

Gender Bias in Scientific Review

Molly Carnes, MD, MS

Professor, Departments of
Medicine, Psychiatry, and
Industrial & Systems Engineering
University of Wisconsin

NIH Director's Pioneer Awards

- 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
- 6/14 second round (43%); 4/13 third round (31%) were women

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

Qx: Is it statistically likely that all 9 would go to male scientists?

Binomial probability test

Pool of potential applicants = 23% ♀

Phase 1 (N=1300) = 20% ♀

$P < .001$

Phase 2 (N=240) = 13% ♀

$P < .01$

Finalists (N=21) = 10% ♀

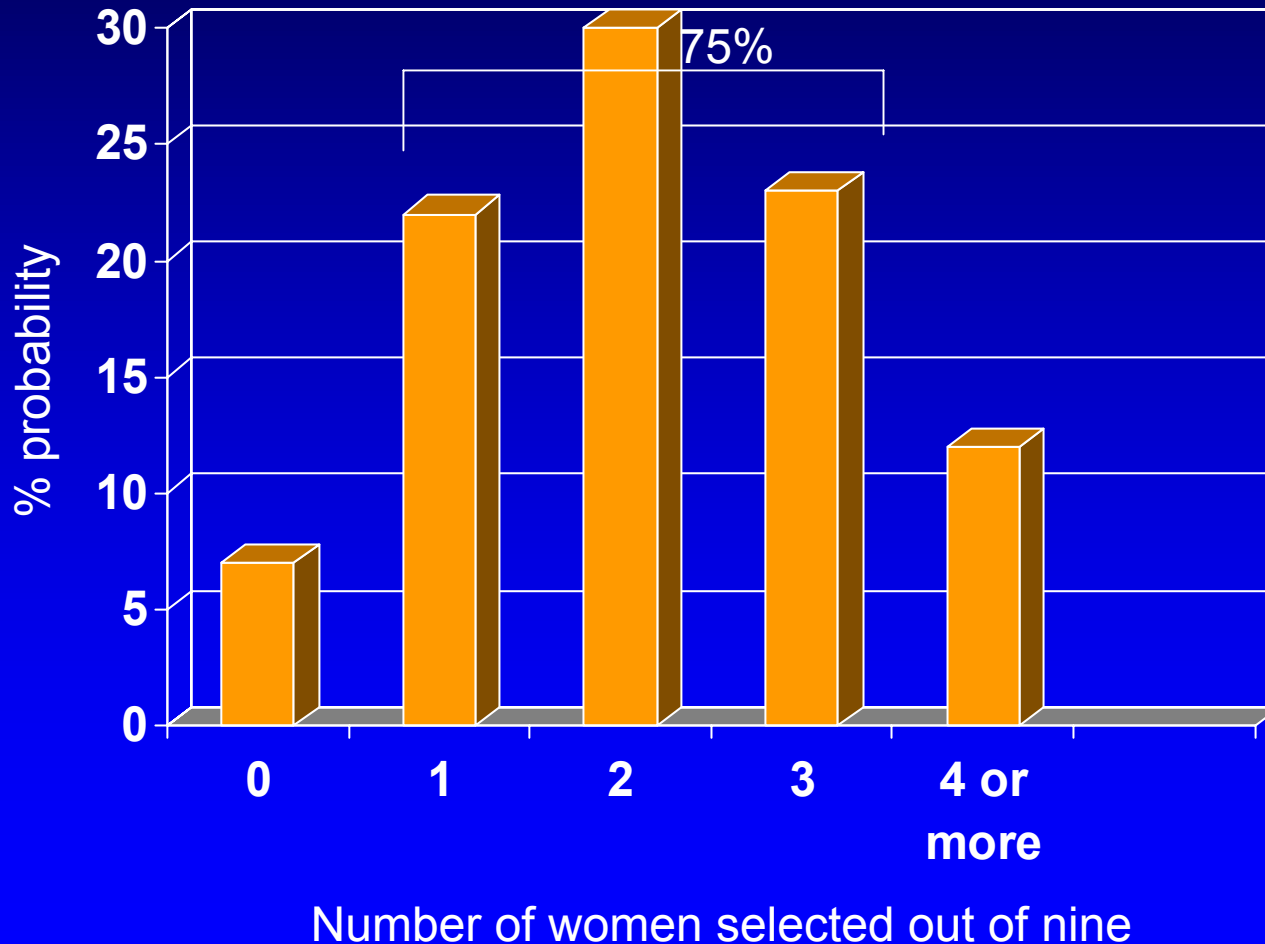
NS

Awardees (N=9) = 0 ♀

NS

Ans: Probably not

If nine people are selected from a population of equally eligible individuals in which 25% are female what is the probability of 0, 1, 2, 3, 4 or more women being chosen?



NIH Director's Pioneer Awards

	2004		2005		2006	
	N	%♀	N	%♀	N	%♀
Phase 1	1300	20	840	26	469	26
Phase 2	240	13			404	25
Interview	21	10	20	35	25	28
Final selection	9	0	14	43	13	31

Probability of occurring in a population of 25% women

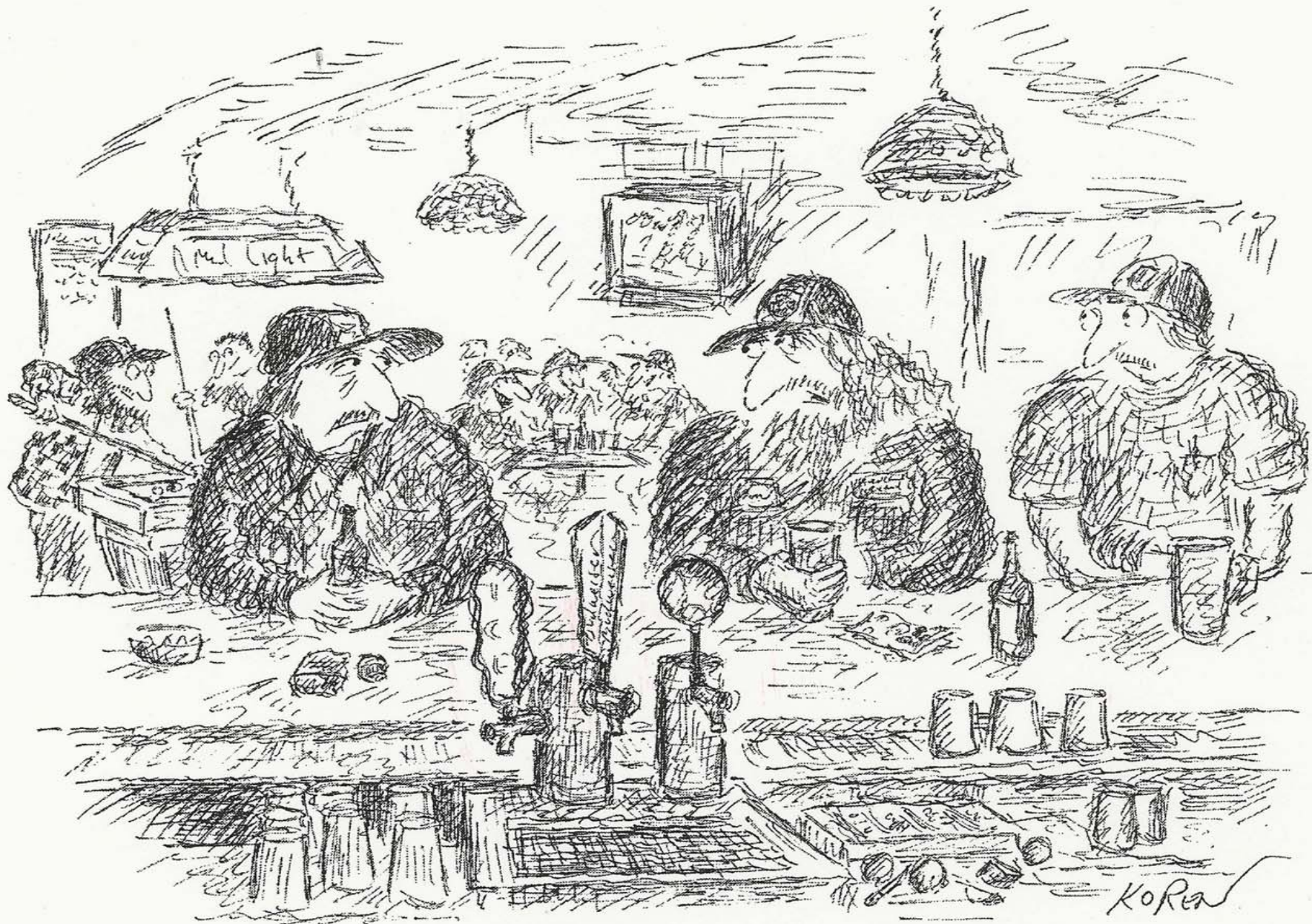
7%

7%

21%

Questions

1. Were there any differences in the solicitation and review processes between the two rounds?
2. If so,
 - Would research on gender and evaluation predict a preferential bias toward the selection of men in 2004 ?
 - Would the changes made in 2005 and beyond be predicted to mitigate this bias?



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

Research on Gender and Evaluation Relevant to Process

- Time pressure and high cognitive demand
- Impact of face-to-face review committee meeting
- Semantic priming
- Focus on intrinsic leadership abilities combined with ambiguous performance criteria
- Proportion of women in applicant pool
- Proportion of women on the review panel
- Social tuning

Background: Gender and Behavior

DESCRIPTIVE: How men and women actually behave

PRESCRIPTIVE: Subconscious 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 (also scientists and pioneers): Decisive, inventive, strong, independent
- Social penalties for violating prescriptive gender assumptions
- Unconscious gender stereotypes are easily and automatically activated

Research on Gender and Evaluation Relevant to Process

- Time pressure and high cognitive demand
- Absence of face-to-face review committee meeting
- Semantic priming
- Focus on intrinsic leadership abilities combined with ambiguous performance criteria
- Proportion of women in applicant pool
- Proportion of women on the review panel
- Counter-stereotype imaging
- Social tuning

Time pressure and high cognitive load enhance activation of automatic gender stereotypes

- 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

Conclusions

- When multi-tasking and pressed for time, evaluation defaults to prescriptive gender assumptions
- In evaluation for a male assumed job, these cognitive “short cuts” increase the likelihood that
 - Men’s evaluations will be inaccurately better
- Removing such pressures increases the likelihood that all applicants will receive a *fair* evaluation of the actual work performed

Corollary: Increasing the fairness of such a process, will decrease the current advantage afforded men

Was there a difference in time pressure and cognitive load between 2004 and subsequent rounds?

Very likely; especially in the first level of winnowing

Research on Gender and Evaluation Relevant to Process

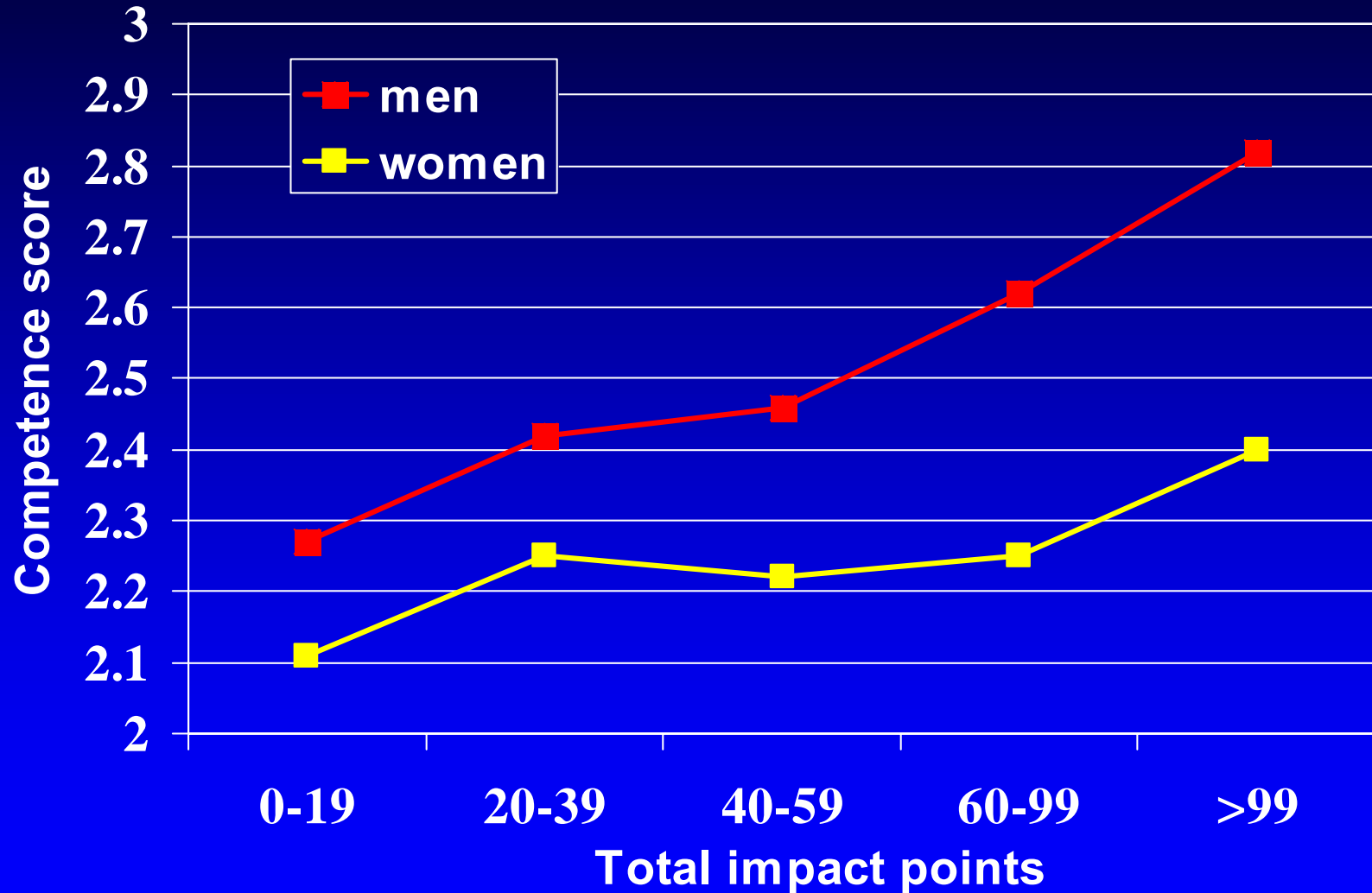
- Time pressure and high cognitive demand
- **Absence of face-to-face review committee meeting**
- Semantic priming
- Focus on intrinsic leadership abilities combined with ambiguous performance criteria
- Proportion of women in applicant pool
- Proportion of women on the review panel
- Counter-stereotype imaging
- Social tuning

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



“Friendship bonus”

- Multivariate models to test contribution of the following on competency ratings:
 - Gender
 - Productivity
 - Nationality (Swedish vs non-Swedish)
 - Field of education (e.g. Medicine, Nursing)
 - Scientific field
 - University affiliation
 - Committee to which application was assigned
 - Postdoc abroad
 - Presence of a letter of recommendation
 - Affiliation with a member of the committee (who themselves could not score) (12-13% for men and women)
- Three had influence: gender, productivity, affiliation
 - Being male worth 64 (CI: 35-93) impact points
 - Committee member affiliation worth 67 (CI: 29-105) impact points
- Being female and not knowing someone on the committee – needed 131 additional impact points

Was there a difference in face to face meeting between 2004 and subsequent rounds?

No, does not appear to have been different

Research on Gender and Evaluation Relevant to Process

- Time pressure and high cognitive demand
- Absence of face-to-face review committee meeting
- **Semantic priming**
- Focus on intrinsic leadership abilities combined with ambiguous performance criteria
- Proportion of women in applicant pool
- Proportion of women on the review panel
- Counter-stereotype imaging
- Social tuning

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

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.”

Was there a difference in semantic priming between 2004 and later rounds?

Yes

Research on Gender and Evaluation Relevant to Process

- Time pressure and high cognitive demand
- Absence of face-to-face review committee meeting
- Semantic priming
- Focus on intrinsic leadership abilities combined with ambiguous performance criteria
- Proportion of women in applicant pool
- Proportion of women on the review panel
- Counter-stereotype imaging
- Social tuning

Unconscious bias in review: Evaluation of Leadership/Competence

Prescriptive Gender Norms

- **Men**

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



- **Women**

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

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

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

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

Were evaluators told to rate applicants on intrinsic leadership qualities in 2004 but not in subsequent rounds?

Yes

Were evaluation criteria more ambiguous in 2004 vs later rounds?

Yes

Research on Gender and Evaluation Relevant to NIH Pioneer Awards

- Time pressure and high cognitive demand
- Absence of face-to-face review committee meeting
- Semantic priming
- Focus on intrinsic leadership abilities combined with ambiguous performance criteria
- Proportion of women in applicant pool
- Proportion of women on the review panel
- Counter-stereotype imaging
- Social tuning

Proportion of women in applicant pool

- Kanter RM. Some effects of proportions on group life: Skewed sex ratios and responses to token women. *Am J Sociol* 1997 82:965-990.
 - Tokens exaggerate differences and activate stereotypes
- Heilman ME. The impact of situational factors on personnel decisions concerning women: Varying the sex composition of the applicant pool. *Organ Behav Hum Perf* 1980; 26: 386-395, 1980.
 - Focal woman applicant evaluated out of group of 8
 - When women \leq 25% rated less qualified, less likely to hire, less potential for advancement
 - When women \leq 25% applicant more stereotypical female traits and this accounted for lower ratings

Were there differences in the proportion of women in the applicant pools between 2004 and later?

Yes

2004:

20%,13%,10%

2005:

26%,35%,43%

2006:

26%,28%,31%

Research on Gender and Evaluation Relevant to NIH Pioneer Awards

- Time pressure and high cognitive demand
- Absence of face-to-face review committee meeting
- Semantic priming
- Focus on intrinsic leadership abilities combined with ambiguous performance criteria
- Proportion of women in applicant pool
- Proportion of women on the review panel
- Counter-stereotype imaging
- Social tuning

The power of numbers in influencing hiring decisions.

Yoder et al. *Gender and Society* 3:269-76, 1989

- Examined hiring of academic psychologists across the country (N = 93 acad positions)
- In depts with < 25% women, men more likely to be hired
- In depts with 36-65% women, men and women equally likely to be hired

Was there a difference in the proportion of women on the review committee between 2004 and later rounds?

Yes

4/64 (6%) vs. 28/64 (44%) vs. 32/79 (40%)

Research on Gender and Evaluation Relevant to NIH Pioneer Awards

- Time pressure and high cognitive demand
- Absence of face-to-face review committee meeting
- Semantic priming
- Focus on intrinsic leadership abilities combined with ambiguous performance criteria
- Proportion of women in applicant pool
- Proportion of women on the review panel
- Counter-stereotype imaging
- Social tuning

Social influence effects on automatic racial prejudice.

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

- 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

2004

2005, 06

Encouragement to specific groups of scientist to apply

No specific encouragement for women and members of underrepresented groups to apply.

Addition of encouragement to apply:

- “Those at early to middle stages of their careers, and women and members of groups underrepresented in biomedical research are especially encouraged to nominate themselves.”

Was social tuning to reduce gender bias more likely to be present in later rounds vs 2004?

Yes

Summary

Feature of process	Predict preferential selection of men	Present in 2004	Present in 2005, 06
Time pressure/no meeting	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 < 25% of reviewers	Yes	Yes	No
Social tuning	No	No	Likely

No. women/total awards (%)

0/9 (0)

6/14 (43);

4/13 (31)

Conclusions

- Even the most objective scientist is susceptible activation and application of unconscious gender stereotypes
- It appears bias in evaluation can be mitigated
- We applaud NIH for evidence-based approach
- We encourage others in similar self-study