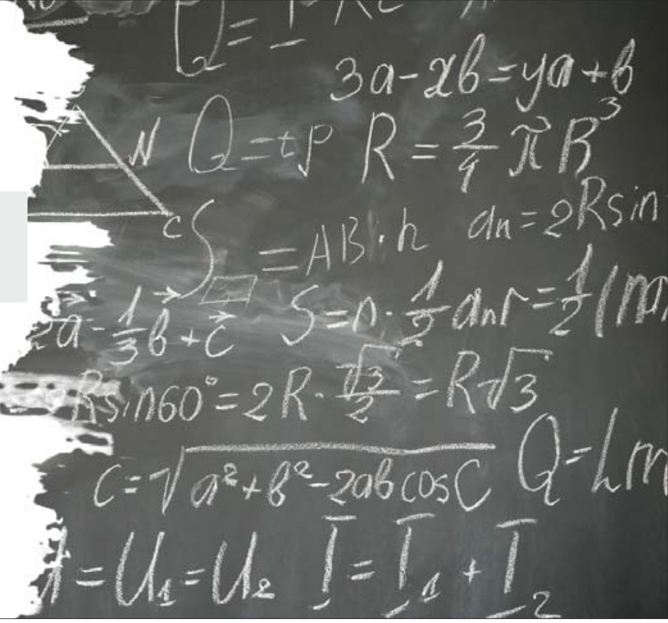
NATIONAL CENTER FOR CASE STUDY TEACHING IN SCIENCE

Everyone Knows Girls are Bad at Math, Right?!



by
Maureen Leonard
Sciences Department
Mount Mary University, Milwaukee, WI

Part I – Introduction

Watch the following video about Mary's experiences with math during her education.

- Video 1: *Everyone Knows Girls Are Bad At Math: Part I* <https://youtu.be/bFFBRu4n22k>

After viewing the video, answer the questions below. Then look up a few articles or a video (at least two sources) that show differences between girls and boys (or women and men). Make sure you bring a hardcopy of your articles or a link for any videos to class.

Questions

Note! Some of the following questions may ask about things you are not comfortable answering so you may skip them.

1. Do you think you're good at math? Why or why not?
2. Have you ever heard that girls are bad at math? What do you think about that claim?
3. What grade did you get in your last math class? What score did you get on your ACT/SAT for the math section? Do you think these are good assessments of your math ability?
4. In the video, what traits and behaviors are being described?
5. What are traits girls have?
6. What are traits boys have?
7. In your articles and the video, how do the sexes differ as described?
8. What causes that difference?
9. How was the cause determined?
10. Do you think those differences are the same for women and men?
11. With what gender do you identify?
12. Do you think you share the traits listed for your gender? How well?

In Class

In the table below record the difference(s) disscussed in class.

Girls/Women	*Boys/Men*

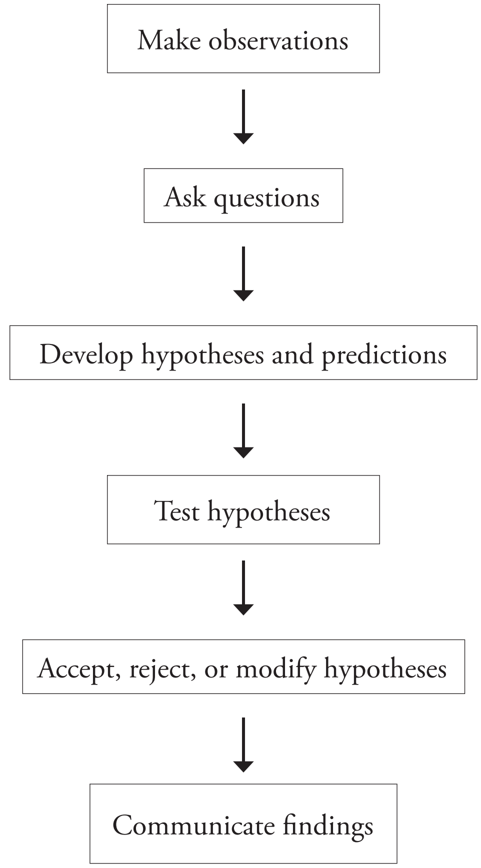
Your instructor should have a sheet for you to record your ACT/SAT math score (anonymously). The instructor will provide the raw data for the class. Draw a graph showing the distribution of grades by sex for everyone to see.

Question

13. Does the graph match what you believed would be the case for your gender? Explain.

Part II – Science and Pseudoscience: Don't Believe Everything You Read!

Mary was taught the scientific method in high school and in her introductory biology class and learned that a key to performing good science is that any hypotheses have to be "falsifiable." That means there has to be a way to demonstrate that the hypothesis is false.



Dr. Metzger, her professor, explained. "A major reason why scientists use the method is to prevent bias from influencing their results and conclusions, but even so there can be implicit bias in the way we ask questions or what kinds of data we collect."

Metzger continued, "A major example of this is found in medical studies. Until relatively recently, female patients were rarely included for fear they might be pregnant (an obvious gender difference), but that meant no sex differences were identified in responses to drugs or other clinical interventions. By expecting the male responses to be universal, this led to differences in care and outcome for female patients."

Mary remembered her grandmother, who suffered a heart attack when she was 64, but the doctors in the emergency room didn't figure out what was happening for a long time because grandma didn't complain of chest pain going down her left arm, like most men experience. Grandma was just tired, out of breath, and her neck hurt. Although her grandma eventually was ok, it was really scary for a while.

Dr. Metzger also talked about another major area where bias can occur is in communicating science, especially to non-science audiences. Mary looked at her articles and realized both of hers were from *Live Science* <https://www.livescience.com/>. When Mary looked at the "About Us" page she saw that only one of the three editors had a science degree and the staff writers also may not have had a science background. Her friend George had an article from the journal *Psychological Bulletin* that was written by scientists with PhDs.

"Sometimes, when an idea is really attractive, and people want it to be true, they may try to convince other people it's true using pseudoscience. Pseudoscience is a process of approaching a claim in a way that looks like science, but is fundamentally flawed. The scientific approach looks at the claim and asks is it even true."

*****

Most of the news articles brought to class were probably not written by the scientists themselves, but by a public relations specialist who may not have fully understood what they were writing about. Perhaps even more concerning is that most headlines are written by a copy editor for the publication rather than by the author.

Other matters of concern include people claiming things as scientific for which they haven't yet done the science, claims that are unfalsifiable, and claims that have already been discredited by science. "Pseudoscience" can sound plausible, especially when it plays into our existing biases.

Let's look at some examples. In your group, look at the articles you brought in. You'll answer the following questions about them and then the class will pool their data.

Questions

1. How many of the articles are about a scientific study?
2. How many of the articles were written by the scientists themselves?
3. How many of the articles were written about a particular scientific study that's named in the article?
4. How many of the articles quote the scientists who did the study?
5. How many of the articles have a conclusion that agrees with the headline completely?
6. How many of the articles don't have a solid conclusion, regardless of the headline?
7. For the articles (if any) that didn't talk about a scientific study, what kind of evidence did they present?
    - *Anecdotal evidence:* an example from one or a few individuals.
    - *Expert opinion:* a report from a recognized expert in the field.
    - *Specialized knowledge:* not common knowledge but knowledge acquired through training or experience.
    - *Documentary evidence:* information that comes from documents.
8. Put your articles in order based on how reliable you think they are.
9. What criteria are you using to order the articles?
10. Would you say any of the articles are actually pseudoscience? What makes you say that?

Part III – Sex and Gender: Not the Same Thing

Dr. Metzger's lesson for the day was about developmental biology and sex determination. She provided the following handout for Mary's class to read and answer.

"Everyone thinks sex is determined just by what chromosomes you have, but scientists keep finding the process is much more complicated than that!"

*****

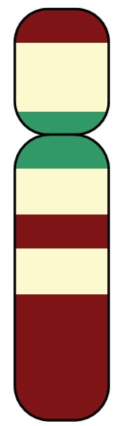
Sex

You may have been taught that sex is determined by the sex chromosomes an individual has. For mammals, XX gives you a female offspring and XY a male one. These chromosomes come together in fertilization, where one is in the egg of the female parent and the other is in the sperm of the male parent.

That's only part of the story though. The Y chromosome is considerably smaller than the X chromosome and is not homologous to the X. That means the genes present are not the same as those found on the X. There is a specific region called the SRY (sex-determining region on the Y chromosome) that is particularly important.

The SRY region contains a gene for a protein called the TDF (testes-determining factor).

The TDF works like a "switch" that turns on all the processes that change the embryonic undetermined gonads into testes between the 4th and 8th week of gestation. Without this protein, the gonads become ovaries. This is the first of many actions that cause embryos to diverge in terms of sex.



SRY location on Y chromosome.

Puberty in adolescence is marked by increased hormonal production and the secondary sexual characteristics arising, leading to what we consider "normal" development for male and female individuals. It should be noted though that men produce estrogen and women produce testosterone—but the relative proportions differ.

There are instances where the chromosomal sex does not match the phenotype of the individual. These disorders of sex development are relatively rare, but provide examples of how one thing alone does not account for the differences between women and men (Table 1).

Table 1. Selected types of mismatched chromosomal sex and gender. Note that this table does not include all types of disorders of sex development, especially those concerning individuals with ambiguous genitalia or those with chromosomal abnormalities.

XY female		*XX male*	
CGD (complete gonadal dysgenesis)	Delayed puberty, amenorrhea	CAH (congenital adrenal hyperplasia)	Testes absent
AIS (androgen insensitivity syndrome)	Delayed puberty, amenorrhea	TRS (testicular regression syndrome)	Testes absent
SRS (Sex reversal syndrome)	Pubertal virilization (male secondary sexual characteristics appear)	SRY translocation	Testes present
TRS (testicular regression syndrome)	Testes absent		
CAH (congenital adrenal hyperplasia)	Testes absent		

Differences in brain anatomy and activity between the sexes have been reported. These differences may start to arise before the gonads are determined, during the first four weeks of embryonic development, but certainly arise by adulthood. For example, some studies show brain lateralization (which side of the brain is more active) to be different between male and female individuals, with the right side (associated with visuo-spatial function) more active in male subjects and the left (associated with language function) in female subjects. How this difference arises is still being researched, though there are marked differences in performance between the genders in spatial rotation tasks and verbal fluency tasks that may result from this lateralization.

An important thing to note is that the studies showing differences in lateralization have not always been successfully replicated, and that earlier studies that claimed large differences in brain anatomy and activity have been replaced with better designed studies showing the early reported differences either are overstated or do not actually exist. Additionally identifying what the subject of a study is, such as humans or rats, is important. Ethically, dissection of human brains can be problematic, so studies done on rats are routinely done. Sometimes the findings of rat studies are applied more broadly, without knowing if the patterns hold across other animals.

Gender

The Merriam-Webster dictionary provides this simple definition for gender: "the state of being male or female." <http://www.merriam-webster.com/dictionary/gender>

The American Psychological Association (APA) provides these definitions for *sex* and *gender:*

> *Sex* refers to a person's biological status and is typically categorized as male, female, or intersex (i.e., atypical combinations of features that usually distinguish male from female). There are a number of indicators of biological sex, including sex chromosomes, gonads, internal reproductive organs, and external genitalia.
>
> *Gender* refers to the attitudes, feelings, and behaviors that a given culture associates with a person's biological sex. Behavior that is compatible with cultural expectations is referred to as gender-normative; behaviors that are viewed as incompatible with these expectations constitute gender non-conformity.
>
> —*The Guidelines for Psychological Practice with Lesbian, Gay, and Bisexual Clients*, 2011.

A biologically deterministic view of sex and gender suggests there are unchanging properties associated with each sex, determining their gender as well. The idea that genes controlling specific behaviors active only in female or only in male individuals or that they evolved to differentiate the sexes is a common one and often arises in pseudoscientific arguments about gendered behavior.

However, many studies on the brain have shown that there is considerable structural variation, but that this variation is not dimorphic—this means there's no "female" brain vs. "male" brain, but it's more of a continuum of variation. How these variations might affect behavior is not well understood, as there are many confounding factors that are difficult to control for. What is known is that, like brain structure, when personality traits, attitudes, interests, and specific behaviors are measured, there is again broad overlap and little consistency even in traits considered to be strictly female or male.

Studies of individuals with disorders of sex development such as those listed in Table 1 above show this pattern as well. Individuals raised as one gender without the knowledge that their chromosomal sex does not align with that gender often behave and believe they are the gender they were raised in.

Questions

1. Looking over the articles you and your peers brought in, do any of them report differences in behavior based on sex hormones?
2. How do sex and gender differ according to the APA?
3. What is meant by "confounding factor"? What kinds of confounding factors might affect behavioral differences seen between the sexes?
4. In the articles and videos you and your classmates collected for Part I, how many of them had specific signifiers of femaleness or maleness? What evidence did these provide for their conclusions? Are they persuasive?
5. Based on the knowledge that there's no male or female brain and that sex and gender are not the same thing, what implications does this have for the idea that sex is either male or female? Do you think it's possible for sex to not be a binary choice (one *or* the other)?

Part IV – What Do We Really Know About Gender Differences in Math Performance?

Mary was upset that so many people seemed to think she was having trouble with math, even though she'd done well in math in school and scored well on the math portion of her SATs. After taking her first exam in Calculus, she found out she got a B on it, which was the lowest score she's ever gotten in math! George had gotten a B+ and was rubbing it in that he did better because "girls are bad at math." She decided to do some research on gender differences in math performance and found a video on it.

*****

Watch the video *Everyone Knows Girls Are Bad At Math: Part IV* <https://youtu.be/wQnLtObO7cE> for homework and answer the following questions.

Questions

1. What differences have been consistently measured between women and men?

2. Fill in the table below. (*Note:* not all studies report both effect size and *VR*.)

Study	*d*	*VR*
Hyde *et al.* 1990		
Hedges and Newell 1995		
NAEP 2008		
Hyde and Mertz 2009		
Else-Quest *et al.* 2010		

3. What trends can you see in reported *d* and *VR* values over time?

4. What do you think accounts for those trends?

Part V – Stereotype Threat: Self-fulfilling Prophecy?

Mary's research showed her that math performance wasn't that different between boys and girls in U.S high schools anymore, so she wanted to see why she did poorly on her math test. She asked Dr. Metzger about stereotype threat, which had been mentioned in the video she had watched. Dr. Metzger decided to give Mary, George, and the rest of the class a paper from the primary literature (an article written by scientists who did an experiment and wanted to communicate the results). The paper was peer-reviewed by other scientists before it was published.

*****

Stereotype threat is defined as reduced performance due to perceived poorer skill by a stereotyped class. You have already examined data showing that while there are real differences between female and male performance in math, those differences are negligible as the overlap between each gender's performance is very large even if mean performance differs slightly in favor of male performance. To explore this idea, we will work through the following primary literature paper:

- Spencer, S.J., C.M. Steele, and D.M. Quinn. 1999. Stereotype threat and women's math performance. *Journal of Experimental Social Psychology* 35: 4–28.

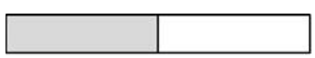
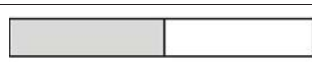
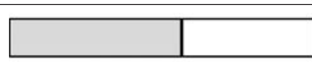
This part of the case has questions throughout for you to answer.

One thing for you to understand is that primary literature is science written for scientists to read, so it is often less clear and more technical than a popular press article. The paper *is* structured around the scientific method, but it isn't always clearly stated. The questions you are answering are intended to help you interpret the article to pull out the components of the method and help you read the article like a scientist would.

Question 1: How have you experienced or observed stereotype threat? What background observations can you make about stereotype threat based on personal experience, popular culture, news media, etc.?

According to Hyde *et al.* 1990, the expected differences in math performance based on gender only appear when the math involved is more advanced.

Table 2. Magnitude of gender differences as a function of math content. (Modified from Table 3, Hyde *et al.* 1990)

Math content	*d*	*Male:Female*
Arithmetic	0.00	
Algebra	0.02	
Geometry	0.13	
Calculus	0.15	

"Explanations of these differences have tended to fall into two camps. Benbow and Stanley (1980, 1983) have argued that they reflect genetically rooted sex differences in math ability. Others (e.g., Eccles, 1987; Fennema & Sherman, 1978; Levine & Ornstein, 1983; Meece, Eccles, Kaczala, Goff, & Futterman, 1982) argue that these differences reflect gender-role socialization, such that males, far more than females, are encouraged to participate in math and the sciences and that the cumulative effects of this differential socialization are most evident on difficult material." —Spencer *et al.* 1999, p. 7.

Question 2: What other observations can you make regarding gender differences in math performance based on the case study including the Part IV video and the above paragraph?

Spencer at al. 1999 examined the effect of stereotype threat on women.

> "In the present research, our central proposition is this: when a stereotype about one's group indicates an important ability, one's performance in situations where that ability can be judged comes under an extra pressure—that of possibly being judged by or self-fulfilling the stereotype—and this extra pressure may interfere with performance." —Spender *et al.* 1990, p6.

Question 3: From the above paragraph, identify Spencer et al. 1990's research question.

> "Widely known stereotypes in this society impute to women less ability in mathematics and related domains (Eccles, Jacobs, & Harold, 1990; Fennema & Sherman, 1977; Jacobs & Eccles, 1985; Swim, 1994). Thus in situations where math skills are exposed to judgment—be it a formal test, classroom participation, or simply computing the waiter's tip—women bear the extra burden of having a stereotype that alleges a sex-based inability. This is a predicament that others, not stereotyped in this way, do not bear. The present research tests whether this predicament significantly influences women's performance on standardized math tests.
>
> "We believe, however, that these processes may also contribute to gender differences in other forms of math achievement as well as test performance (and to achievement deficits in other groups that face stereotype threat, e.g., Steele & Aronson, 1995). For example, the stereotype threat that women experience in math-related domains may cause them to feel that they do not belong in math classes. Consequently they may "disidentify" with math as an important domain, that is, avoid or drop the domain as an identity or basis of self-esteem—all to avoid the evaluative threat they might feel in that domain (Major, Spencer, Schmader, Wolfe, & Crocker, 1998; Steele, 1992, 1997). Such a process, then, originating with stereotype threat, may influence women's participation in math-related curricula and professions, as well as their test performance."
> —Spender *et al.* 1990, pp. 6–7.

Question 4: From the above paragraphs, identify Spencer et al. 1990's hypotheses.

Spencer *et al.* 1999's first study was to replicate the previous studies. Controlling for age, previous math performance, self-evaluation of math ability, etc., women and men from an introductory psychology class were given a type of GRE exam, either the general exam (less advanced) or the specific GRE subject test for mathematics exam (more advanced). Figure 3 (next page) presents the results.

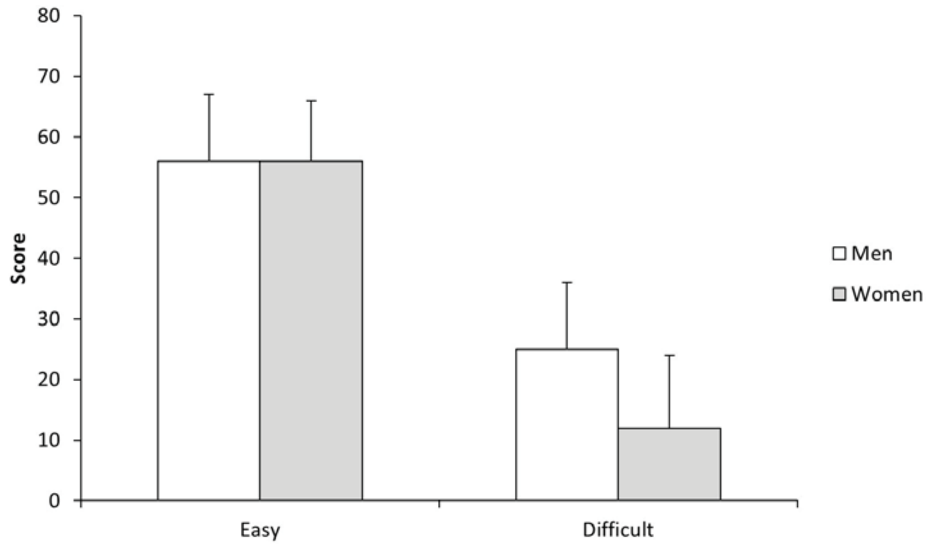


Figure 1. Performance on a math test by sex of subject and test difficulty. Differences in performance on the difficult test were significant (ANOVA F = 3.99, $p < 0.05$). Modified from Spencer *et al.,* 1999.

An ANOVA (analysis of variance) is a statistical measure of the differences between groups in terms of the measured variable. Statistical significance (where $p < 0.05$) means scientists are 95% sure the difference is real.

Question 5: How would you interpret Figure 1?

Question 6: What are the possible causes for the data seen in Figure 1?

Spencer *et al.* 1999 then did a second experiment where all participants, chosen similarly to the first study, were given an advanced math exam like the difficult one used in the first study but the participants were informed in advance of taking the test that it had or had not shown gender differences in the past depending on experimental treatment.

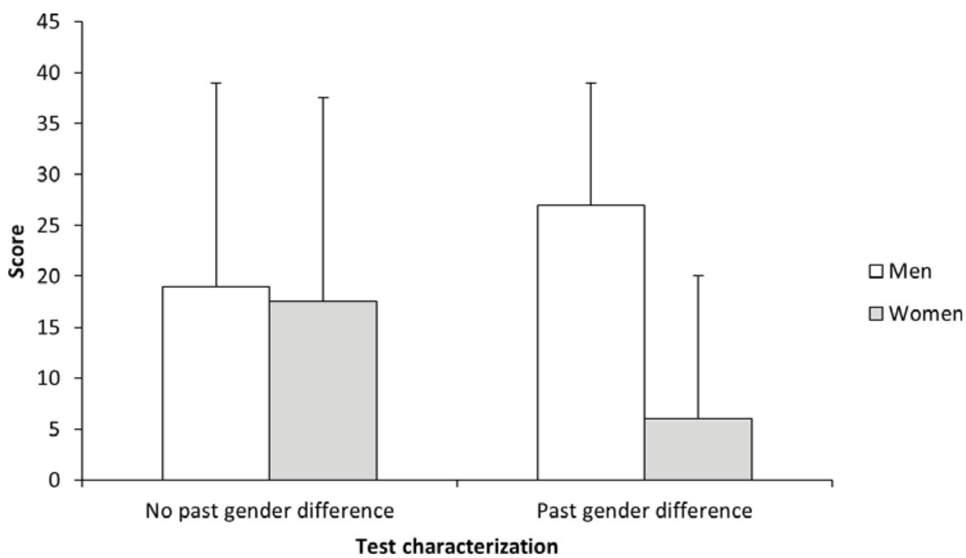


Figure 2. Performance on a math test by sex of subject and test characterization. Women's performance on the test characterized as having a past gender difference were significantly different from all the others (ANOVA F = 4.18, $p < 0.05$). No other means differed. Modified from Spencer *et al.,* 1999.

Question 7: How would you interpret Figure 2?

Question 8: What are the possible causes for the data seen in Figure 2?

Spencer *et al.* 1999 did a third study with subjects that were less highly selected (fewer requirements were needed to be a participant, especially with lower level math requirements) and a more diverse test. Additionally a control group where gender performance was not mentioned at all was added and the experimental group was only told there was no past difference in performance.

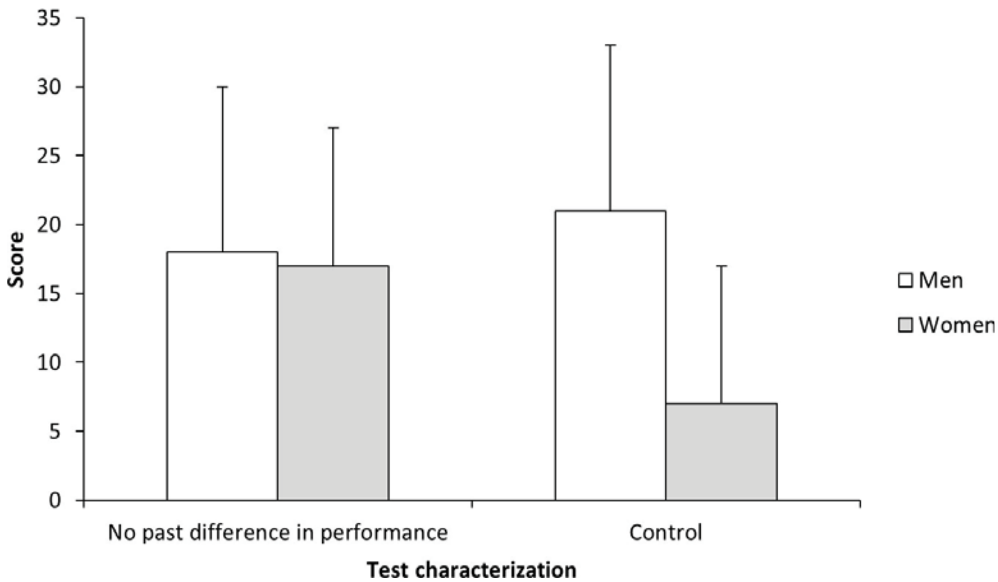


Figure 3. Performance on a math test by sex of subject and test characterization. Women's performance on the test with no mention of gender difference was significantly different from all the others (ANOVA F = 4.78, $p < 0.05$). No other means differed. Modified from Spencer *et al.,* 1999.

Question 9: Do the changes for the third study improve the test of the hypothesis? Explain.

Question 10: How would you interpret Figure 3?

Question 11: What are the possible causes for the data seen in Figure 3?

> "The crux of our argument is that collectively held stereotypes in our society establish this kind of threat for women in settings that involve math performance, especially advanced math performance. The aim of the present research has been to show that this threat can quite substantially interfere with women's math performance, especially performance that is at the limits of their skills, and that factors that remove this threat can improve that performance.
>
> "The three experiments reported here provide strong and consistent support for this reasoning. Study 1 replicated the finding in the literature that women underperform on advanced tests but not on tests more within their skills. Study 2 attempted to directly manipulate stereotype threat by varying how the test was characterized—as one that generally found gender differences or as one that did not. Representing test performance as unaffected by gender, we reasoned, would make the gender stereotype irrelevant as an interpretation of test performance, preclude stereotype threat, and thereby allow women to match the performance of equally qualified men. This is precisely what happened in this condition, while in the condition where the same test was represented as affected by gender, women again underperformed in relation to men. Finally, Study 3 replicated the results of Study 2 with a less highly selected population…" —Spencer *et al.* 1999, pp. 21–22.

Question 12: What are Spencer *et al.* 1999's conclusions?

Question 13: Do you agree with Spencer *et al.* 1999? Explain.

> "The experience of the testing situation itself may be dramatically different for women and men. As the present research shows, stereotype threat as a feature of this situation can undermine women's performance, precisely when the test is difficult. … Thus we may not need to look to the earlier experience of these girls or to their biology to explain their performance. The critical factor may be the stereotype threat of the immediate test-taking situation. ...
>
> "This process may also contribute to women's high attrition from quantitative fields, especially math, engineering, and the physical sciences, where their college attrition rate is 2 1/2 times that of men (Hewitt & Seymour, 1991). At some point, continuously facing stereotype threat in these domains, women may disidentify with them and seek other domains on which to base their identity and esteem. …
>
> "Embedded in our analysis is a certain hopefulness: the underperformance of women in quantitative fields may be more tractable than has been assumed. It attempts to understand the math performance of women not in terms of internal characteristics (e.g., abilities or internalized cultural orientations) but in terms of the interaction between the individual and a threatening predicament posed by societal stereotypes. Predicaments are circumstantial and thus should be easier to change than internalized characteristics. And though our experimental manipulations have yet to establish broadly generalizable strategies for changing this predicament, they do show that it can be changed." —Spencer *et al.* 1999, pp. 24–26.

Question 14: What are Spencer *et al.* 1999's applications of their work?

Question 15: Do you agree with Spencer *et al.* 1999? Explain.

Part VI – What Are the Real Differences Between the Genders?

Mary felt much better about her own math ability after Dr. Metzger's classes on sex and gender. On her next calculus test, she studied more and also went to her professor's office hours to see how to improve her performance by looking over where she made mistakes and how to avoid them. She got an A! (George didn't change what he did to prepare and got another B+.)

*****

Many studies have been done examining what we "know" to be true about the genders, but very few have found real, large differences between them. This has led J.S. Hyde, a prominent researcher in the field, to propose the gender similarity hypothesis in 2005, which states that female and male individuals are similar on most psychological variables, and that's what should be the assumption from the start before we claim there are differences.

Hyde collected meta-analyses of studies on gender differences and found that 78% showed little to no difference between the genders, including mathematics performance and verbal ability, two areas where "everybody knows" women and men are different. Many studies are showing interesting differences between how women and men do some of these tasks, but little to no difference in outcome.

There are some areas of consistent and large differences, such as measures of strength and throwing distance, where men outperformed women by a large margin. Certain measures of sexuality such as incidence of masturbation and attitudes toward casual sex are also different, again with male individuals having higher incidence and more positive attitudes than female individuals. Men are more likely to be physically aggressive than women. Beyond these few things, differences are either highly variable by study or small.

The strength parameters are undoubtedly affected by physiology, especially the comparative levels of testosterone. For those psychological differences however, there is also evidence that they may be less "real" than reported. One example is a study that attempted to allow a person to "forget" their individual identity showed that the rates of aggression were the same between women and men. Contextual differences within studies, such as whether there were onlookers, if the onlookers were known or strangers, etc., also affect the degree of difference measured.

While there are real differences between the genders, many of the ones "everybody knows" are based on nothing but stereotypes rather than evidence. The causes of those real differences are also not clear as yet, but are less likely to be as "hard-wired" as previously assumed. Ignoring the real differences is not useful, but neither is pigeon-holing people based on assumed characteristics or in many cases violating the gender norms associated with their sex due to these often incorrect stereotypes.

Questions

1. What are the costs of societal stereotypes on women? On men?

2. What are the benefits of societal stereotypes on women? On men?