Exposure to Scientific Theories Affects Women's Math Performance

Ilan Dar-Nimrod and Steven J. Heine*

On 14 January 2005, Lawrence Summers, then president of Harvard University, speculated that one reason why women are underrepresented in science and engineering professions is because of a "different availability of aptitude at the high end" (*1*). These remarks were met with much outcry by some critics of President Summers, and social scientists were divided in their reaction to his comments. The question of sex differences in math in the context of the nature-versus-nurture debate is not new and remains contentious. For this paper, we did not explore whether such innate sex differences exist. Instead, we investigated how women's math performance is affected by whether they are considering genetic or experiential accounts for the stereotype of women's underachievement in math. Such a question is relevant to how people respond to scientific arguments and science education more generally.

Stereotype threat is a phenomenon in which the activation of a self-relevant stereotype leads people to show stereotype-consistent behavior, thereby perpetuating the stereotypes (*2*). For example, African Americans perform worse on intelligence tests when their race is highlighted (*2*), and women's math performances decrease when their gender is made salient (*3*). Stereotype threat can be reduced when people focus on the malleability of the traits at hand (*4*).

Past research reveals that people respond differently to genetic and experiential accounts of behaviors. Undesirable behaviors with experiential causes are seen as more voluntary and blameworthy than behaviors with genetic causes (*5*). Experiential causes, in contrast to genetic ones, appear to be viewed as less impactful and more controllable. We reasoned that stereotypes about one's groups are often perceived as inescapable, because many stereotypes are viewed in essentialized terms (*6*). That is, people may view the origin of some stereotypes as resting on the perceived genetic basis that distinguishes these groups. If individuals share the same genetic foundation at the base of the stereotype, they might feel that the stereotype applies to them and hence are vulnerable to stereotype threat. In contrast, we propose that people might react differently if the origins of the group differences were perceived to rest on the specific experiences that people's groups have had. People may reason that their own experiences are different or that they can resist the effects of their experiences.

Our studies manipulated participants' beliefs regarding the source of gender differences in math and measured their subsequent math performance (Fig. 1). In study 1 (*7*), women undertook a Graduate Record Exam–like test in which they completed two math sections separated by a verbal section. The verbal section contained the manipulation in the form of reading comprehension essays. Each test condition used a different essay. Two of the essays argued that math-related sex differences were due to either genetic (G) or experiential causes (E). Both essays claimed that there are sex differences in math performance of the same magnitude. Two additional essays served as a traditional test of stereotype threat. One essay, designed to eliminate underperformance, argued that there are no math-related gender differences (ND). The other essay, designed as a standard stereotype-threat manipulation (S), primed sex without addressing the math stereotype. Controlling for performance on the first math section, we used analyses of covariance to demonstrate that women in the G and the S conditions exhibited similar performances on the second math test ($F < 1$). Women in the E and the ND conditions, although not different from each other ($F < 1$), significantly outperformed women in G and S conditions (all P values ≤ 0.01).

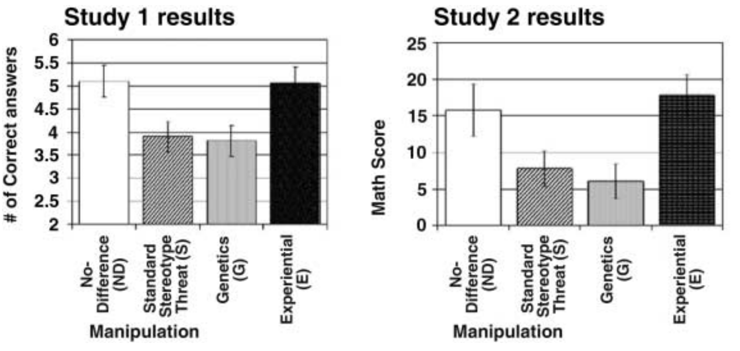


Fig. 1. (**Left**) Study 1 results. Scores on second math test (controlling for scores on first test) after reading essays. (**Right**) Study 2 results. Scores on math test after hearing manipulation.

These findings were replicated in a second study (*7*) that used a different experimental design. An analysis of variance identified significant performance differences between the conditions [$F(3,88) = 4.15$, $P < 0.01$]. Fisher probable least-squares difference (PLSD) comparisons revealed that women in G and S conditions performed comparably ($P > 0.50$) but significantly worse than women in E and ND conditions (all P values < 0.02), which did not differ ($P > 0.50$).

These studies demonstrate that stereotype threat in women's math performance can be reduced, if not eliminated, when women are presented with experiential accounts of the origins of stereotypes. People appear to habitually think of some sex differences in genetic terms unless they are explicitly provided with experiential arguments. It remains to be seen whether the results generalize to stereotypes about other groups and abilities.

Whether there are innate sex differences in math performance remains a contentious question. However, merely considering the role of genes in math performance can have some deleterious consequences. These findings raise discomforting questions regarding the effects that scientific theories can have on those who learn about them and the obligation that scientists have to be mindful of how their work is interpreted. What President Summers perhaps intended to be a provocative call for more empirical research on biological bases of achievement may inadvertently exacerbate the gender gap in science through stereotype threat.

References and Notes

1. L. H. Summers, "Remarks at NBER conference on diversifying the science and engineering workforce," 14 January 2005, www.president.harvard.edu/speeches/2005/nber.html.
2. C. M. Steele, *Am. Psychol.* **52**, 613 (1997).
3. S. J. Spencer, C. M. Steele, D. Quinn, *J. Exp. Soc. Psychol.* **35**, 4 (1999).
4. J. Aronson, C. Fried, C. Good, *J. Exp. Soc. Psychol.* **38**, 113 (2002).
5. J. Monterosso, E. B. Royzman, B. Schwartz, *Ethics Behav.* **15**, 139 (2005).
6. D. A. Prentice, D. T. Miller, *Psychol. Sci.* **17**, 129 (2006).
7. Materials and methods are available on *Science* Online.

8. J. Lau, J. Sim, and R. Vella-Zarb provided assistance. E. Buchtel, E. Dunn, and A. Norenzayan commented on drafts. S.J.H. acknowledges funding from National Institute of Mental Health in the USA (R01 MH60155-01A2) and from Social Sciences and Humanities Research Council of Canada (410-2004-0795).

Supporting Online Material

www.sciencemag.org/cgi/content/full/314/5798/435/DC1
Materials and Methods

9 June 2006; accepted 15 August 2006
10.1126/science.1131100

2136 West Mall, University of British Columbia, Vancouver, BC V6T 1Z4, Canada.

*To whom correspondence should be addressed. E-mail: heine@psych.ubc.ca

www.sciencemag.org/cgi/content/full/314/5798/435/DC1

Supporting Online Material for

Exposure to Scientific Theories Affects Women's Math Performance

Ilan Dar-Nimrod and Steven J. Heine*

*To whom correspondence should be addressed. E-mail: heine@psych.ubc.ca

Published 20 October 2006, *Science* **314**, 435 (2006)
DOI: 10.1126/science.1131100

This PDF file includes:

Materials and Methods

Supporting Online Materials for

Exposure to Scientific Theories Affects Women's Math Performance

Ilan Dar Nimrod
Steven J. Heine
University of British Columbia

Includes Methods, Materials

Study 1

Method

Participants

One hundred and thirty-three female university students (mean age =20.83, SD=3.4) participated in the study. The participants were pre-screened for math proficiency by selecting only students that had completed either grade 12 and/or university level math courses. Twenty-two participants were dropped due to below chance performance (p=.22) on the first math test (prior to the manipulation).

Materials

*Test***.** The test was designed to resemble a GRE test, which was constructed of two sections of math (order counterbalanced) each containing eighteen difficult questions taken from a pool of practice questions. A verbal section, which included the manipulation in form of a reading comprehension test, was inserted in between the math sections. The dependent variable was the performance on the second math test controlling for performance on the first test. Four different reading comprehensions served as the "source of stereotype" (SoS) manipulation. These bogus essays made the following claims:

No Gender Difference (ND) Condition. A meta-analysis across multiple countries revealed that males and females performed equally well on math tests.

Standard Stereotype Threat (S) Condition. The role of the female body in the arts was discussed with relation to women's identity.

Experiential (E) Condition. Males perform 5 percentile points better on math tests than women because teachers biased their expectations during early school formative years.

Genetic (G) Condition. Males perform 5 percentile points better on math tests than women because of some genes that are found on the Y chromosome.

Procedure

Participants were recruited for an experiment titled "Improving grad schools' admission criteria". Participants were randomly assigned to one of the four SoS conditions and received a test package. All the participants started with a math section (Math 1) followed by a verbal section and then with a second math section (Math 2). Following the exam participants were thoroughly debriefed about the deception that was used in the study.

Study 2

Method

Participants

Ninety-two female students (mean age =20.4, SD=3.57) were recruited from a psychology subject pool.

Materials

Test. A 15-minute quantitative test containing 15 difficult questions was constructed. In addition, participants received a 10-minute verbal exam containing 15 questions to decrease suspicion. To reduce the effect of guessing, participants were informed that the exam would be scored by assigning 5 points for each correct answer and deducting 1 point for each incorrect answer. Our analysis utilized this score on the math test as the dependent variable.

Fake Debriefings. The experimenter provided a verbal post-experimental debriefing to a confederate posting as another participant to which the participant was “accidentally” exposed. She read one of 4 different accounts of research findings which were variations of the previously mentioned conditions (i.e., ND, S, G, and E conditions).

Procedure

Upon arrival participants were greeted by a female experimenter who apologized that a previous unrelated study was running late and the participant was asked to wait in the lab room until the alleged previous participant would finish in the exam room. Then, the experimenter left the room and returned with a female confederate who was supposedly the previous participant. The experimenter read the fake debriefing to the confederate out loud, forcing the participant into the role of an audience. The participant was then chaperoned to the experiment room where she took the test. Finally, participants were thoroughly debriefed about the study.

A different group of participants rated the plausibility of our G, E, and ND manipulations in study 2. An ANOVA revealed that the arguments in the manipulations were rated as similarly plausible [$F(2,129)= 1.14$, *ns*].

* All materials are available from the authors upon request and are available online at http://www.psych.ubc.ca/~heine/stereotypematerials.html.

Additional supporting Online Materials for

Experiential Accounts of the Origins of Stereotypes Reduce Stereotype Threat

Ilan Dar Nimrod

Steven J. Heine

University of British Columbia

Includes Methods, Materials, Pilot Study

Study 1

Method

Participants

One hundred and thirty-three female participants took part in the study in exchange for either course credit or two movie passes. The participants were pre-screened for minimum math proficiency by selecting only students that have completed either grade 12 and/or university level math courses. Twenty-two participants were dropped from the central analysis, due to below chance performance on the first math test (prior to the manipulation). These participants were removed as we are unable to distinguish their knowledge from guesses. Participants were 17-39 years old (M=20.5, SD=3.0).

Materials

Test

The test was designed to resemble a GRE template. Following an instruction page, participants received the test, which was constructed of two sections of math (order counterbalanced) each containing twenty difficult questions taken from a pool of practice questions (see Appendix 1).

A verbal section, which included the manipulation, was inserted in between the math sections. The verbal section contained analogies, sentence completion, and reading comprehensions. Four different reading comprehensions served as the "source of stereotype" (SoS) manipulation. These bogus articles made the following claims (see Appendix 2 for the full versions):

No Gender Difference (ND) Condition. An extensive meta-analysis across multiple countries revealed that males and females performed equally well on math tests.

Standard Stereotype Threat (S) Condition. The role of the female body in the arts was discussed with relation to women's identity.

Genetic (G) Condition. Males perform 5 percentile points better on math tests than women because of some genes that are found on the Y chromosome.

Experiential (E) Condition. Males perform 5 percentile points better on math tests than women because teachers biased their expectations during early school formative years.

All essays were given in the early part of the verbal section to insure participant's exposure to the manipulation. All the participants answered the questions about the essays that contained the manipulation. An additional unrelated reading comprehension essay was embedded at the end of the verbal section to decrease suspicion. At the end of the experiment participants completed a demographic questionnaire.

Procedure

Participants were recruited for an experiment titled "Improving grad schools' admission criteria." Participants were run in either small groups (n=2-6) of females only to reduce "minority status" effect (1) or larger groups (n=8-15) containing both males and females to increase ecological validity by creating conditions that are common in regular education settings. This condition did not have a significant main effect $F(1,102) = 1.49$, *ns,* nor did it interact with the SoS variable $F < 1$ and was therefore dropped from the main analysis. Upon arrival participants were greeted by a female experimenter. They were randomly assigned to one of the four SoS conditions and received a test package that contained sample questions of quantitative and verbal reasoning, which the experimenter used to demonstrate what they would be asked to do on the exam. Following the sample questions they were given 15 minutes to complete each section with no intervals between them. All the participants started with a math section (Math 1) followed by a verbal section and then with a second math section (Math 2). The two math sections were counterbalanced. Following the exam participants were asked to complete a demographic questionnaire after which they were thoroughly debriefed about the deception that was used in the study.

Pre-test Study

Method

Participants

One hundred and thirty-two female participants took part in the study for partial course credit.

Materials and Procedure

Participants signed up for a study titled "Evaluation of scientific research and performance". Upon arrival at the lab, participants were asked to view 2 short video clips, featuring a "scientific presentation" and a "humanistic presentation" and provide feedback. The "scientific presentation" video contained footage of a well dressed woman in her thirties who was presented as a researcher from Yale. The "researcher" presented her international research team's findings which corresponded to the ND, G, and E manipulations (the same manipulations that were later adapted and used for study 2 fake debriefings, see Appendix 4). Following a couple of unrelated tasks the participants rated the reliability of the scientific presentation (while the poem was rated on creativity) using a 7-point scale (1- very unreliable, 7- very reliable).

Results

The reliability rating scores were analyzed using a one-way analysis of variance to assess whether the different manipulations (i.e., ND, G, and E) were perceived as equally reliable. The analysis showed no significant differences in the reliability ratings of the ND (M=4.64, SD=1.28), G (M=4.35, SD=1.57), and E (M=4.75, SD=1.22) manipulations $F(2,129)=1.14$, *ns*.

Study 2

Method

Participants

Ninety-two female students at a Canadian university were recruited from the psychology subject pool in exchange for course credit. Participants were between 17 and 48 years old (M=20.4, SD=3.6).

Materials

Personality questionnaire package. To support the cover story that was given to the participants regarding the purpose of the study, participants completed a number of personality and affective questionnaires while waiting for the confederate to finish her earlier, unrelated study. These questionnaires were not a focus of this study and will not be discussed further.

Test. A 15-minute quantitative aptitude test containing 15 questions obtained from the ETS practice GRE exams as well as an instruction cover page was constructed (see Appendix 3). Questions of an intermediate difficulty level were selected (approximately 50% of the examinees answered them correctly), a level that should represent a hard exam for arts students. In addition, participants received a 10-minute verbal aptitude exam containing 15 GRE questions to decrease suspicion. To reduce the effect of guessing, participants were informed that the exam would be scored by assigning 5 points for each correct answer and deducting 1 point for each incorrect answer. Our analysis utilized this score as the dependent variable.

Fake Debriefings. The experimenter provided a verbal post-experimental debriefing to which the participant was "accidentally" exposed. She read one of 4 different accounts of research findings that corresponded to the previously mentioned (i.e., ND, S, G, and E) conditions. Appendix 4 contains the full debriefings). Those in the ND condition heard that a large scale international survey had demonstrated that there were no gender differences in math performance. Those in the S condition heard a debriefing designed to prime thoughts about their gender, by describing how women perceived gender-ambiguous characters as women in recall tests. Those in the G condition heard a debriefing that men outperformed women on math tests by 5 percentile points because of genetic factors (the precise account was slightly different from the one offered in Study 1). Those in the E condition heard that men outperformed women on math tests by 5 percentile points because mothers communicate numbers differently to young girls than to boys.

Procedure

Participants arrived at the lab to participate in a study that was advertised as a project designed to investigate correlations between personality (hence the personality questionnaires) and performance. Upon arrival they were greeted by a female experimenter who apologized that a previous unrelated study was running late and the participant was asked to wait in the lab room until the alleged previous participant would finish in the exam room. In the meantime, the experimenter asked the participant to complete the personality questionnaires at the only available table in the room. Then, the experimenter left the room and returned with a female confederate who was supposedly the previous participant. The confederate set in the only available seat, next to the participant, and the experimenter joined them, apologizing to the participant that she had to debrief and release the confederate. She then turned to the confederate and read her debriefing out loud, forcing the participant into the role of an audience. Once the debriefing was concluded the experimenter thanked the confederate and released her. The participant was then chaperoned to the experiment room where she was given an explanation about the test (along with some sample questions) and asked to complete it. Upon completion, the participants provided some demographic information. Finally, participants were thanked for their participation and thoroughly debriefed about the study.

References

1. M. Inzlicht, T. Ben-Zeev, *Psychol. Sci.* **11**, 365 (2000).

Appendix 1: Source of Stereotype Manipulations, Study 1

No Gender Difference Condition (ND)

There are no gender differences in mathematical abilities, Researchers Say

By DR. ERIN A. GOODEY

The environmental camp in a longstanding controversial issue, which has drawn a lot of attention over the past few decades, has received the most convincing support to date in results released today from an international group of psychology researchers. The researchers claim to find that mathematical reasoning differences, which were assumed to exist between men and women, actually have no hold in reality. The results show that there are no innate differences between males and females in mathematical reasoning.

The new research is the largest published study of differences among males and females in mathematical reasoning. The research was conducted over 8 years in which the participants were followed and their performance closely observed. Unlike previous research in the field, the present study followed both a genetic research design (to look for inherit differences) and a survey research design. In the genetic paradigm, using top of the line instruments (F-MRI, DNA analyzers, and messenger RNA blockers), the researchers failed to find any gender differences on mathematical tasks.

Using the largest sample (over 50 million people!) ever recorded, the researchers compared grades in mathematics and physics in 113 countries around the world. The grades were taken from national comparison exams in elementary schools and high schools. The results showed that males and females were performing just as well on the math sections. The same results were found among college students in 26 countries.

The research was supported by the National Institute of Health (NIH), which provided the international team of researchers, led by Dr. Mark Goldstein from the Harvard Gender Research Institute, with a grant of an unprecedented 35 million dollars to fund a 6-year study of gender differences in education. The results that appeared today in Child Development, one of the leading journals of the American Psychology Association, are only the start of many that will follow in the coming years from this prolific team.

Dr. Thomas Schmidt, speaking for the team, concluded that "another bubble of prejudice had exploded. The sheer magnitude of the study is a guarantee of its results. If there would have been even the slightest difference between males and females in math, the study would have been able to detect it, up to a .001 difference in performance."

"This study is both statistically and clinically significant," said the leading author, Dr. Karen Dinear, director of child and adolescent psychiatry at the University of Wisconsin Medical Branch. "Its magnitude sheds new light on a long discourse concerning gender roles and math performance. I hope that teachers and student will modify their expectations accordingly."

Dr. Laura Wehr, from the University of North Dakota Microbiology and Genes Unit suggested that the results are not as sound as other may claim due to the size of the sample used in the genetic conditions of the study (950 females and 875 males). Other experts predicted more criticism in the coming weeks and months once more researchers in the field have a chance to review the findings.

6. What is the main argument of this article?

a. Gender differences can only be detected by extremely sensitive tests

b. A large sample population strengthens experiment results

c. Males are not always better at math than females

d. Gender differences are negated by social factors

e. Mathematical skills are learned

7. How was mathematical reasoning represented in the experiment?

Comparing algebra and problem solving questions

Comparing mathematic grades

Comparing mathematic and physics grades

Comparing physics grades

Comparing calculus grades

8. Why is this research the most convincing evidence to date?

A large population was tested

Use of advanced technological equipments is more reliable now

Both innate and cognitive factors were tested

Teachers opinions are more valued than others

Experts conducted the experiment

9. Why are experts so excited about the experiment results?

a. More funding will be provided

b. Females will no longer shy away from mathematics

c. Males and Females are equally capable of mathematical reasoning

d. Gender differences can no longer be a good excuse for failing math class

e. Many hours of hard work paid off

Standard Stereotype Threat Condition (S)

Women’s body in art; Women’s unique experience

By ERIN A. GOODEY

The woman’s body has been the subject of numerous artifacts throughout history. Sculptors, painters, designers and even musicians have dedicated their lives trying to capture the mystery of this unique anatomical entity.

Representations of the female anatomy can be traced back to an era before writing was invented, and they are still considered a central theme of contemporary art. The centrality of the theme throughout the ages raises an intriguing question: what is it about women’s body that attracts such lasting attention?

The answer you get depends on who you ask. Dr. Roy J. Stephan, a fine arts professor at the University of Colorado in Bolder, provides us with his perspective: "As human beings we have a built-in curiosity towards things that are different from us. Most artists throughout history were men, and as such, they found the female body with its different curves and shapes to be of unparalleled appeal. Capturing the mystery of the unfamiliar became a life-long quest for them. It became their holy grail."

Dr. Susan Monash from the Department of Women’s Studies at the University of New South Wales, Australia, has a different point of view: "Being a woman is a defining part of identity for every woman in the world. One cannot escape the constant reminders of our gender. We get it in from the stares that are sizing us up as we walk down the street, listening to the lyrics of popular music, reading books, filling out a demographic questionnaire that allows you to omit your race but never your gender, and every time you turn on the TV. I’m sure that the women reading this article will know what I mean here. Women’s body in art is just another case of art imitating life. The centrality of being a woman in life is mirrored in this theme in the arts."

Sara Costner is working as a drawing model in art classes on the Upper East Side in New York. Ms. Costner reveals her outlook, which corresponds to both Dr. Monash’s and Dr. Stephan’s: "I find the way that male artists look at my body in class to be the same as the looks I get from my bank teller. The only difference is that my bank teller tries to hide these looks. They all have this insatiable curiosity about my body that goes beyond its outlines. They all seem to be trying to understand what it is to be a woman." She smiles, and adds "It’s like belonging to an exclusive club."

In a 2003 survey conducted by Dr. Joanne Chan at the State University of New York in Albany, found that most art scholars interviewed seemed to agree that sexuality has an important role in the centrality of the female body in art, but none of them claimed that the centrality originated exclusively from sexuality. These findings, published in the February issue of *American Fine Arts Magazine*, were certainly surprising in light of previous theories that attributed the theme’s centrality merely to sexuality.

Whether in the arts or in reality, it seems that being a woman carries with it a cluster of distinctive experiences and expectations in an array of gender-related themes. It is almost as if we are talking about a separate species that goes beyond human beings. Perhaps we should call them "woman beings".

6. What is Dr. Monash’s explanation for the central role of women’s body in the arts?

a. Women's bodies are different than men's and therefore attract their attention.

b. Being a woman is a defining experience in life and the arts are imitating that.

c. Women's body is important in the arts because of its sexual implications.

d. Popular music's lyrics emphasize the women's body.

e. Consumers of art are interested in the theme and therefore the artist supply the demand.

7. Ms. Costner feels that

a. artists look at her differently than other people.

b. men's interest in her is due solely to her body.

c. men try to understand what it is like being a woman.

d. being a drawing model is an important job.

e. her bank teller acts inappropriately.

8. Past theories regarding the centrality of women's body in the arts assumed a bigger role for

a. sexuality .

b. women's defining experiences.

c. the mystery associated with the body.

d. the influence of the Renaissance.

e. historical changes.

9. Whose views does the final paragraph agree with the most?

a. Dr. Stephan.

b. Dr. Monash.

c. Ms. Costner.

d. Dr. Chan.

e. None of the above.

Genetic Condition (G)

Genes are involved in mathematical abilities, Researchers Say

By DR. ERIN A. GOODEY

The biological camp in a longstanding controversial issue, which has drawn a lot of attention over the past few decades, has received the most convincing support to date in results released today from an international group of genetic researchers. The researchers claim to find genetic bases for well-documented gender differences in mathematical reasoning abilities. The study shows that innate differences exist between males and females in mathematical reasoning.

The new research is the largest published study of polygenetic effects to test the interaction between different genes and higher cognitive functions. One of the main findings demonstrates an interaction of 2 genes located on the Y chromosome (which is found only in males) with genes on chromosome 5 and chromosome 7. This interaction produces hormonal changes guided by the hypothalamus. The onset of the hormonal release is guided by activation of the Brotically area in the frontal lobe. This area is activated when processing mathematical oriented problems. F-MRI scans show these hormonal changes create an increase in the amount of ATP (the body's currency of energy) molecules directed to the hippocampus when a person is engaged in higher mathematical reasoning tasks. The increased energy to this area of the brain, considered the "working memory organ", enables the person to retain more accessible short term memory information while concentrating, a critical element in mathematical reasoning capabilities. This genetic difference seems to explain the findings that boys show superior performance by having on average a grade 5 percentile points higher than girls.

The research was supported by the National Institute of Health (NIH), which provided the international team of researchers, led by Dr. Mark Goldstein from the Harvard Microbiology Research Institute, with a grant of an unprecedented 35 million dollars to fund a 6-year study of polygenetic effects on brain capacities. The results that appeared today in the Journal of American Medical Association are only the start of many that will follow in the coming years from this prolific team.

Dr. Thomas Schmidt, speaking for the team, concluded that "manipulating hormonal state in the same way that the polygenetic effect does, may enable us in the future to elevate females' mathematical reasoning abilities to be in-line with those of males".

"This study is both statistically and clinically significant," said the leading author, Dr. Karen Dinear, director of child and adolescent psychiatry at the University of Wisconsin Medical Branch. "Its magnitude sheds new light on a long discourse concerning the role that genes and the environment play in the finding that, in general, males have higher mathematical reasoning abilities than females."

Other experts said the study was important in adding to the limited knowledge about the effects of different hormones on brain functions.

Dr. Laura Wehr, from the University of Aiwa Microbiology and Genes Unit, suggested that the results are not as sound as other may claim due to the size of the sample used in the study (63 females and 58 males). Other experts predicted more criticism in the coming weeks and months once more researchers in the field have a chance to review the findings.

6. What is the main argument of this article?

a. Males are better at math than females

b. Females are better at math than males

c. Males have a genetic math disadvantage over females

d. Males have a genetic math advantage over females

e. Males and Females both are genetically equipped for math

7. How does the "math gene" work?

a. Through clearer visual representations

b. Flow of energy allows longer short term memory retention

c. Higher levels of cognitive thinking are encoded differently

d. Hormones alter the structure of the brain

e. More areas of the brain are triggered for enhanced mathematical attention

8. According to the article, what is not the cause of math differences between the sexes?

a. Pituitary Gland

b. Hormones

c. Genes

d. Hypothalamus

e. Frontal lobe

9. According to this article, in the future how can females improve their math skills?

a. Taking herbal supplements

b. Asking more questions during math class

c. It is not possible for females to improve their math skills

d. Spending more time on their math homework

e. Altering hormone secretions

Experiential Condition (E)

Expectations are responsible for gender differences in mathematical abilities, Researchers Say

By DR. ERIN A. GOODEY

The environmental camp in a longstanding controversial issue, which has drawn a lot of attention over the past few decades, has received the most convincing support to date in results released today from an international group of psychology researchers. The researchers claim to find reasons for well-documented gender differences in mathematical reasoning abilities. The results show that there are no innate differences between males and females in mathematical reasoning.

The new research is the largest published study of differences among males and females in mathematical reasoning. The research was conducted over 8 years in which the participants were followed and their performance closely observed. Unlike previous research in the field, the present study followed both a genetic research design (to look for internal factors to explain the difference) and a cognitive research design (to look for external factors to explain the difference). In the genetic paradigm, using top of the line instruments (F-MRI, DNA analyzers, and messenger RNA blockers), the researchers failed to find any gender differences on mathematical tasks.

Using an ingenious cognitive paradigm, the researchers manipulated the teachers' expectations of students in 64 elementary school classes in 18 cities and towns around the country. In the experimental condition, the researchers visited schools as educational psychologists and gave students a bogus mathematical test at the beginning of the year. Afterwards, they provided the teachers with fake reports that illustrated that the girls in the class were better in mathematics. Observing the teachers through a video camera in the class, it became apparent that teachers were paying more attention to the girls, were more praising towards them and were more dismissive of the boys. In the control condition, where no manipulations of teachers' expectations had taken place, the opposite pattern was observed. Teachers were more attentive towards boys, were praising them more and were more dismissive of girls. The findings showed that the girls in the experimental condition were superior to the boys if the teachers' expectations were manipulated in one of the first three years of elementary school followed by two more years. Manipulation of teachers' expectations after the third year seems to mitigate the effects of teacher's expectations in the first years of school, but not to enough to turn them around as in the case of a manipulation during the first three years. In the control conditions, boys showed superior performance by having on average a grade 5 percentile points higher than the girls throughout the 8 years of the experiment, providing more support to the general stereotype. The critical period for the students' self-expectations construct seems to be in the beginning of the formal education.

The research was supported by the National Institute of Health (NIH), which provided the international team of researchers, led by Dr. Mark Goldstein from the Harvard Gender Research Institute, with a grant of an unprecedented 35 million dollars to fund a 6-year study of gender differences in education. The results that appeared today in Child Development, one of the leading journals of the American Psychology Association, are only the start of many that will follow in the coming years from this prolific team.

Dr. Thomas Schmidt, speaking for the team, concluded that "manipulating teachers' expectations in the same way that the stereotype does, shows that the construct of mathematical abilities that is apparent in teachers' minds and behavior may as well be the factor that explains gender differences in math". The current research joins a long line of research showing the effect of teachers' expectations on students' performance.

"This study is both statistically and clinically significant," said the lead author, Dr. Karen Dinear, director of child and adolescent psychiatry at the University of Wisconsin Medical Branch. "Its magnitude sheds new light on a long discourse concerning the role that genes and the environment play in the finding that, in general, males have higher mathematical reasoning abilities than females."

Dr. Laura Wehr, from the University of North Dakota Microbiology and Genes Unit, suggested that the results are not as sound as other may claim due to the size of the sample used in the genetic conditions of the study (950 females and 875 males). Other experts predicted more criticism in the coming weeks and months once more researchers in the field have a chance to review the findings.

6. What is the main argument of this article?

a. Mathematics should not be taught in co-ed classes

b. Gender differences cannot be accounted for by innate qualities

c. No reasonable explanations can account for differences in mathematical abilities

d. Teachers should be aware of gender differences

e. Girls are not putting enough effort into their math studies

7. Why is this research the most convincing evidence to date?

a. A large population tested and the time spent on observation

b. Use of advanced technological equipments is more reliable now

c. Both innate and cognitive factors were tested

d. Teachers opinions are more valued than others

e. Children were unaware of the manipulation

8. According the article, how do math differences occur amongst boys & girls?

a. Teachers' high expectations led girls to be more anxious and boys to be more determined

b. Boys were disruptive affecting girls' concentration

c. Girls did not show as much interest in math as boys

d. Teachers were much more likely to help and praise boys than girls

e. Boys played with toys that involved more mathematical reasoning

9. In the future, how can females improve their math skills?

a. Take herbal supplements

b. Ask more questions during math class

c. Find teachers who praise them more

d. Teachers should be aware of their own interactions with students

e. Believe in their abilities

Appendix 2: Study 2 Fake Debriefings

Standard Stereotype Threat Condition (S)

The study that you participated in was designed to add to the recent findings made by an international team of researchers led by scientists from Yale and Oxford who have found that females are constantly aware of their gender, even if only at a subconscious level.

The researchers of this study had 1500 females from 73 countries read several stories with ambiguously gendered characters and then two weeks later had them recall what they remembered of the stories and describe the main characters. Almost all of the women reported the stories back with the main characters being female. In a follow up study they had participants of the same cohort perform a word completion task whereby words could be completed as feminine words or gender-neutral terms. They found that in almost all of these tasks, women completed the words as feminine words. For example _ _ IRT was more likely to be completed as SKIRT by women than completed as SHIRT.

What is most surprising about this study is the consistency with which the effect was found. Almost every country tested found that females showed a heightened awareness of their gender regardless of the task they were assigned.

In our study we are attempting to add to these researchers' findings so that Canada can be included as the 74th country to support this research. That is why we had you do the two word search puzzles, one with feminine terms and one with gender neutral terms. So far we seem to get exactly the same results.

Do you have any questions?

Thank you

No Gender Difference Condition (ND)

The study that you participated in was designed to add to recent findings made by an international team of researchers led by scientists from Yale and Oxford who found that, contrary to some popular beliefs, there are no gender differences in math ability. The researchers of this study used a genetic research method, which was designed to rule out genetic differences relevant to math performance, and a cognitive research method, one that focused on the environmental and experiential aspects unique to each gender.

Despite covering the nature and nurture camps as well as using the latest technology available, the researchers were not able to find any considerable gender differences. Their results showed that each gender's mathematical performance is a bell-shaped curve, with a few students at the top of the class, a few at the bottom, but most of them in the middle. These distributions look exactly the same for each gender with no statistically significant differences.

What is most surprising about this study is the consistency with which the effect was found. Almost every country tested found that males and females showed no difference in mathematic performance or ability. There were few countries in which males performed better on average (like Germany and New Zealand) and there were countries in which females performed better overall (like Sweden and Australia). However, on average, males and females performed equally well and tests analyzing their mathematical ability.

In our study we are attempting to add to these researchers' findings so that Canada can be included as the 74th country to support this research. We are trying to show the same results experimentally since they were already replicated in a survey method with Canadian 10th grade students. So far we seem to get exactly the same results.

Do you have any questions?

Thank you.

Genetic Condition (G)

The study that you participated in was designed to add to the recent findings made by an international team of researchers led by scientists from Yale and Oxford who have found that gender differences in mathematical ability are caused by genetics. More specifically they found that an interaction between two genes on chromosome 3 and a gene on the Y chromosome (which as I'm sure you know is found only in males) causes the pituitary gland to increase the release of certain hormones whenever the left medial temporal lobe is activated. This area of the brain is engaged whenever math-related functions are attempted. Some researchers have even termed this area the math area.

The net result of this interaction is that attempting math problems activates the left medial temporal lobe, which in turn causes the pituitary gland to release hormones. These released hormones cause an increase in the production of ATP in that left medial temporal lobe. ATP molecules serve as the brain's energy source. The more energy that is available to this area of the brain, the more information the individual can take in and use for the mental task at hand. The hard numbers that result from this whole system are that males have, on average, a score that is 5 percentile points higher than that found in females who are matched for age, education, and SES background.

What is most surprising about this study is the consistency with which the effect was found. Almost every country tested found that males performed better in math than females by around 5 percentile points, and this difference was completely accounted for by genetics.

In our study we are attempting to add to these researchers' findings so that Canada can be included as the 74^{th} country to support this research (that's why we took a saliva test at the beginning of the experiment). So far we seem to get exactly the same results.

Do you have any questions?

Thank you.

Experiential Condition (E)

The study that you participated in was designed to add to recent findings made by an international team of researchers led by scientists from Yale and Oxford who found that gender differences in mathematical ability are caused by early childhood experiences. More specifically they found that early selective exposure to count nouns and mass nouns is responsible for the difference. Put simply, differences in math ability between males and females are determined by experiences and are not innate.

Research in linguistics has shown that through the first 3 years of a child's life mothers use both count nouns (e.g., tables, dolls) and mass nouns (e.g., sand, milk) in their verbal interaction with their child. The researchers found an astonishing difference between count nouns and mass nouns used in interactions with daughters and sons. The ratio between mass nouns and count nouns between mothers and their sons was more than 3 times the nouns' ratio of interactions between mothers and their daughters. Following up on the infants' math performance in the first few years of school, children who were exposed to a higher ratio of mass to count nouns as infants and toddlers performed substantially better regardless of their gender. Children with equivalent levels of exposure to the nouns did not differ in their math performances. One especially interesting finding was that even girls who were exposed to more count nouns performed better on average than boys who were exposed to more mass nouns.

What is most surprising about this study is the consistency with which the effect was found. Almost every country tested found that males performed better in math than females by around 5 percentile points, and this difference was completely accounted for by the differences in the exposure to nouns in early age.

In our study we are attempting to add to these researchers' findings so that Canada can be included as the 74^{th} country to support this research. That is why your parents were contacted last week as part of the experiment. So far we seem to get exactly the same results.

Do you have any questions?

Thank you.

```
Math / Gender Data..
(Ilan Dar-Nimrod and Steven Heine, Science 4314 20 Oct 2006, p 435)

proc format ;
value codes 1="G" 2="E" 3="ND" 4="S";
run;
data a;   * 1=G  2=E 3=ND 4=S ;
array ic(4) G E ND S;
*infile "unix:mathdata.txt";
infile "Macintosh HD:Users:jameshanley:Documents:Courses:626:MathGender:mathdata.txt";
input c math1 math2;
do i =  1 to 4; ic(i)=(c=i); end;
math1c = math1 - 4.9099099;
run;


proc means n min mean max; format c codes. ;  var math1 math1c math2; run;

The SAS System 11:20 Tuesday, October 24, 2006

           Variable    N       Minimum          Mean       Maximum
           -------------------------------------------------------
           MATH1     111     1.0000000     4.9099099    13.0000000
           MATH1C    111    -3.9099099   9.9099106E-9    8.0900901
           MATH2     111             0     4.4414414    10.0000000
           -------------------------------------------------------

proc means n min mean max; format c codes.; class c; var math1c math2;

      C  N Obs  Variable    N       Minimum          Mean       Maximum
     -----------------------------------------------------------------
     G      28  MATH1C     28    -2.9099099    -0.4456242     5.0900901
                MATH2      28     1.0000000     3.5714286     9.0000000

     E      27  MATH1C     27    -2.9099099     0.3123123     3.0900901
                MATH2      27     2.0000000     5.2222222     9.0000000

     ND     27  MATH1C     27    -3.9099099    -0.3913914     4.0900901
                MATH2      27     2.0000000     4.8888889    10.0000000

     S      29  MATH1C     29    -3.9099099     0.5038832     8.0900901
                MATH2      29             0     4.1379310     8.0000000
     -----------------------------------------------------------------

proc reg data=a; model math2 = S G E ;

Dependent Variable: MATH2

                            Analysis of Variance

                            Sum of         Mean
   Source          DF      Squares       Square      F Value       Prob>F

   Model            3     45.73062     15.24354        3.473       0.0187
   Error          107    469.63875      4.38915
   C Total        110    515.36937
```

Root MSE	2.09503	R-square	0.0887
Dep Mean	4.44144	Adj R-sq	0.0632
C.V.	47.17003		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	4.888889	0.40318855	12.126	0.0001
S	1	-0.750958*	0.56027753	-1.340	0.1830
G	1	-1.317460	0.56508076	-2.331	0.0216
E	1	0.333333	0.57019472	0.585	0.5601

```
proc reg data=a; model math2 = S G E math1c;
```

Dependent Variable: MATH2

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	4	185.10374	46.27594	14.852	0.0001
Error	106	330.26563	3.11571		
C Total	110	515.36937			

Root MSE	1.76514	R-square	0.3592
Dep Mean	4.44144	Adj R-sq	0.3350
C.V.	39.74247		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	5.103533	0.34121364	14.957	0.0001
S	1	-1.241939**	0.47772816	-2.600	0.0107
G	1	-1.287718	0.47612189	-2.705	0.0080
E	1	-0.052588	0.48386265	-0.109	0.9137
MATH1C	1	0.548414	0.08199698	6.688	0.0001

**Example of adjusted difference... (see class notes on confounding by jh)

S vs ND(ref) ..

unadjusted difference = (4.1379310 - 4.8888889) = -0.750979*

adjusted difference = -0.750979 - 0.548414 * (0.5038832 - (-0.3913914))

= -0.750979 - 0.548414 * 0.8952746

= -0.750979 - 0.490981

= 1.2419**

===== raw data (courtesy of 1st author) ========

OBS	C	MATH1	MATH2	G	E	ND	S
1	2	8	6	0	1	0	0
2	3	9	8	0	0	1	0
3	3	4	4	0	0	1	0
4	1	3	1	1	0	0	0
5	1	5	2	1	0	0	0
6	1	5	4	1	0	0	0
7	1	6	5	1	0	0	0
8	2	2	2	0	1	0	0
9	2	8	5	0	1	0	0
10	2	5	7	0	1	0	0
11	1	5	2	1	0	0	0
12	2	3	4	0	1	0	0
13	2	6	7	0	1	0	0
14	2	3	3	0	1	0	0
15	2	4	9	0	1	0	0
16	2	5	3	0	1	0	0
17	2	5	5	0	1	0	0
18	3	4	5	0	0	1	0
19	3	4	4	0	0	1	0
20	3	4	3	0	0	1	0
21	1	2	6	1	0	0	0
22	1	4	4	1	0	0	0
23	1	4	4	1	0	0	0
24	1	7	6	1	0	0	0
25	1	3	3	1	0	0	0
26	1	3	4	1	0	0	0
27	1	4	2	1	0	0	0
28	1	3	2	1	0	0	0
29	2	4	5	0	1	0	0
30	2	6	6	0	1	0	0
31	2	7	7	0	1	0	0
32	2	3	2	0	1	0	0
33	2	8	5	0	1	0	0
34	2	6	6	0	1	0	0
35	2	4	6	0	1	0	0
36	3	3	4	0	0	1	0
37	3	5	3	0	0	1	0
38	3	5	6	0	0	1	0
39	3	3	3	0	0	1	0
40	1	2	3	1	0	0	0
41	1	2	3	1	0	0	0
42	1	10	2	1	0	0	0
43	2	5	3	0	1	0	0
44	1	3	1	1	0	0	0
45	1	5	5	1	0	0	0
46	1	3	2	1	0	0	0
47	1	2	1	1	0	0	0
48	1	2	2	1	0	0	0
49	3	2	4	0	0	1	0
50	3	1	2	0	0	1	0
51	1	9	8	1	0	0	0

52	2	7	8	0	1	0	0
53	2	2	5	0	1	0	0
54	2	5	3	0	1	0	0
55	1	5	5	1	0	0	0
56	3	6	10	0	0	1	0
57	1	4	3	1	0	0	0
58	1	8	9	1	0	0	0
59	2	5	6	0	1	0	0
60	2	6	4	0	1	0	0
61	3	7	9	0	0	1	0
62	3	7	8	0	0	1	0
63	3	4	5	0	0	1	0
64	3	9	6	0	0	1	0
65	3	3	2	0	0	1	0
66	4	2	3	0	0	0	1
67	4	7	6	0	0	0	1
68	4	1	1	0	0	0	1
69	4	4	5	0	0	0	1
70	4	6	3	0	0	0	1
71	4	5	7	0	0	0	1
72	4	6	5	0	0	0	1
73	4	6	3	0	0	0	1
74	4	6	2	0	0	0	1
75	4	9	8	0	0	0	1
76	4	5	5	0	0	0	1
77	4	4	4	0	0	0	1
78	4	6	8	0	0	0	1
79	4	8	2	0	0	0	1
80	4	9	6	0	0	0	1
81	4	6	2	0	0	0	1
82	4	5	2	0	0	0	1
83	3	2	2	0	0	1	0
84	4	4	4	0	0	0	1
85	4	4	4	0	0	0	1
86	4	4	3	0	0	0	1
87	3	2	4	0	0	1	0
88	3	4	3	0	0	1	0
89	2	5	5	0	1	0	0
90	2	7	8	0	1	0	0
91	2	6	5	0	1	0	0
92	2	6	6	0	1	0	0
93	1	7	6	1	0	0	0
94	1	4	3	1	0	0	0
95	1	5	2	1	0	0	0
96	4	4	0	0	0	0	1
97	4	6	5	0	0	0	1
98	4	13	7	0	0	0	1
99	4	7	6	0	0	0	1
100	3	6	10	0	0	1	0
101	3	5	6	0	0	1	0
102	3	3	6	0	0	1	0
103	3	6	3	0	0	1	0
104	4	5	4	0	0	0	1
105	4	3	3	0	0	0	1
106	4	3	6	0	0	0	1
107	3	5	3	0	0	1	0
108	4	4	2	0	0	0	1

109	3	6	6	0	0	1	0
110	4	5	4	0	0	0	1
111	3	3	3	0	0	1	0

=== jh 2006.11.02 ===

Confounding: Reducing it by Regression

Preamble

- Don't overlook classical, "non-regression" methods
- Regression methods are more "synthetic" (i.e. "artificial")
- Cf chapter 3 by Anderson et al. (c622; readings from aahovw)

Definitions ... / synonyms

Original (statistical, in design of experiments)

- inability to estimate higher order interactions (so typically assume they are zero)

- "mixed up with other effects" or "inextricable"

Epidemiological

- (osm)

Other terms

- "Lurking" (i.e. "hidden") variable

- "Simpson's Paradox" is the most extreme form

(see collection of Simpson's paradox examples under ***Other Resources*** *on* ***c626****)*

Examples...

- Does using a Macintosh lead to sloppier writing? [a]
- Better Service from Canada Post after "Major Restructuring"[a]
- Salaries of Master's and PhD's [a]
- Outcomes of Pregnancy during Residency for women and wives of their male classmates• Admissions of Males & Females to Berkeley Graduate Schools [b]
- Percentage of White & Black Convicts Receiving Death Penalty [a]
- Intelligence Quotient (IQ) - Mother's Milk; Other Variables [a]
- Lung Function of Vanadium Factory Workers <u>Other resources, **c697**</u>
  - vs. reference group (matched for smoking and age) that was 3.4 cm different in ave. height
- Blood Pressure and Altitude - age; height; weight; country [b]
- Longevity - Sexual Activity; thorax size <u>**c622**</u>
- Fatalities & Speed Limit Change - Time [a]
- NEURODEVELOPMENT OF CHILDREN EXPOSED IN UTERO TO ANTIDEPRESSANT DRUGS [b]
- What Does It Take to Heat a New Room? <u>**dataset, c697**</u>

[a] notes on Ch 2, c607 [b] resources this course (678), session 5

Adjustment via regression ...

- "Outcome" **Y**
- Contrast with respect to **X** ("Exposure" variable)

  (for now, say X is binary **X=1** and **X=0**)
- Confounder **C**

CRUDE CONTRAST:

via $E\,[Y \,|\, X\,] = b_0 + b_X\, X$

b_X = crude difference = $\bar{Y}_{X=1} - \bar{Y}_{X=0}$

ADJUSTED CONTRAST:

$E\,[\,Y \,|\, X\,\underline{\underline{\mathbf{C}}}\,] = b_0^{*} + b_X^{*}\, X + b_C\, C$

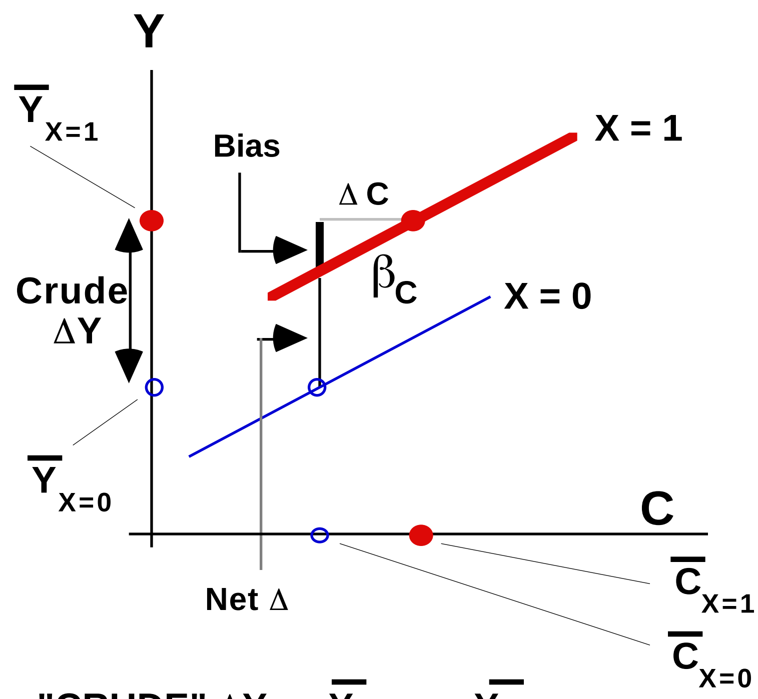
b_X^{*} = adjusted difference

= $\bar{Y}_{X=1} - \bar{Y}_{X=0}$ ***(CRUDE Δ)***

minus

$b_C\,(\,\bar{C}_{X=1} - \bar{C}_{X=0}\,)$ ***(ADJUSTMENT)***

In Pictures... (cf Anderson et al. chapter)



$$\text{"CRUDE" } \Delta Y = \bar{Y}_{X=1} - \bar{Y}_{X=0}$$

$$\Delta C = \bar{C}_{X=1} - \bar{C}_{X=0}$$

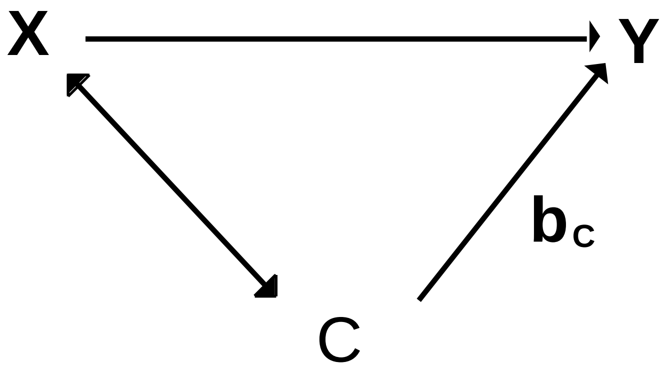
$$\textbf{Bias} = \beta_C \times \Delta C$$

$$\text{"Net" } \Delta Y = \bar{Y}_{X=1} - \bar{Y}_{X=0} - \beta_C \times \Delta C$$

cf Notes re "First Visit to Mars" (G&S Ch 1+) on c678 page

Anatomy of the "Adjustment"

$b_C \left(\bar{C}_{X=1} - \bar{C}_{X=0} \right)$



- for a NON-ZERO ADJUSTMENT...

b_C NON ZERO

AND

$\left(\bar{C}_{X=1} - \bar{C}_{X=0} \right)$ NON ZERO

Special issues

1.

- Adjustment uses a <u>LINEAR</u> relation Y <--> C

If Y <--> C relationship not linear, using a linear relation will not produce correct adjustment

e.g. Y= birthweight and C = Age in residents' study

2.

- If Y <--> C relationship not same at different levels of X

(ie if C is a modifier of X<->Y rel'n,
or X is a modifier of C<->Y rel'n
i.e. if X<-->C "interaction")

then cannot make a unique "adjustment"
(adjustment different at different levels of C)

e.g. gender D's in salary (C = # years experience)
c.f. Miettinen diagram (covariate as a modifier, confounder, or both)

3.

- **<u>Inappropriate Adjustment</u>...**

X ---> C ---> Y

X ---> Y ---> C