Spring 2011

Gender And The Evaluation Of Physicists

Amy Lisa Graves
Swarthmore College, abug1@swarthmore.edu

E. Hoshino-Browne

K. Lui

Unconscious Gender Bias in the Classroom

By Amy Bug, Etsuko Hoshino-Browne and Kris Lui

Why should gender be an issue in physics? Philosopher of science Fox Keller asked this more than 25 years ago in her groundbreaking book *Reflections on Gender and Science*. Astrophysicist Urry made a similar point in her 2008 contribution to the volume *Gendered Innovations in Science and Engineering*, writing "photons have no gender." Of course, while photons have no gender, physicists do. Happily, when even the oldest of us (AB) was in graduate school, the days of legally-sanctioned discrimination and harassment recalled by several eminent women physicists on the 20th anniversary of the CSWP (Gazette Vol. 12, no. 2, 1992) were over. By 2005 the infamous "pipeline" had even stopped leaking women at almost all career junctures (Feder, Ivie). NSF and AIP figures show that fraction of women in physics has continued to rise; between 2000 and 2008 Ph.D.s awarded to women went from around 13% to almost 19%.

On the other hand, the "pipeline" has already delivered an excellent gender balance in professions like medicine and in life and social sciences. Using U.S. physics Ph.D. numbers as a metric, Murnane estimates that we will achieve equity by gender almost a century from now. Prospects for racial balance are even more dire (Ivie, Horton). Data reveal that women have poorer job quality: less satisfaction, lower pay, later promotions, ... and that children have a negative impact on their mother's success, but not their father's.

Are today's inequities due solely to past discrimination — and will they resolve without any further action on our parts? This was certainly AB's belief when she was a student. Attending a girl's school, then Williams College with great professors like Stuart Crampton and David Park, then M.I.T. insulated her from any thought that being a woman could harm her career. The difficulties that arose only when she became an assistant professor align with the concept of "stereotype activation," and can be exacerbated for people in so-called gender or race "stereotype-incongruent" fields (e.g. female physicist or male nurse). An understanding of the effects of subtle, or unconscious bias has arisen thanks to decades of research in social psychology and gender studies. Unconscious gender bias when evaluating people — the topic of the research which we describe below — is well understood and accepted by social scientists. Our concern, though, is that it is not only foreign to most physicists, but that it is tough to accept as well. Acknowledging that we are biased is at odds with the way we physicists conduct business as usual in the lab or at the blackboard. Faith in the ability to be wholly objective in our judgments is the *sine qua non* of physics.

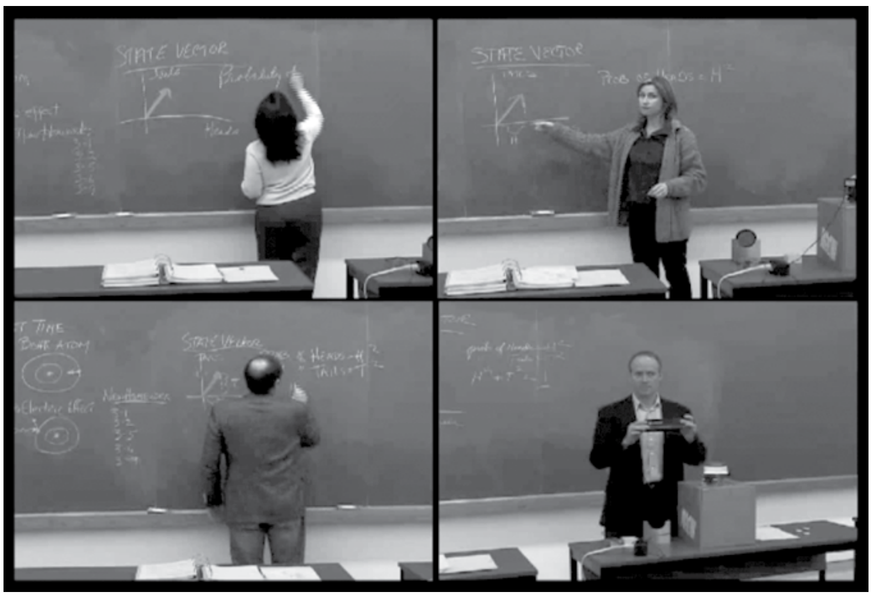
Our research probed a double standard of evaluation — effects similar to ours are well documented in the social science literature. More than one study has shown that teaching evaluations can be quantifiably different according to the gender of the candidate, and that the name (e.g. John, Jane, or J.) on otherwise identical CVs or preprints resulted in different ratings of the document. In grant competitions women can be less successful or receive lesser awards. In 2003 Trix and Psenka analyzed adjectives which appear in the letters of recommendation of male and female candidates for medical fellowships. In 2010 Hebl et al. documented a similar effect in letters in tenure dossiers — and claimed that negative ratings result from female-identified adjectives like "nurturing" and "kind" appearing. (However, it is known that people find it tough to judge competent women as nice, nice women as competent — and to hire women who aren't both. So we might assume that women's letters were downgraded not because they indicated candidates were nice, but because they lacked additional adjectives that signaled agency and competence.)

It is surely not the case that gender bias is active in all hirings, grant competitions, job niches, and so on. One of us (KL) teaches at a community college, with many women physical science professors. Freshmen are surprised to hear that there is gender inequality in the physical sciences. It is only in sophomore classes that they start noticing the lack of women — one which occurs among their classmates. Gender studies is a field guaranteed to annoy a physicist — where the devil is forever in the details, and where unifying principles, when they exist at all, offer much less in the way of accurate predictive power than do the laws of physics.

One such unifying principle is that negative stereotypes are activated when there is some motivation — as in the 2000 study of Sinclair and Kunda.

continued on page 12

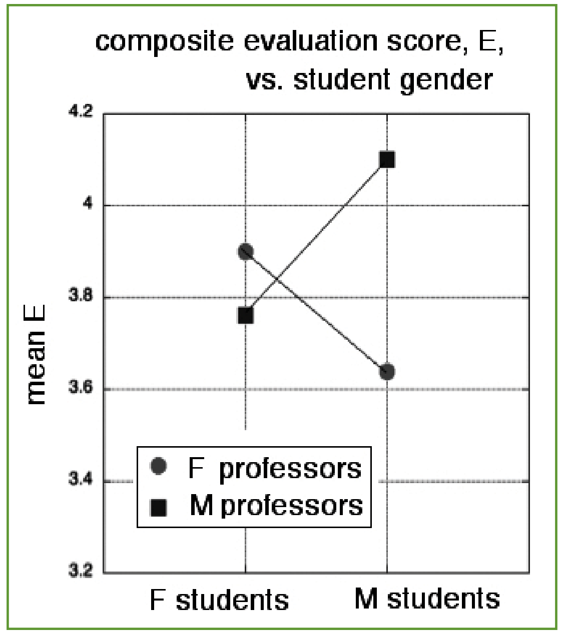
Our study was designed to determine whether male and female physicists giving equivalent classroom lectures would be evaluated differently by students who heard them.



In this Mellon Foundation-funded study, we used videotaped lectures in which professional actors, two male and two female, played the role of physics professors. None of the actors were trained in physics. Each of 126 physics students was randomly assigned to view a single videotape featuring only one of the four "professors." They were not informed that the study was related to gender, or that there was more than one version of the lecture being shown. Students then rated various aspects of the lecture.

Gender Bias in the Classroom, *continued from page 9*

Another is that ambiguity acts as a trigger, as when a candidate is neither a superstar nor a failure, but somewhere in the middle. A higher standard of proof is required for a middling candidate if they are gender- or race-incongruent. In a 2010 study by Brescoll et al, when mistakes in job performance were included in a fictive resume, women/men in typically male/female jobs received lower competency ratings. As a young professor, AB used to worry about what students would think if she made any mistakes at the blackboard, despite reassurance of her male colleagues that this happens to everyone, and it would not matter. Unfortunately, research now suggests otherwise.



Mean "composite evaluation score" for female and male professors (dot female, square male) segregated according to gender of student doing the evaluation. Data shown correspond to N=55 students at one of two institutions studied. Data from second institution is omitted here for visual clarity, but it also has "scissors" form showing interaction between professor and student gender, and will appear in a future publication.

Motivated both by previous studies and by real-world anecdotes, our study (APS March Meeting 2010, *Physics World* August 2010, preprint in progress) was designed to determine whether male and female physicists giving equivalent classroom lectures could be evaluated differently by students who heard them. In this Mellon Foundation-funded study, we used videotaped lectures in which professional actors, two male and two female, played the role of physics professors. None of the actors were trained in physics. Each of 126 physics students was randomly assigned to view a single videotape featuring only one of the four "professors," so student responses were independent. They were not informed that the study was related to gender, or that there was more than one version of the lecture being shown. Students then completed a survey in which they rated various aspects of the lecture using a 5-point scale. They were also invited to write additional comments. They were asked for some personal information, but not their own gender, which was recorded covertly.

While female students rated female professors slightly better, male students rated male professors vastly better. This result is reminiscent of numerous studies of real course evaluations, both for college and high school science teaching.

Our experimental design precluded any difference in the knowledge of physics exhibited by the "physics professors," the scripted words spoken or symbols chalked on the board. Though clearly the four individuals were different in myriad ways, our study eliminated the variability in the intellectual content of the lecture — making it an interesting compliment to studies that look at surveys of actual physics teachers who author their own lectures and speak with genuine authority. For those outside of this field of research, it is natural to wonder if using only two actors of each gender is statistically suspect. This type of social science experiment characteristically uses only one actor — at the very most two — of each "type." (If two, one can statistically test for effects that depend on identity a lecturer, rather than his/her gender.) For example, the 2010 study on customer service representatives of Hekman et al. used one white man, one white woman and one black man — all actors performing the same script. We used actors of the same race, matched for attractiveness and quality of acting resume, and rehearsed in a group setting. In this way, we tried our best to standardize performative aspects of their lectures.

One null hypothesis for the experiment was that student responses would be statistically indistinguishable between the two genders. Responses on fifteen survey questions were combined to create a "composite evaluation score" for a professor. There were three "overall quality" questions at the end (rating of lecture, of lecturer, and whether they should be hired if a job candidate). There was strong correlation between the overall quality and our composite evaluation score. Our statistical tools were common ones in the social sciences: the t-test and analysis of variance (ANOVA). These predict how likely it is for the group means to differ quantitatively by the amount seen if a null hypothesis is true. Further, ANOVA reveals the degree to which independent variables "interact." In other words, the value of one variable, say professor gender, might meaningfully influence what transpires when another, say student gender, is varied. The null hypothesis was upheld for the three questions of overall quality — but only for the female students. According to them, not only was quality independent of gender, but independent of identity ... that all four actors were of indistinguishable quality. The male students disagreed, preferring the male actors. The fact that these results divided sharply along gender lines supports the idea that our attempts to standardize teaching/acting ability were successful, that a distribution of acting abilities among the four actors was not a compelling explanation for our results.

We found several independent variables which were predictive of the mean. These were of institution, professor identity and (marginal effect) professor gender. ANOVAs indicated that there was only one significant interaction - between the independent variables of professor gender and student gender, as mentioned to above. While female students rated female professors slightly better, male students rated male professors vastly better. This result is reminiscent of numerous studies real course evaluations, both for college and high school science teaching.

We also looked at subgroups of questions that one might argue relate to gender-stereotypical attributes. Questions related to having a "solid grasp of the material", being knowledgeable, and being good with equipment yielded a distinct gender bias, in that both male and female students rated male professors as better. Female students were more equivocal; the difference in scores failed to achieve significance for them, but did for male students. On the other hand, questions asking whether the professor "teaches in a way that really helps students learn", was well organized, and interacts well with students produced an own-gender bias. Female/male students rated the female/male professor better.

We believe our results show that in the physics classroom, the gender of the professor can, on average, make a difference in how the class is received, and what sorts of strengths and weaknesses students attribute to the professor. Further, the gender of the student has a role. Clearly, no single study like ours can guarantee that gender is the causative factor rather than, say, the sparse hair or lower voices of our male actors versus the abundant hair or higher voices of our female actors. On the other hand, we would argue that "gender" embraces a multitude of specific features like these, and it would be beside the point to costume the actors identically, graphically equalize their voices, etc. Depicting naturalistic people as physicists was our goal.

What prevents physicists from applying our analytical skills and "throwing away the unimportant terms" related to gender as we teach, recruit, mentor, evaluate, or reward colleagues? We can try, but "gender schemas" — the set of associations suggesting what a person is like based on their gender - are deeply embedded, and tend to govern virtually every interaction we have with others (and even with inanimate objects). (See https://implicit.harvard.edu/implicit/demo/ for the online test of these biases.) Schemas were naturally beneficial to humans in the wild. However, schemas about gender, race, disability, ... create a societal problem by seeping into professional interactions and judgments. The tiny, unconscious biases that infiltrate evaluations can be amplified, say at in the final stage of a hiring decision, when an all-or-nothing decision is made between candidates.

Further, Valian argues that the accumulation of tiny disadvantages will, over time, have dramatic career consequences. This model suggests to us a Monte Carlo simulation. Our careers are like random walks, biased by both gender-independent and gender-dependent terms in the Hamiltonian. To extend the physics analogy a bit further, each walker (person) has a Hamiltonian which has a unique spectrum of such terms (Is the walker a white man? Black woman? Mother? Father?) In each interaction with a teacher, an advisor, a grant committee, ... a Boltzmann factor determines in which direction the walk is likely to proceed. This leads in the usual way, to a steady-state flux. The walk can be projected onto a subspace like professional achievement or personal fulfillment. The walker-averaged flux of men in the positive directions (career success, fulfillment) is higher than that of women, all thanks to gender-biased terms, small though they may be.

In closing, we assert that we do not believe in a single cause for the "Why so few; why less successful?" issue of women in physics. Quite unlike physicists, feminists tend to resist monocausal explanations. Philosophers of science have discussed highly theoretical causes like feminist epistemologies, and symbolic linkages between masculinity and power, logic, and mathematical thinking. Educational psychologists have pointed to loss-of-confidence issues among girls at critical stages in their math and science education. Sociologists and practicing physicists alike have described institutional policy failings, and effects of a "chilly climate" for women. An interesting distinction regarding barriers to women in math-intensive fields is drawn by Ceci and Williams. In 2010 they implicate the choices that women themselves make, but assert that there are two kinds: some free and some constrained by society. We feel that all of these factors and more are part of the complicated pastiche that represents causes and suggests cures for issues of women in physics today.

We hope that our data on gender stereotypes in the physics classroom will encourage action that faces up to the challenge, and contributes to a climate of gender equity in physics.

Much wonderful progress has been made with remedies that take their cues from of these different root causes, to foster equity in the scientific community. Though Wenneras and Wold created a stir with their work on nepotism and sexism in 1994 Swedish Medical Institute fellowships, a study from 2004 could find no such gender bias (but still some nepotism). Along with research on stereotype activation and threat, there is research on reducing it. Moody's 2007 paper is a rich resource in which she both identifies and shows how to avoid the "cognitive shortcuts" that produce biases in hiring and promotion. The AAUW 2010 booklet, *Why So Few?* is another exemplary resource that documents problems and proposes remedies. In conclusion, we hope that our data on gender stereotypes in the physics classroom will not discourage our community, but will encourage action that faces up to this challenge, and contributes to a climate of gender equity in physics. ■