as girls intend to go into science, but that sex difference is not explained by sex differences in test scores. If anything, taking scores into account exacerbates the sex difference. Top-scoring girls are a particular casualty of science and math education in the United States—not a desirable result; we do not succeed in nurturing girls' talent.

A different kind of measure focuses on the educational outcomes for young adolescents who demonstrate early high performance on the quantitative SAT (Benbow, Lubinski, Shea, & Eftekhari-Sanjani, 2000). Two groups of 12- to 14-year-olds who scored in the top 1% of their age group in the 1970s were surveyed 20 years later at age 33; in the initial group, boys outnumbered girls by a little more than 2:1. Let us look at how the percentages of males and females who achieved undergraduate degrees in various fields compared with the percentages who achieved PhDs in those same fields. The natural sciences and math lost large numbers of high-scoring males and an even larger proportion of high-scoring females. The higher education system in the United States wastes female talent.

Take mathematics as an example. The same percentage (about 10%) of high-scoring boys and girls received undergraduate degrees in math. Because there were a little more than twice as many boys as girls among the high scorers to begin with, about 42 boys compared with 18 girls earned a bachelor's degree in math. At the PhD level, the comparison is considerably worse: About 9 boys and 1 girl ended up with a PhD in math. An initial 2.3:1 difference in BAs becomes a 9:1 difference in PhDs (calculated from Benbow et al., 2000, Table 1, Cohort 1). High-scoring girls are not being retained in math and science at the same rate as high-scoring boys. Thus, whether we look at population data (Xie & Shauman, 2003) or top scorers (Benbow et al., 2000), we see the same picture. The U.S. education system does not retain girls at the same rate as boys. Furthermore, there is greater attrition of both boys and girls in math and the natural sciences than in other fields.

In summary, test performance cannot explain the low representation of women in math and natural science.

Are Women Less Interested in Professional Careers Than Men Are?

Another reason for women's underrepresentation in the sciences may be that women are less interested in a professional career—in science or any other field—than men are. Again, there are unspoken assumptions in the phrasing of the hypothesis. One is that people make their choices in an unconstrained manner and are unaffected by the support and encouragement they receive or fail to receive. Another is that having a high-powered professional life and having a rich personal life are incompatible; you have to pick one or the other. A third is that anyone who is talented and works hard will be successful, regardless of their sex, race, age, and so on.

Because people are affected by encouragement, support, and expectations (see Valian, 1998, for a review of relevant experiments), it is difficult to evaluate people's choices. What would men do if they, like women, were expected and encouraged to take care of children? What would women do if they, like men, were expected and encouraged to have a professional life?

One reason to question the sufficiency of an explanation that emphasizes women's interest in a family is that women pay a price in rate of advancement for being women, even if they do not have children. Women without children do not advance as fast as men (for academic science, see Long, 2001; for law, see Wood, Corcoran, & Courant, 1993). Thus, although fathers do less and mothers do more than their fair share of child care, that alone does not account for the fact that women without children are less successful than men. There is a professional cost to women and to society of women's having children—absence from full-time employment (Long, 2001; Xie & Shauman, 2003). But women with children who remain as full-time academics publish the same amount as women without children (Long, 2001; Valian, 1998). They also have careers that are very similar to those of childless women. In summary, women's interest in a full life cannot explain their sparse numbers in science or at the top of different professions.

GENDER SCHEMAS AND ACCUMULATION OF ADVANTAGE

Why, then, are women underrepresented in science and underrepresented at the top in all professions? This is where gender schemas and the accumulation of advantage come in. *Schemas* are hypotheses that are used to interpret social events (Fiske & Taylor, 1991). Schemas are similar to stereotypes, but the term *schema* is more inclusive, more neutral, and more appropriate because it brings out the protoscientific nature of social hypotheses. Social schemas are necessary. One cannot treat every person one meets as if the social group to which they belong is irrelevant; it often is relevant and provides valuable information. But schemas—being schematic—oversimplify and thus can lead to mistakes.

Gender schemas are hypotheses about what it means to be male or female, hypotheses that all people share, male and female alike. Schemas assign different psychological traits to males and females (Martin & Halverson, 1987; Spence & Helmreich, 1978; Spence & Sawin, 1985). As folk psychologists, we think of males as capable of independent action, as oriented to the task at hand, and as doing things for a reason. We see females as nurturant, expressive, and behaving communally. In brief, men act; women feel and express their feelings. And our beliefs have support. In questionnaires, men endorse more "instrumental" characteristics and women endorse more "expressive" characteristics. The sexes overlap, as they do on every measure of behavior, perception, and cognition; but there are broad differences.

The main answer to the question of why there are not more women at the top is that gender schemas skew our perceptions and evaluations of men and women, causing us to overrate men and underrate women. Gender schemas affect judgments of people's competence, ability, and worth.

Consider two experiments on judgments of women's competence. Both were published in 2004 and demonstrate some of the effects of gender schemas. The first investigated how males and females rated people who were described as being "assistant vice presidents" in an aircraft company (Heilman, Wallen, Fuchs, & Tamkins, 2004). The evaluators read background information about each person, the job, and the company. In half of the cases, the person was described as about to have a performance review; thus, evaluators did not know how well the person was doing in the job. In the other half of the cases, the person was described as having been a stellar performer. The evaluators' job was to rate how competent the employees were and how likeable they were.

When no information was given about how well people were doing in the job, evaluators rated the man as more competent than the woman and rated them as equally likeable. When the background information made clear that the woman was extremely competent, however, the ratings changed. Evaluators now rated the man and the woman as equally competent, but they rated the woman as much less likeable than the man. They also perceived the woman as considerably more hostile than the man.

Thus, in evaluating a woman in a male-dominated field, observers saw her as less competent than a similarly described man unless there was clear information that she was competent. In that case, they saw her as less likeable than a comparable man. Notably, as is the case in almost all such experiments, there were no differences between male and female raters. Both male and female raters saw competence as the norm for men and as something that has to be demonstrated unequivocally for women. Both male and female raters saw competent men as likeable. Neither male nor female raters saw competent women in male-dominated positions as likeable.

Does likeability matter? In a follow-up experiment, the experimenters described targets as high or low in competence and as high or low in likeability. People rated the targets who were high in likeability as better candidates for being placed on a fast track and as better candidates for a highly prestigious upper level position. One cannot tell women just to be competent, because likeability can make the difference in whether people get rewards. Again, there were no differences between male and female raters.

The second study demonstrated how people shift their standards to justify a choice that seems a priori reasonable to them (Norton, Vandello, & Darley, 2004). In this experiment, gender schemas determined what seemed reasonable. The experiments asked male undergraduates to select a candidate for a job that required both a strong engineering background and experience in the construction industry. The evaluators rated five people, only two of whose resumes were competitive. One candidate had more education—both an engineering degree and certification from a concrete masonry association—than the other, who only had an engineering degree. The other candidate had more experience—9 years—than the first, who only had 5 years.

In the control condition, the candidates were identified only by initials. Here, the evaluators chose the candidate with more education three fourths of the time, and education was most often cited as the most important determinant of their decision. In one of the experimental conditions, a male name was given to the resume that had more education and a female name to the resume that had more experience. Here, too, evaluators chose the candidate with more education three fourths of the time and also rated education as very important. In the second experimental condition, a female name was given to the resume with more education and a male name to the resume with more experience. In this condition, less than half the evaluators picked the person with more education and less than a quarter said that education was the most important characteristic.

Men look more appropriate than women for the job of construction engineer, whether they have more education or more experience. The standards by which people judge others will shift depending on a priori judgments about their goodness of fit. Gender schemas help determine goodness of fit. Shifting standards are in operation at work, when job candidates are evaluated, and at home, when heterosexual couples make decisions about whose profession is more important. Thus, even though most people are genuinely meritocratic and egalitarian, and even though people's estimates of sex differences correlate well with psychological measures of sex differences (Swim, 1994), people's implicit evaluations of performances are in tune with gender schemas.

These experiments and others like them are relevant to women's underrepresentation in most natural science fields and underrepresentation at the top in every field. To take the second issue first, the experimental data show that both men and women slightly overrate men and underrate women in professional domains. Women appear to both men and women to be less competent. The small imbalances in evaluation and perception add up to advantage men and disadvantage women. It is like interest on an investment. If X has a slightly higher interest rate than Y, X will end up with more money than Y down the line, thanks to the "miracle" of compound interest.

A similar "miracle" happens in the professional world. Success is largely the accumulation of advantage, parlaying small gains into bigger ones (Merton, 1968). If you do not receive your fair share of small gains because of the social group you belong to, you—and your group—will be at a disadvantage. The miracle of compound interest means that someone who receives an extra quarter percent of interest than you will be in better shape than you 10 years later.

A computer simulation (Martell, Lane, & Emrich, 1996) showed the importance of even tiny amounts of bias. The researchers simulated an eight-level hierarchical institution with a pyramidal structure. They staffed this hypothetical institution with equal numbers of men and women at each level. The model assumed a tiny bias in favor of promoting men, a bias accounting for only 1% of the variability in promotion. After many series, the top level was 65% male. Even very small amounts of disadvantage accumulated.

Thus, even in a work environment in which everyone intends to be fair—and believes they are being fair—men are likely to receive advantages in evaluations that women do not. Over time, those advantages mount up, so that men reach the top faster and in greater numbers than women do. Each individual event in which a woman does not get her due—is not listened to, is not invited to give a presentation, is not credited with an idea—is a molehill. Well-meaning observers may tell the woman not to make a mountain out of a molehill. What they do not understand is what the notion of the accumulation of advantage encapsulates. Mountains are molehills, piled one on top of the other.

Because gender schemas and the accumulation of advantage operate in all the professions, men in general will have an easier time than women getting to the top in all the professions. The role of gender schemas can be extended to the natural sciences at all levels. The schema of a natural scientist is more compatible with the schema for men than the schema for women. Consider again the specifics of the schemas. People see males as capable of independent action, as oriented to the task at hand, and as doing things for a reason. They see females as nurturant, expressive, and behaving communally. The natural sciences seem particularly compatible with the schema for males and out of keeping with the schema for females.

Where do gender schemas come from? I have proposed that, like race schemas (Hirschfeld, 1997), they are cognitive in origin rather than motivational or emotional (Valian, 1998). Humans create categories, and the ability to do that is an important step in scientific reasoning. In addition, I propose, humans aim first for binary categories that have nonoverlapping characteristics. We are of course capable of creating more categories and of noticing overlap, but on grounds of simplicity, we prefer to create two categories and to have those two be as distinct as possible. A two-category system is fast and efficient, even if it sometimes leads to error. In the case of sex, I suggest that the two categories are based on observation of a qualitative physical difference between the sexes. Females give birth and physically nurture their young. People draw an analogy from the physical to the mental and see females not only as physically nurturing but as metaphorically nurturing. The next step is to add other traits that seem highly compatible with nurturance to the schema for females and to construct an "alternative" set of characteristics for men (Parsons & Bales, 1955). Although the reasoning behind gender and race schemas makes sense, schemas lead to errors in evaluation. Worse, they can lead to the development of behaviors and traits that appear to show the validity of the initial distinction.

Many false beliefs are eventually dispelled by education. A flat earth looks like a natural hypothesis, but education provides data and theory to show that the earth is round. There is a limit to how far astray one can go with a belief about a flat earth even if one keeps it forever; one can never make the earth flat. In the psychological domain, however, one can make the data fit the theory. One can discourage females from high professional achievement and discourage males from nurturing parenthood. One can create limits where none intrinsically exist. The aim of this essay is to present the alternative—to use scientific experiments and reasoning to erase arbitrary limits.

REFERENCES

Benbow, C. P., Lubinski, D. S., Shea, D. L., & Eftekhari-Sanjani, H. (2000). Sex differences in mathematical reasoning ability at age 13: Their status 20 years later. *Psychological Science, 11*, 474–480.

Benbow, C. P., & Stanley, J. C. (1983, December 2). Sex differences in mathematical reasoning ability: More facts. *Science, 222*, 1029–1030.

Brody, L. E., & Mills, C. J. (2005). Talent search research: What have we learned? *High Ability Studies, 16*, 97–111.

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition* (2nd ed.). New York: McGraw-Hill.

Georgi, H. (2005, January 21). *Talent, skills in math and science hard to quantify*. Retrieved November 26, 2005, from http://www.thecrimson.com/article.aspx?ref=505377

Halpern, D. F., & LaMay, M. L. (2000). The smarter sex: A critical review of sex differences in intelligence. *Educational Psychology Review, 12*, 229–246.

Hedges, L. V., & Nowell, A. (1995, July 7). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science, 269*, 41–45.

Heilman, M. E., Wallen, A. S., Fuchs, D., & Tamkins, M. M. (2004). Penalties for success: Reactions to women who succeed at male gender-typed tasks. *Journal of Applied Psychology, 89*, 416–427.

Hirschfeld, L. (1997). The conceptual politics of race: Lessons from our children. *Ethos, 25*, 63–92.

Levy, S. R., Plaks, J. E., Hong, Y., Chiu, C., & Dweck, C. S. (2001). Static versus dynamic theories and the perception of groups: Different routes to different destinations. *Personality and Social Psychology Review, 5*, 156–168.

Long, J. S. (Ed.). (2001). *From scarcity to visibility: Gender differences in the careers of doctoral scientists and engineers*. Washington, DC: National Academy Press.

Lubinski, D. S., & Benbow, C. P. (1992). Gender differences in abilities and preferences among the gifted: Implications for the math–science pipeline. *Current Directions in Psychological Science, 1*, 61–66.

Martell, R. F., Lane, D. M., & Emrich, C. (1996). Male–female differences: A computer simulation. *American Psychologist, 51*, 157–158.

Martin, C. L., & Halverson, C. (1987). The roles of cognition in sex role acquisition. In D. B. Carter (Ed.), *Current conceptions of sex roles and sex typing: Theory and research* (pp. 123–137). New York: Praeger.

Merton, R. K. (1968, January 15). The Matthew Effect in science. *Science, 159*, 56–63.

Mullis, I. V. S., Martin, M. O., & Foy, P. (2005). *TIMSS 2003 international report on achievement in the mathematics cognitive domains: Findings from a developmental project*. Retrieved November 24, 2005, from http://timss.bc.edu/PDF/t03_download/T03MCOGDRPT.pdf

Norton, M. I., Vandello, J. A., & Darley, J. M. (2004). Casuistry and social category bias. *Journal of Personality and Social Psychology, 87*, 817–831.

Parsons, T., & Bales, R. (Eds.). (1955). *Family, socialization, and interaction process.* New York: Free Press.

Robinson, E. S., & Mee, C. L. (2004). Salary survey. *Nursing, 34*, 36–39.

Shucking, E. L. (1994). Review of Einstein lived here. *Physics Today, 47*, 70–71.

Spence, J. T., & Helmreich, R. L. (1978). *Masculinity and femininity: Their psychological dimensions, correlates, and antecedents.* Austin: University of Texas Press.

Spence, J. T., & Sawin, L. L. (1985). Images of masculinity and femininity: A reconceptualization. In V. E. O'Leary, R. K. Unger, & B. S. Wallston (Eds.), *Women, gender, and social psychology* (pp. 35–66). Hillsdale, NJ: Erlbaum.

Swim, J. K. (1994). Perceived versus meta-analytic effect sizes: An assessment of the accuracy of gender stereotypes. *Journal of Personality and Social Psychology, 66*, 21–36.

Valian, V. (1998). *Why so slow? The advancement of women.* Cambridge, MA: MIT Press.

Valian, V. (2005a). *Sex disparities in advancement and income.* Retrieved November 24, 2005, from http://www.hunter.cuny.edu/genderequity/equityMaterials/numbers.pdf

Valian, V. (2005b). *Tutorials for change: Gender schemas and science careers.* Retrieved November 24, 2005, from http://www.hunter.cuny.edu/gendertutorial/

Wood, R., Corcoran, M., & Courant, P. (1993). Pay differences among the highly paid: The male–female earnings gap in lawyers' salaries. *Journal of Labor Economics, 11*, 417–441.

Xie, Y., & Shauman, K. (2003). *Women in science: Career processes and outcomes.* Cambridge, MA: Harvard University Press.