

Careers in Academic Medicine: Evaluation at Gatekeeping Junctures

Molly Carnes, MD, MS

Professor, Depts of Medicine,
Psychiatry, and Industrial & Systems
Engineering

University of Wisconsin-Madison

AAMC Benchmark Data, 2005-06

Medical students: 49% (33% in 1985; 21% 1975)

Residents: 43%

IM: 22% (42%), Peds: 16% (67%); Ob/Gyn: 9%(74%);
Ortho: 1% (11%)

Full-time faculty: 32%; 17% full profs; 38% assist profs

5% AA, 4% H, 0.1% NA, 13% Asian

Dept chairs: 102 basic science (avg 1/school); 174 clinical
(avg 1/school)

Deans at US Medical Schools: 11/125, <10%

Are women physicians “leaking out?”

Deans at top 25
medical schools:

Year med school graduation	1971 (1960 -1981)
% women MDs graduating 1971	9%
% women deans (N=2)	8%

Dept Medicine
Chairs at top 25:

Year med school graduation	1974 (1965 -1984)
% women MDs graduating 1974	16%
% women chairs (N=0)	0

What about geriatrics?

Discrete Geriatrics
Section or Division
at top 25, N=16

Year med school graduation	1980
% women chiefs or heads of geriatrics (N=3)	19%
% women MD geriatric fellows 1980	35%

National Study of Internal Medicine Manpower

What Do Women Want?

Basically the same thing men want

- More protected research time
- More institutional support
- Better clarification of expectations of employment
- Improved feedback

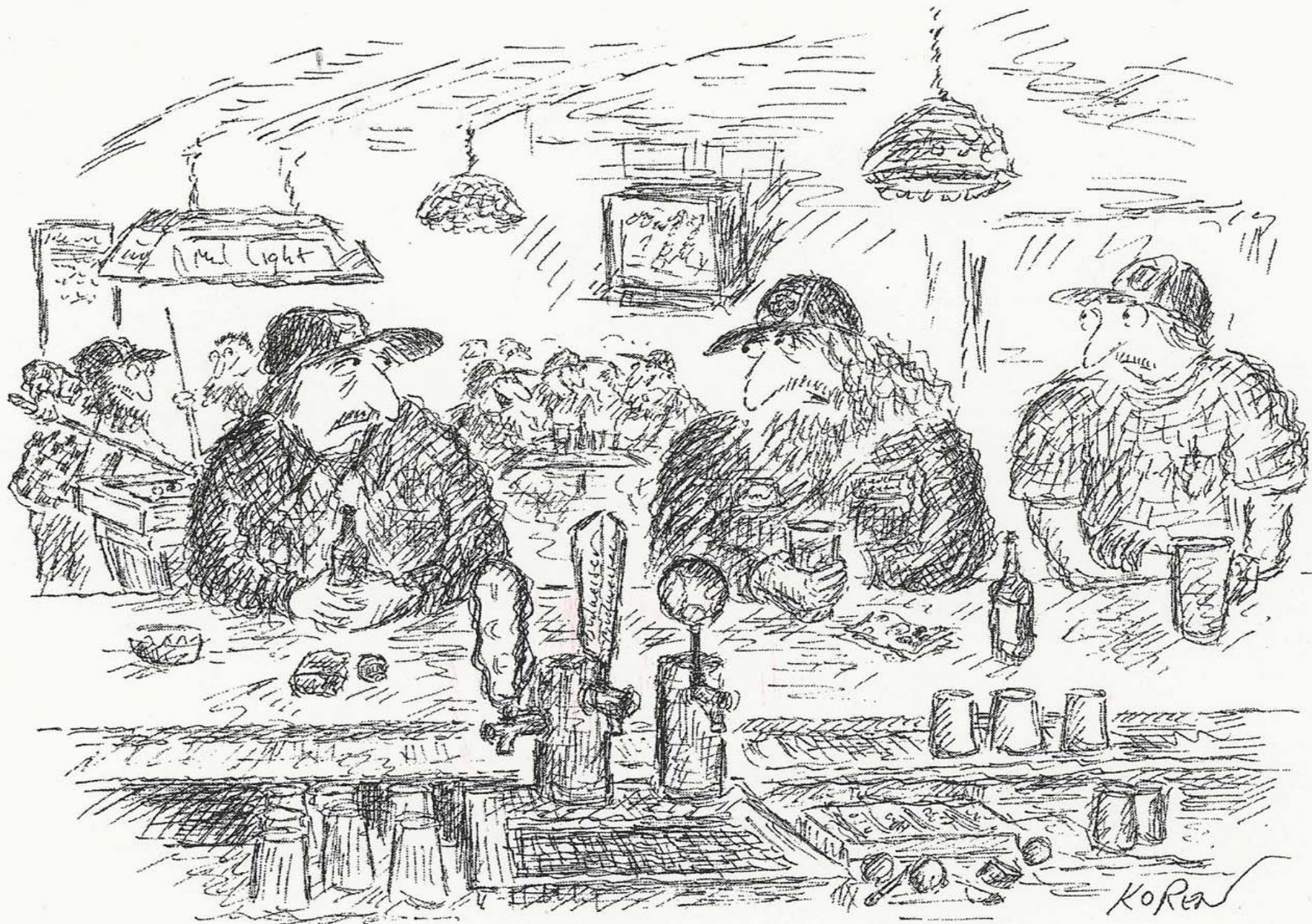
Broadus & Feigel, Chest 105:1858, 1994

WHY?

Attacking the issue with the tools of our trade:

Research and evidence-based action

- Gather data on numbers – AAMC benchmarking, pay equity evals, %women faculty at ranks locally
- Survey perceptions of reasons for lack of advancement - numerous studies tell a consistent story
- Turn to research methods and findings from other fields



"Are you just pissing and moaning, or can you verify what you're saying with data?"

WISELI.engr.wisc.edu

click on Library; extensive
annotated bibliography

What is “unconscious bias”

- Unconscious bias and assumptions
- Previously held beliefs about a social category
- Schemas
- Stereotypes
- Mental models
- Cognitive shortcuts
- Statistical discrimination
- Implicit associations
- Spontaneous trait inference

The tendency of our minds to judge *individuals* based on characteristics (real or imagined) of *groups*

Background: Gender and Behavior

DESCRIPTIVE: How men and women actually behave

PRESCRIPTIVE: Unconscious assumptions about the way men and women in the abstract “ought” to behave:

- Women: Nurturing, communal, nice, supportive, helpful, sympathetic
- Men: Decisive, inventive, strong, forceful, independent, “willing to take risks”

RELEVANT POINTS:

- Leaders, scientists, pioneers: Decisive, inventive, strong, independent
- Social penalties for violating prescriptive gender assumptions
- Unconscious gender assumptions are easily and automatically activated and applied

Language can activate
assumptions about a social
category

Semantic priming activates unconscious gender stereotypes

- Unrelated exercise: unjumble sentences where actions reflect *dependent*, *aggressive* or *neutral* behaviors; e.g.:
 - P alone cannot manage a
 - M at shouts others of
 - R read book by the
- “Reading comprehension” experiment with Donna or Donald engaging in dependent or aggressive behaviors
- Rated target on series of traits (Likert, 1-10)

-
- Gender of target determined influence of semantic priming:
 - Neutral primes – Donna and Donald same
 - Dependent primes – only Donna more dependent
 - Aggressive primes – only Donald more aggressive

Banaji et al., J Pers Soc Psychol, 65:272 1993

Time pressure and high cognitive load
enhance application of unconscious
assumptions

Evaluation of Police Officers Engaging in Competent and Incompetent Behaviors

- 202 undergrads (77 male, 125 female)
- Subjects randomly assigned to 1 of 8 experimental conditions (2x2x2 factorial):
 - Male or female version of police officer's performance
 - Hi or low attentional demands (concurrent task demand and time pressure)
 - Hi or low memory demand

Ratings:

- Competence, job performance, potential for advancement, likely future success → work performance scale
- Adjective scales of gender-related attributes (e.g. dominant-submissive, strong-weak) → composite score

Martell RF. J Applied Soc Psychol, 21:1939-60, 1991

- No effect of evaluator sex
- No impact of memory demand on evaluation
- *Low attentional* demand:
 - Men and women comparable
- *High attentional* demand:
 - Work performance
 - Men higher than women
 - Women same
 - Men higher than men under low attentional demand
 - Gender-related characteristics
 - Men more stereotypically masculine
 - Women same

In a traditionally male job,
ambiguous performance criteria or
evaluation based on “potential” cause
evaluators to fall back on unconscious
assumptions and consistently
disadvantages women

Ambiguous performance criteria in traditionally male jobs favors evaluation of men : “glass escalator”

- 48 subjects (20 men)
- Job description; Assist VP; products made suggested male (e.g. engine parts, fuel tanks). Male and female rated in two conditions:
 - Performance clear
 - Performance ambiguous

Heilman et al., J Applied Psychol 89:416-27, 2004

Competence Score:

Competent - incompetent

Productive - unproductive

Effective - ineffective

Achievement-related
Characteristics:

Unambitious - ambitious

Passive - active

Indecisive - decisive

Weak - strong

Gentle - tough

Timid - bold

Unassertive - assertive

Interpersonal Hostility:

Abrasive - not abrasive

Conniving - not conniving

Manipulative - not manipulative

Not trustworthy - trustworthy

Selfish - not selfish

Pushy - accommodating

Likeability:

Likeable - not likeable

How much do you think
you would like to work
with this person?

Very much - not at all

Comparative Judgment:

Who is more likeable?

Who is more competent?

Results

- Performance clear
 - Competence comparable
 - Achievement-related characteristics comparable
 - Women less liked
 - Women more hostile
- Performance ambiguous
 - Likeability and hostility comparable
 - Men more competent
 - Men more achievement-related characteristics

- Study 2 – women only less liked in male gender type jobs
- Study 3 – Likeability and competence independently linked to recommendation for organizational rewards

Only women were deemed unlikeable for being competent at their job!

Evaluators can re-construct the value of identical accomplishments to get the hire that aligns with assumptions

Redefining Merit to Justify Discrimination

- Mock hiring situation – 3 studies
- Male and female applicants with identical credentials confirmed by ratings
- Police Chief – criteria constructed to favor male applicant, sign. for male evaluators
- Women's Studies Professor – criteria constructed to favor female applicant, sign. for female evaluators
- Self-perceived objectivity predicted gender bias

Uhlman and Cohen, 2005

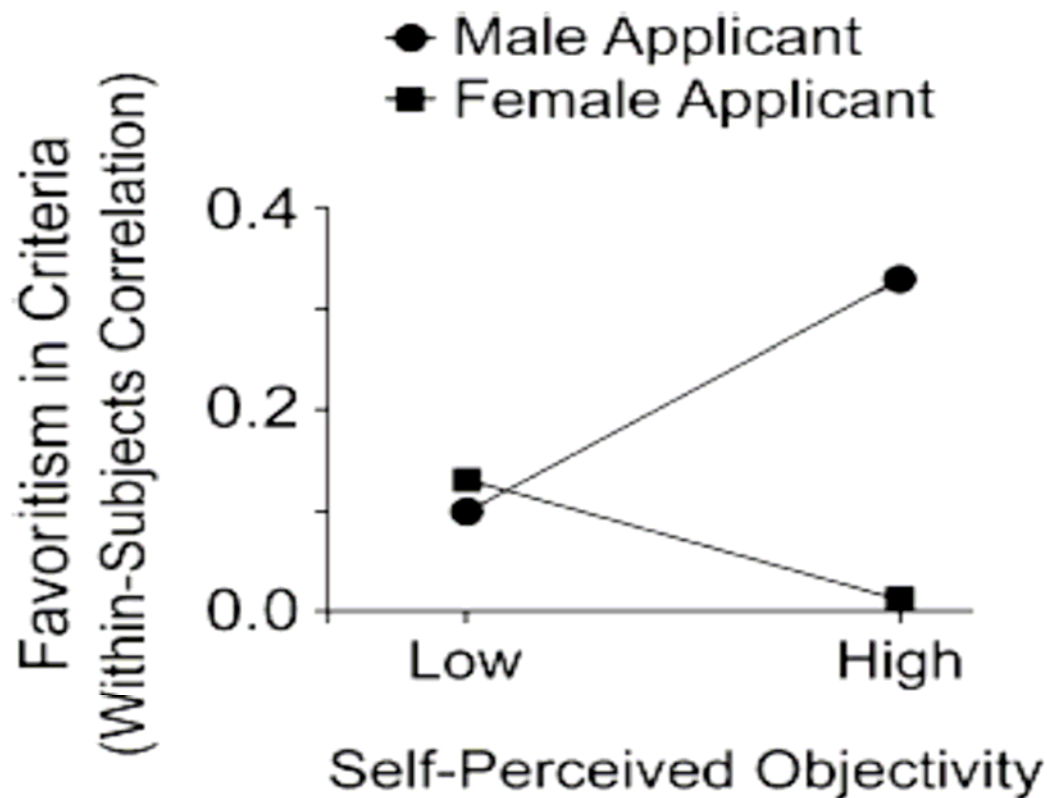


Fig. 2. Results from Experiment 1: the interaction of applicant's gender and self-perceived objectivity in predicting biased criteria. Low self-perceived objectivity is defined as one standard deviation below the mean; high self-perceived objectivity is defined as one standard deviation above the mean. Higher numbers indicate greater favoritism toward the applicant.

Redefining Merit to Justify Discrimination

- Half of the evaluators rated importance of criteria *before* seeing applications (commitment vs no-commitment)
- No-commitment: Criteria constructed to favor male applicant
- Commitment: Male and female applicants – similar hiring evaluations

Conclusion: To prevent the unconscious re-construction of merit to favor the person they want to hire, agree on criteria before seeing any applicants.

Uhlman and Cohen, 2005

Presence of a member of a social category can influence behavior through “social tuning”

Social influence effects on automatic racial prejudice

- Series of experiments measuring automatic prejudice
- Significant interaction of results with race of experimenter (less anti-black prejudice with black experimenter)
- When given instruction to avoid prejudice, further reduction in anti-black automatic prejudice

Lowery et al. J Pers Soc Psych 81:842, 2001

Progress in an Academic Career: Gatekeeping Events

- Prestigious research awards
- Hiring into a faculty position
- Achieving tenure

Progress in an Academic Career: Gatekeeping Events

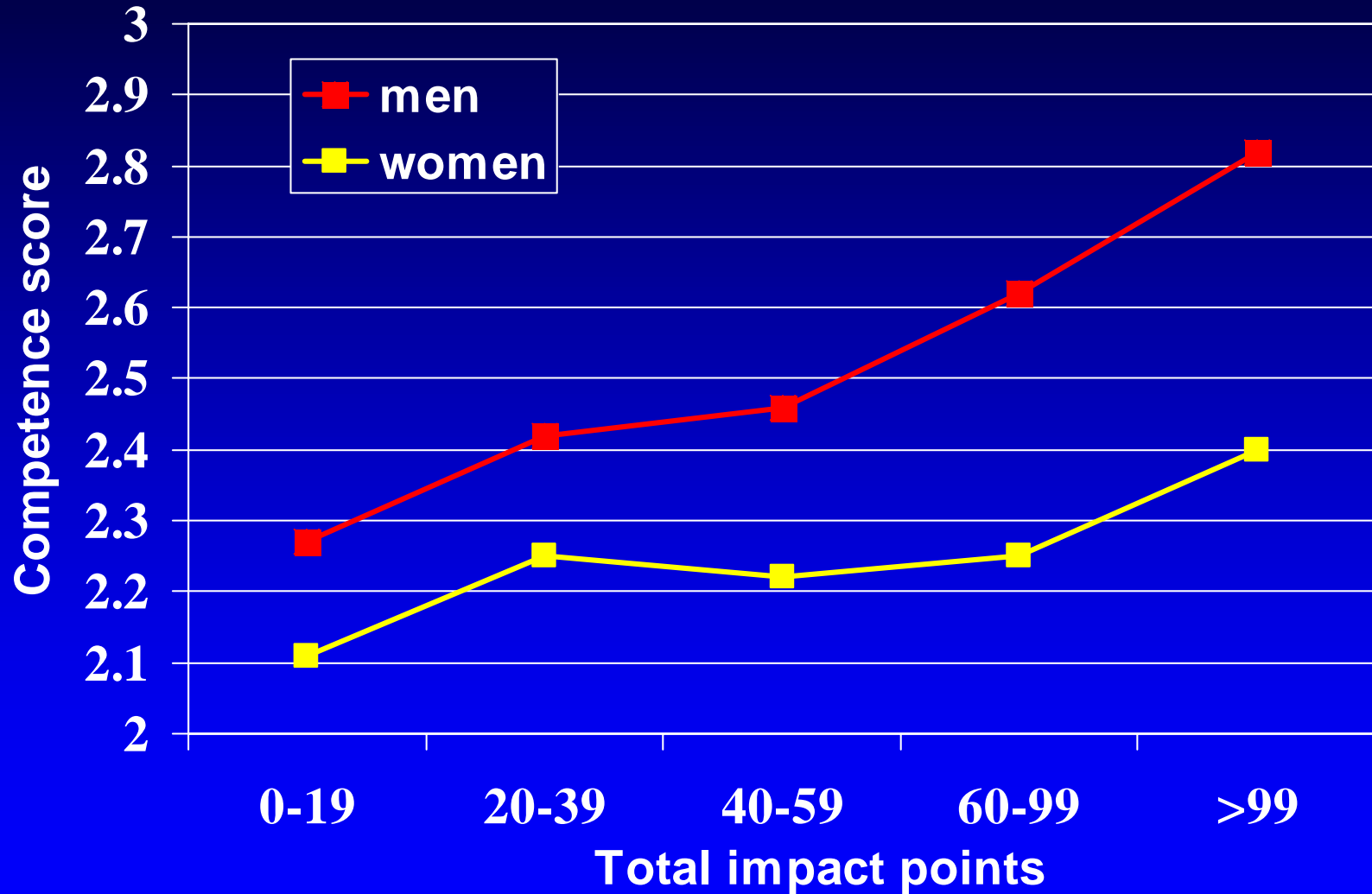
- Prestigious research awards
- Hiring into a faculty position
- Achieving tenure

Real Life Example: Swedish Postdoc study

Wenneras and Wold, Nature 387:341; 1997

- 114 applications for prestigious research postdocs to Swedish MRC (52 women)
- Reviewers' scores vs standardized metric from publication record = impact points
- Women consistently reviewed lower, especially in "competence"
- Women had to be 2.5x as productive as men to get the same score
- To even the score, women needed equivalent of 3 extra papers in a prestigious journal like Science or Nature

Wenneras and Wold, Nature, 1997



NIH Director's Pioneer Award:

Real life example of activation and application of unconscious bias?

- First NIH Roadmap initiative to be rolled out
- Intended to accelerate innovative research unsupported through traditional NIH funding mechanisms
- \$500,000/yr for 5 years
- Drew from all institutes
- New protocol for submission and review
- None of 9 awarded first round were women

Carnes, et al. JWH, 2005

Potential Pool of Women Applicants

Women earn:

- 45% PhD's in biological sciences
- 20% HHMI awards
- 50% MacArthur genius awards
- 25% of R01 applicants
- 23% of all NIH grants

Subsequent Review Cycles

- 2005: 6 out of 14 women (43%)
- 2006: 4 out of 13 women (31%)

Were women doing better science after 2004 or were biases favoring male scientists minimized?

Male semantic primes were present in the announcement and review criteria in 2004 and were removed in subsequent rounds

2004

2005, 06

Characteristics of target scientist and research

Risk-taking emphasized:

- “exceptional minds willing and able to explore ideas that were considered risky”
- “take...risks”
- “aggressive risk-taking”
- “high risk/high impact research”
- “take intellectual risks”
- URL includes “highrisk”

Emphasis on risk removed:

- “pioneering approaches”
- “potential to produce an unusually high impact”
- “ideas that have the potential for high impact”
- “highly innovative”
- URL no longer includes “risk”

Description of recommendations from outside consultants

Technological advances highlighted as desirable:

- “support the people and projects that will produce tomorrow’s conceptual and technological breakthroughs”

Mention of technological breakthroughs removed; human health added:

- “encourage highly innovative biomedical research with great potential to lead to significant advances in human health.”

Social tuning to avoid anti-female bias more likely after 2004

- Huge public outcry
- Many more women present on review committee:
 - 2004: 6/64 (6%)
 - 2005: 28/64 (44%)
 - 2006: 32/79 (40%)
- Wording added to encourage women and minority applicants

Time pressure and cognitive load likely less after 2004

- 2004 – 1300 applications, unfamiliar process
- 2005 – 840 applications, more experience with process
- 2006 – 469 applications

In 2004 evaluation focused on
intrinsic leadership qualities
and performance criteria
emphasized *potential*

2004

2005, 06

Evaluation criteria

Intrinsic qualities stressed:

- “*Potential* for scientific leadership”
- “Testimony of *intrinsic* motivation, enthusiasm, and intellectual energy”
- Reviewers told to look at *potential* for future work

Focus on intrinsic abilities removed:

- “Relevance of the research and impact on the scientific field and on the NIH mission”
- “Motivation/enthusiasm/intellectual energy to pursue a challenging problem.”
- Reviewers encouraged to look at accomplishments as evidence

Progress in an Academic Career: Gatekeeping Events

- Prestigious research awards
- Hiring into a faculty position
- Achieving tenure

Preference for male applicants in academic settings

- 238 academic psychologists sent a curricula vitae with either male or female name
 - Entry level: more likely to vote to hire man, more likely to indicate man had adequate teaching, research, and service experience
 - High level: no gender differences
 - No differences between male and female evaluators
 - More write-in comments for women

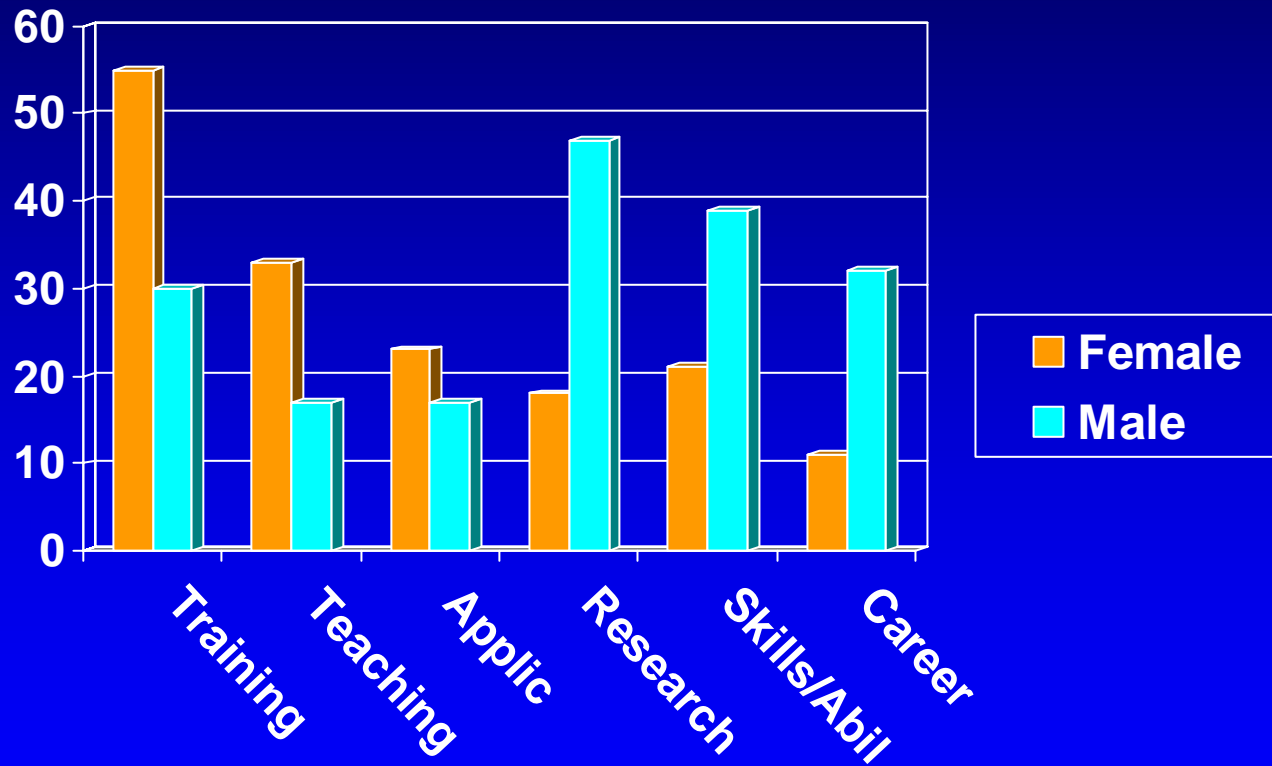
Steinpreis, Anders, and Ritzke 1999

Subtle gatekeeping bias – letters of recommendation

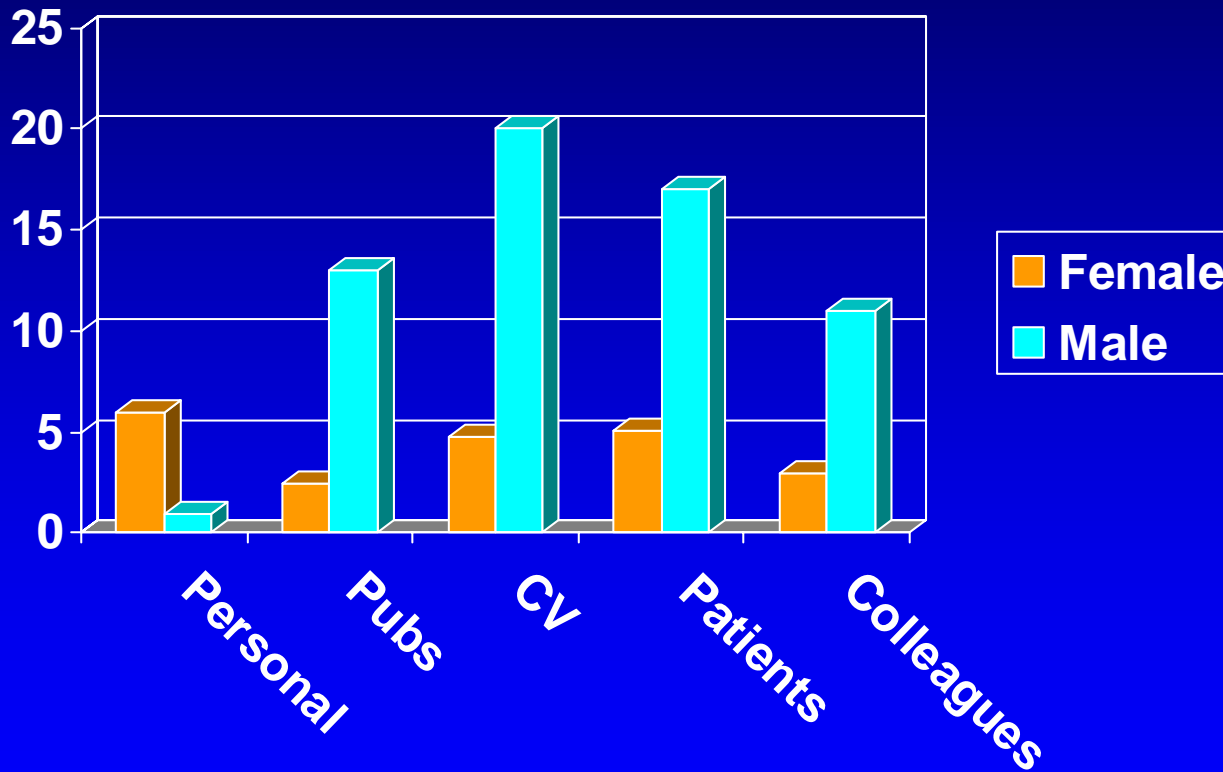
Trix and Psenka, *Discourse & Soc* 14:191 2003

- 312 letters of rec for medical faculty hired at large U.S. medical school
- Letters for women vs men:
 - Shorter
 - 15% vs 6% of *minimal assurance*
 - 10% vs 5% with *gender terms* (e.g. “intelligent young lady”; “insightful woman”)
 - 24% vs 12% *doubt raisers*
 - *Stereotypic adjectives*: “Compassionate”, “related well...” vs “successful”, “accomplished”
 - Fewer *standout adjectives* (“outstanding” “excellent”)

Semantic realms following possessive (e.g. “her training”; “his research”)



Distinctive semantic realms following possessive



Progress in an Academic Career: Gatekeeping Events

- Prestigious research awards
- Hiring into a faculty position
- Achieving tenure

Bias in Evaluation of Leadership/Competence

“Think-manager-think-male phenomenon”

Prescriptive Gender Norms

- **Men**

- Strong
- Decisive
- Assertive
- Tough
- Authoritative
- Independent



- **Women**

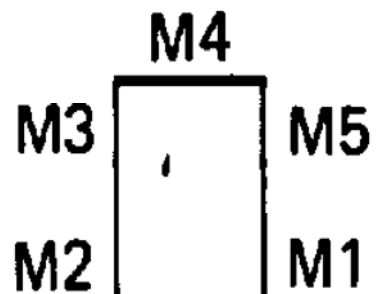
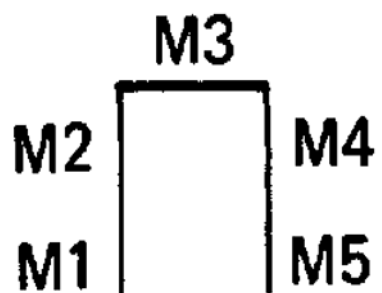
- Nurturing
- Communal
- Nice
- Supportive
- Helpful
- Sympathetic

Evaluation of Leadership/Competence

- Students seated around the table—
when is the head of the table identified
as the “leader?”

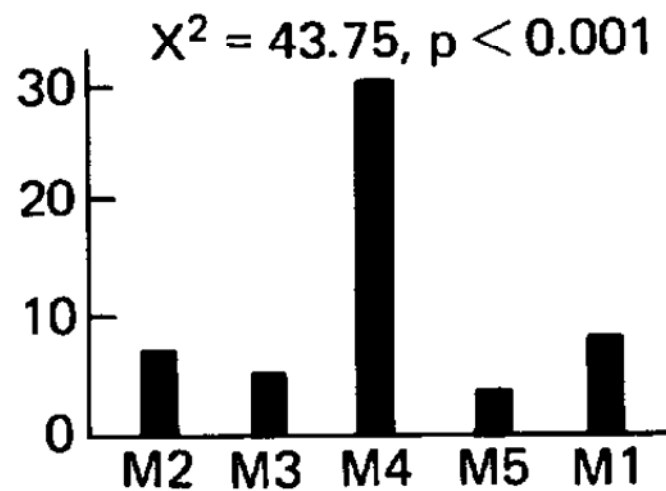
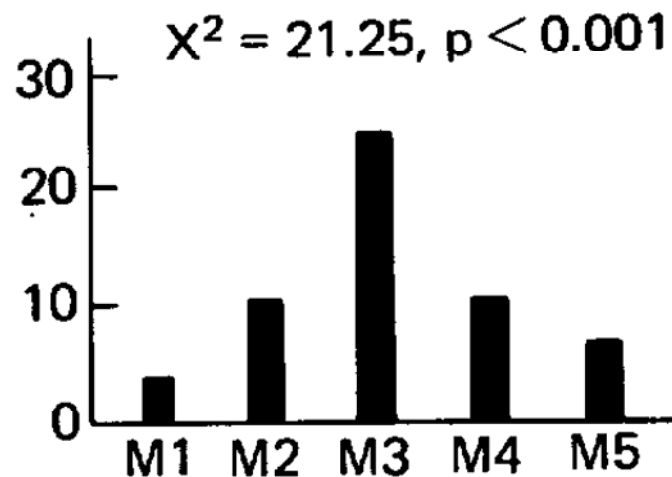
Porter & Geis 1981

MALE

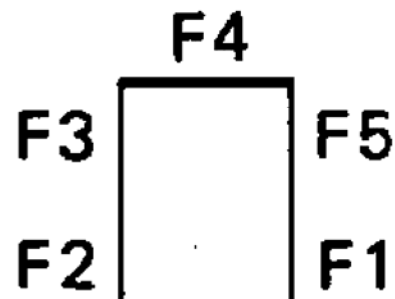
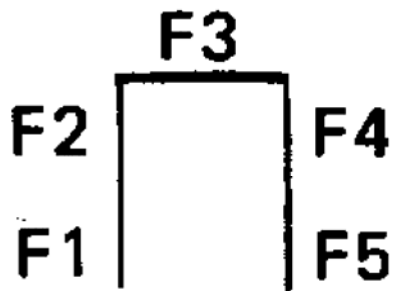


MALE

SAME-SEX STIMULUS GROUPS

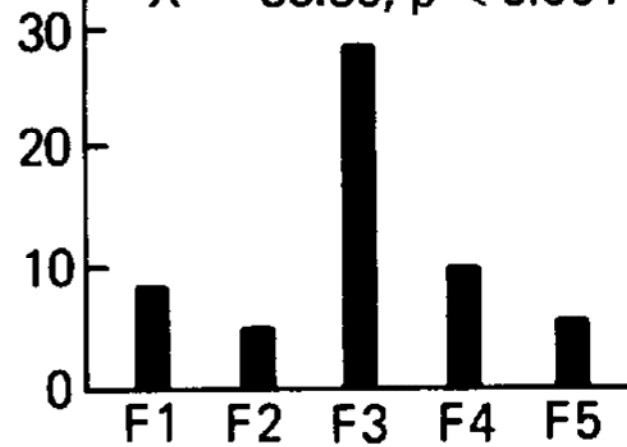


FEMALE

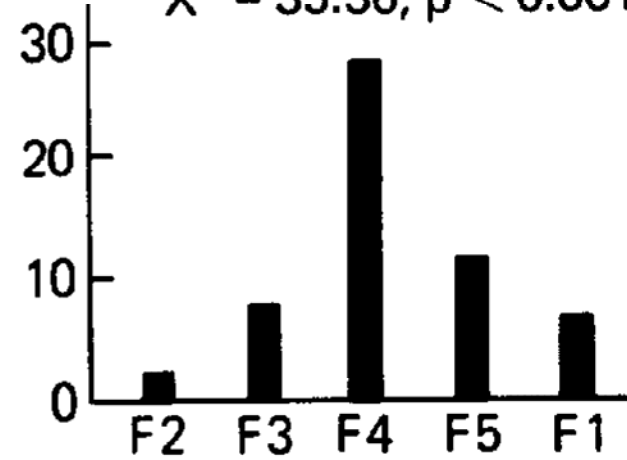


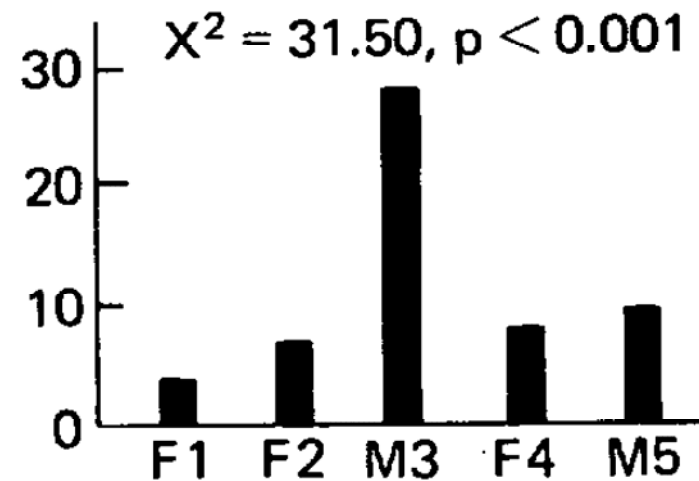
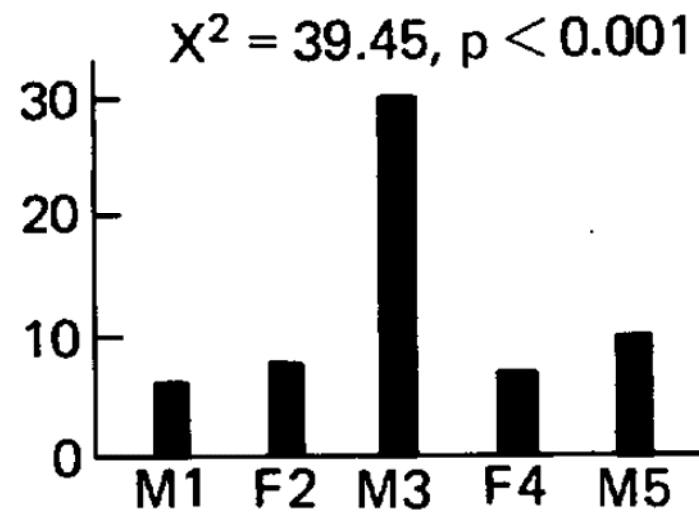
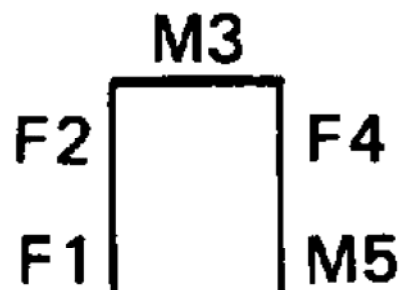
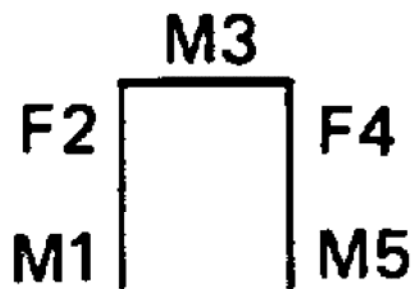
FEMALE

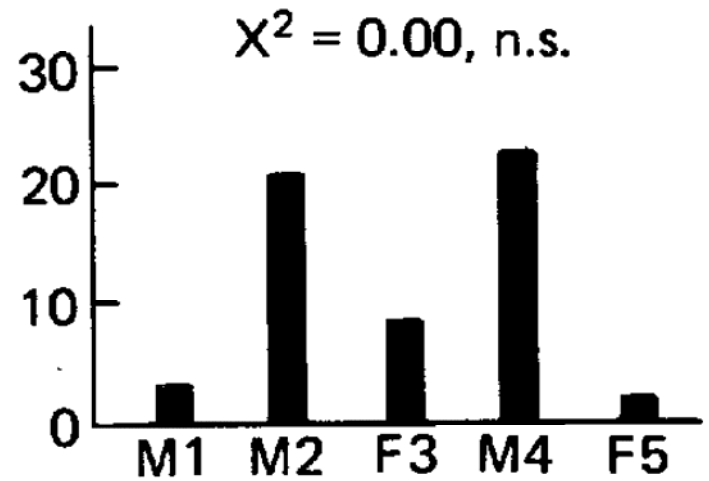
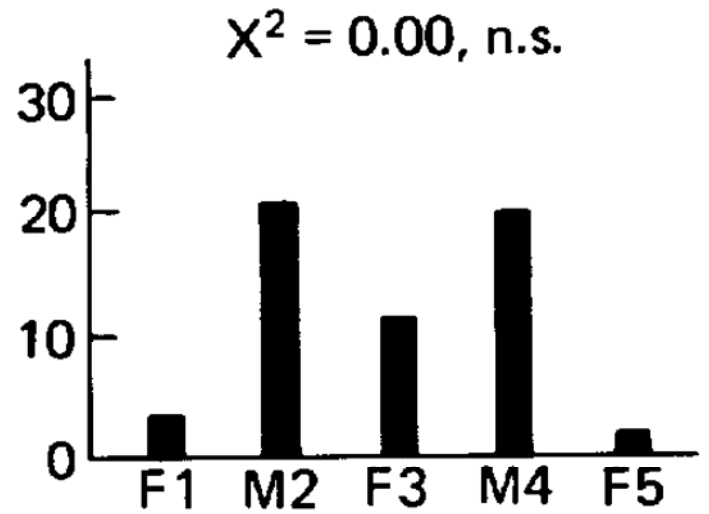
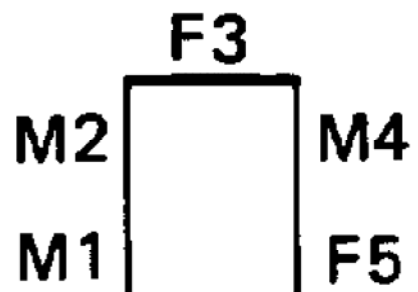
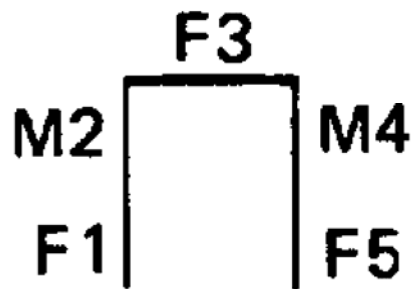
$\chi^2 = 35.36, p < 0.001$



$\chi^2 = 35.36, p < 0.001$







Characteristics of Effective Leadership – is there a basis for the prejudice favoring male leaders?

- Transformational **
- Transactional
- Laissez-faire

Are Men better Leaders than Women?

- Metanalysis of 45 studies measuring leadership effectiveness – Eagly et al., 2002
- Leadership effectiveness of 16 male and 6 female deans – Rosser et al., 2003

Conclusion: little difference between men and women. When differences emerged, women exhibited more transformational and men more laissez-faire leadership behaviors

“Leader” in tenure criteria

- 25 top research academic medical centers
- Tenure criteria from websites
- Scanned for “Leader”
- Also scanned for other Bem Sex Role Inventory male, female, neutral words
- Slopes of regressions for annual % faculty who are tenured women x 7 years
- “Leader” = OR 6.0 (1.02, 35.37; $p=0.04$) for slope below median compared to those without

Carnes et al. 2007

Stereotypically male traits valued for tenure

Male

- Analytical
 - Competitive
 - Defends
 - Independent
 - Individualistic
 - Leadership
 - Risk
-

Neutral

Friendly
Helpful
Inefficient
Truthful

4 schools

Total 5

Female

Sensitive
Understanding
Yielding

3 schools

Total 3

Med 5.5/school; 2-50

Total 183

Conclusions

- Even the most well-intentioned person committed to gender equity has unconscious biases about social categories
- These assumptions can disadvantage women at gatekeeping junctures in fields traditionally occupied by men such as academic medicine

Recommendations

- Acknowledge that we all have biases and assumptions
- Examine language and other processes at gatekeeping junctures in the context of research from social psychology
- Continue to raise awareness of the fact that:
 - “fixing the women” is not enough to achieve gender equity
 - With women comprising 50% of the pipeline, we cannot afford a system that utilizes only half of the potential talent in academic medicine



Abstract language reinforces and transmits stereotypes

- Subjects: 72 from University community (36 women/36 men)
- Subjects asked to write 4 stories about a male or female friend behaving in pos and neg stereotypically male and female way
- Read other's stories
- Dispositional inference:
 - Repetition likelihood
 - Situation attribution
 - Person attribution
 - Situation-person attribution
- Rate behavior stereotypically male, female, desirable, undesirable (Likert, 1-7)
- Level of abstractness computed from verbs and adjectives

Abstract language reinforces and transmits stereotypes

- No effect of evaluator sex or desirability of behavior
- Writers' description of behavior more abstract when gender-congruent (expectant)
- Readers' – when behavior rated gender-congruent (expectant) \Rightarrow greater dispositional inference
- Readers' dispositional inference \Rightarrow accounted for by level of linguistic abstractness of the story

Conclusion: Expected information is communicated at a higher level of abstractness than unexpected information and this effectively maintains gender stereotypes in recipients

Abstract language reinforces and transmits stereotypes

- Subjects = 72 Dutch from University community
- Stories stated concrete or abstract gender-congruent behaviors based on traits:
 - Male: independent, handy, adventurous, technical
 - Female: careful, considerate, emotional, spontaneous
- Subjects asked to rank target stereotypical male or female (Likert 1-7, *not at all* → *very much*)
- Abstract stories led to stronger dispositional inferences regardless of content

Conclusion: linguistic expectancy bias may lead to subtle, undetected forms of discrimination

Three Examples

- Semantic priming
- Linguistic expectancy bias
- Language that ignores or blames women

NIH Extramural Nexus, January, 2007

“The disproportionate *difficulty women have* as Principal Investigators of large grants was obvious in the first round of the Clinical and Translational Science Awards (CTSAs) applications, where none of the applicants were women.”

Women are not people

- Sixth Report of the Joint National Committee on Prevention, Detection, Evaluation and Treatment of High Blood Pressure. Arch Intern Med, 1997
 - “No more than 1 drink per day in women and lighter-weight persons.”

Letter to the editor with apology and promise to be more careful in JNC7.
- Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. JAMA, 2006
 - “No more than 1 drink per day in women and lighter-weight persons.”

Letter to the editor not accepted for publication.

Women's continued invisibility in clinical research

- [Wooley and Simon, NEJM 343:1942-1950 2000](#) – review on managing depression in medical outpatients

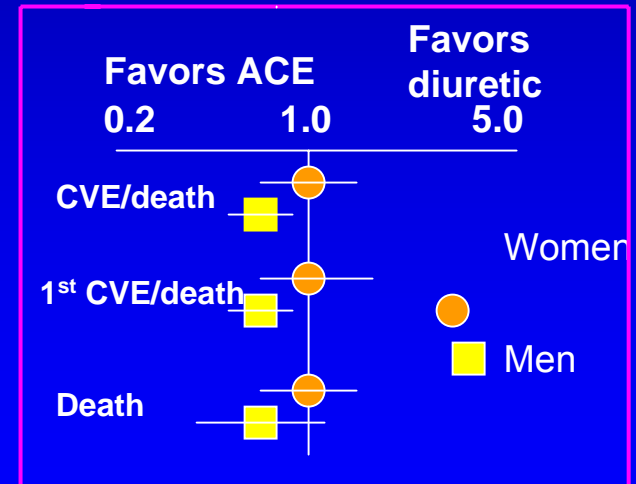
No mention of:

- greater prevalence in women;
- postpartum depression;
- safety of antidepressants during pregnancy or nursing;
- how to counsel women on rx who want to get pregnant;
- childhood sexual abuse, domestic violence, or sex and gender-based harassment in the work place as risk factors

- [Wing et al NEJM 348:583-92 2003](#) – ACE vs diuretic for HT in elderly outpatients

- results in older women ignored;
- results extrapolated to “the elderly”

- [McFalls et al NEJM 351:2795-804 2004](#) – RCT CABG before elective vascular surgery
- 98% men
- results extrapolated to “patients”



Inadequate compliance with NIH guidelines to include women in clinical trials

- NIH Revitalization Act of 1993 – NIH required to include women; other fed agencies followed
- 9 high impact medical journals in 2004
- 46 studies, not sex-specific
 - 70% enrolled \geq 30% women
 - 40 (87%) did not report outcomes by sex or include sex as a covariate in modeling
 - None acknowledged limits of generalizability (including 7 studies with $<$ 20% women)

Summary

Feature of process	Predict preferential selection of men	Present in 2004	Present in 2005
Time pressure	Yes	Yes	Less
Semantic priming in RFA	Yes	Yes	No
Intrinsic leadership + ambiguous criteria	Yes	Yes	No
Women <25% applicant pool	Yes	Yes	No
Women < 35% of reviewers	Yes	Yes	No
Instruction to avoid stereotype	No	No	Likely

No. women/total awards (%)

0/9 (0)

6/14 (43)

Women afforded a narrower range of behaviors to be taken “seriously”

valuators of corporate leaders more likely to attribute the success of attractive women to luck and unattractive women to ability

- Heilman ME and Stopeck MH. Attractiveness and corporate success: Different causal attributions for males and females. J App Psychol 70: 379-388 1985.

men more inclined to like and be influenced by a competent woman speaker when she is also sociable than when she is merely competent

- Carli et al. Nonverbal Behavior, Gender, and Influence J Personal & Soc Psychol 68:1030-42, 1995

behavior perceived as too feminine – activates assumptions of less competence; perceived as too un-feminine – activates social penalties for violating proscriptive gender assumptions

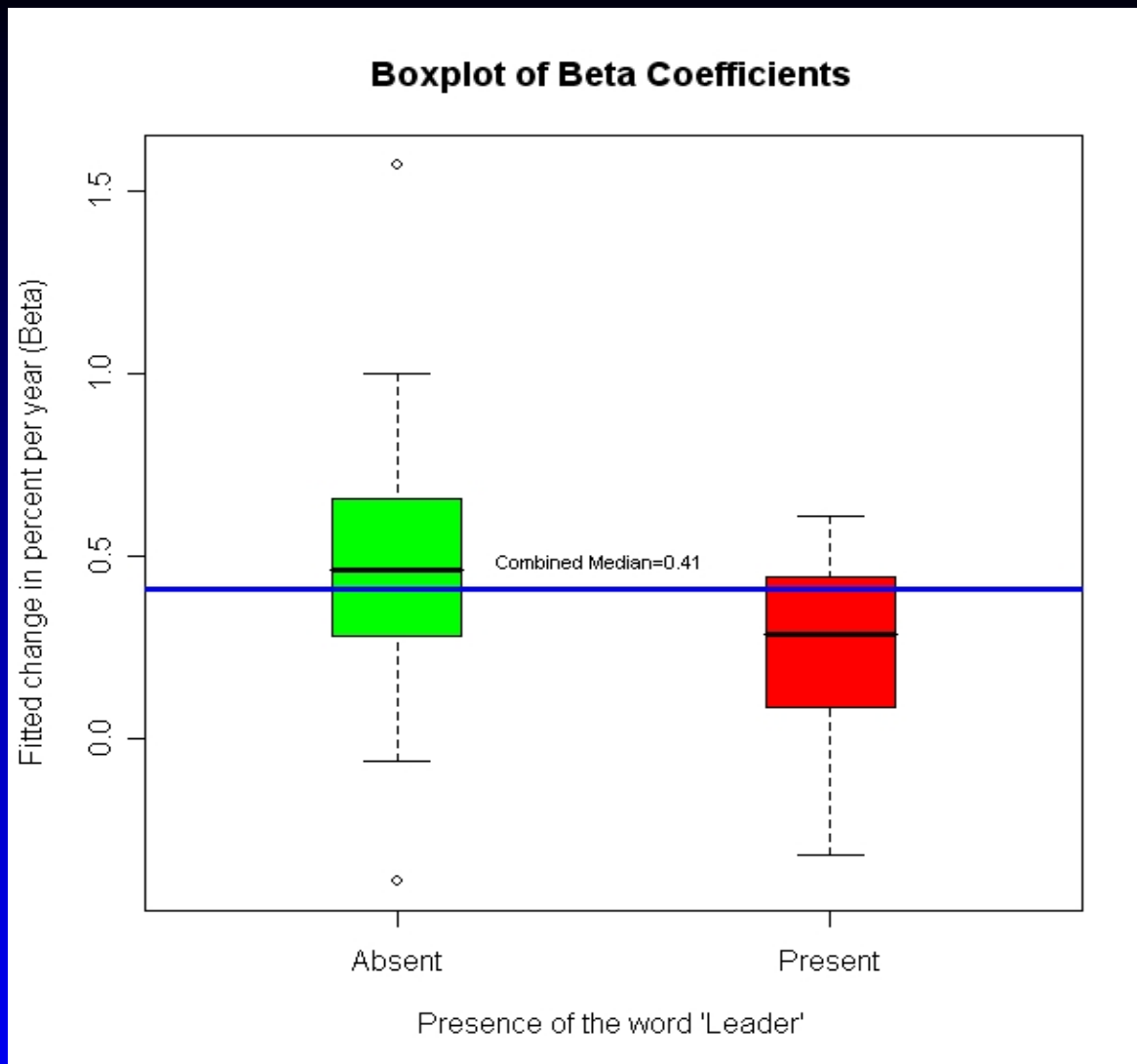


Figure 1. Box plots of beta coefficients (slopes of regression lines) for annual change in percent faculty who are tenured women over 7 years. Schools with the word “leader” in tenure criteria have significantly higher odds of having a slope below the median slope for all institutions ($p = 0.04$).

Linguistic Expectancy Bias (LEB)

Stereotype-congruent (i.e. expected) behavior is described more abstractly than stereotype-incongruent (i.e. unexpected) behavior

Stereotype	Concrete	Abstract
Men are better at lab science than women	“John dropped the beaker”	“Mary is clumsy in the lab”
Men are leaders	“She pulled together the proposal”	“He inspired the writing team”
Women are more communal than men	“Steve said hello to all the employees”	“Janet is nice”

Linguistic Category Model

- Levels of linguistic abstractness:
 - Level 1 – descriptive action verb
 - Most concrete, specific behavior
 - Level 2 – interpretive-action verb
 - More abstract, class of behaviors
 - Level 3 – State verb
 - More abstract, emotional state
 - Level 4 – Adjective
 - Most abstract, generalize across events

Abstract language reinforces and transmits stereotypes

- No effect of evaluator sex or desirability of behavior
- Writers' description of behavior more abstract when gender-congruent (expectant)
- Readers' – when behavior rated gender-congruent (expectant) \Rightarrow greater dispositional inference
- Readers' dispositional inference \Rightarrow accounted for by level of linguistic abstractness of the story

Conclusion: Expected information is communicated at a higher level of abstractness than unexpected information and this effectively maintains gender stereotypes in recipients

Clinical and Translational Science Awards (CTSA)

- PI will be elite leader:
 - Enormous institutional power
 - Massive budget up to \$70 million
 - No previous performance criteria
 - Leader of leaders: CTSA subsumes several other independent programs

We predicted that it would be unlikely for women to be represented as CTSA PIs

In fact, all 35 applications had male PIs.

Carnes & Bland Acad Med, 2007

Male applicants consistently better evaluated, more likely to be hired

- Metanalysis 49 studies
 - Davidson & Burke. Sex discrimination in simulated employment contexts: A metanalytic investigation. *J Voc Beh* 56:225-248, 2000
- Selection of musicians for symphony orchestras
 - Golden & Rouse. The impact of “blind” auditions on women musicians. *Amer Econ Rev* 90:715-741, 2000

Prejudice favoring male leaders is strong

- Subjects: German students and faculty
- Task:
 - read short description of person;
 - shown photograph (pre-tested masculine or feminine)
 - Rate 5 leadership abilities (exper 1)
 - Confidence in remembering traits in story (exper 2)
- Masculine appearing individuals (even among men):
 - More competent leaders
 - Greater false recognition of leadership competence
 - Independent of likeability