

# Evaluations and Equity



**JEAN STOCKARD  
UNIVERSITY OF OREGON**

# Sociology of Science



- Merton – “functional goal of science” – the expansion of “true” and secure knowledge
- Ethos (culture) of science – norms and values
  - Ideally, is universalistic (allocation of rewards based on merit)
  - Not particularism – when allocation of rewards is based on individual characteristics apart from merit
- Legitimacy and trust is key

# Sociology of Professions



- Monopoly on knowledge, highly trained, long period of apprenticeship
- Self-control (judgment by peers)
- Peer review thought central to maintaining integrity of discipline and promoting quality, innovation, and fairness
- Matthew effect – inequalities within professions

# Some Sociological Maxims



- Environments → behaviors and attitudes
- Think like a scientist (a social scientist), not a philosopher.
- Everything can be subject to empirical test.
- We can never be 100% sure of our results, but repeated patterns promote strong confidence.
- Sociological knowledge can make the world (and science) better (better quality, more innovative, and fairer to all).

# Peer Review – Today's Focus



- What is it, why and when used?
- Are there different types of peer review?
- What is the empirical evidence regarding its relationship to innovation, quality, and fairness of the profession?
- What is evidence related to legitimacy and trust?
- Are there implications of these results?
- A disclaimer

# Defining Peer Review



- **Judgments by those with equal or greater professional knowledge**
- **Used in all areas of professional life**
- **Defines winners and losers – may contribute to “Matthew effect”**
- **Serves as incentive for nature and content of individuals’ scholarly work**

# Types of Peer Review



- **Single blind**
- **Double blind**
- **No blinding**
- **Open reviewing**
- **Combinations**

# Ideal Functions of Peer Review



- Promotes quality, innovation, fairness (universalism)
- Systematic nature (rules norms – the ethos of science) engenders trust/legitimacy
- Determines dominance of paradigms and perspectives (site of power struggles) and thus potentially controls pace of innovation



# Empirical Evidence on Quality (a)



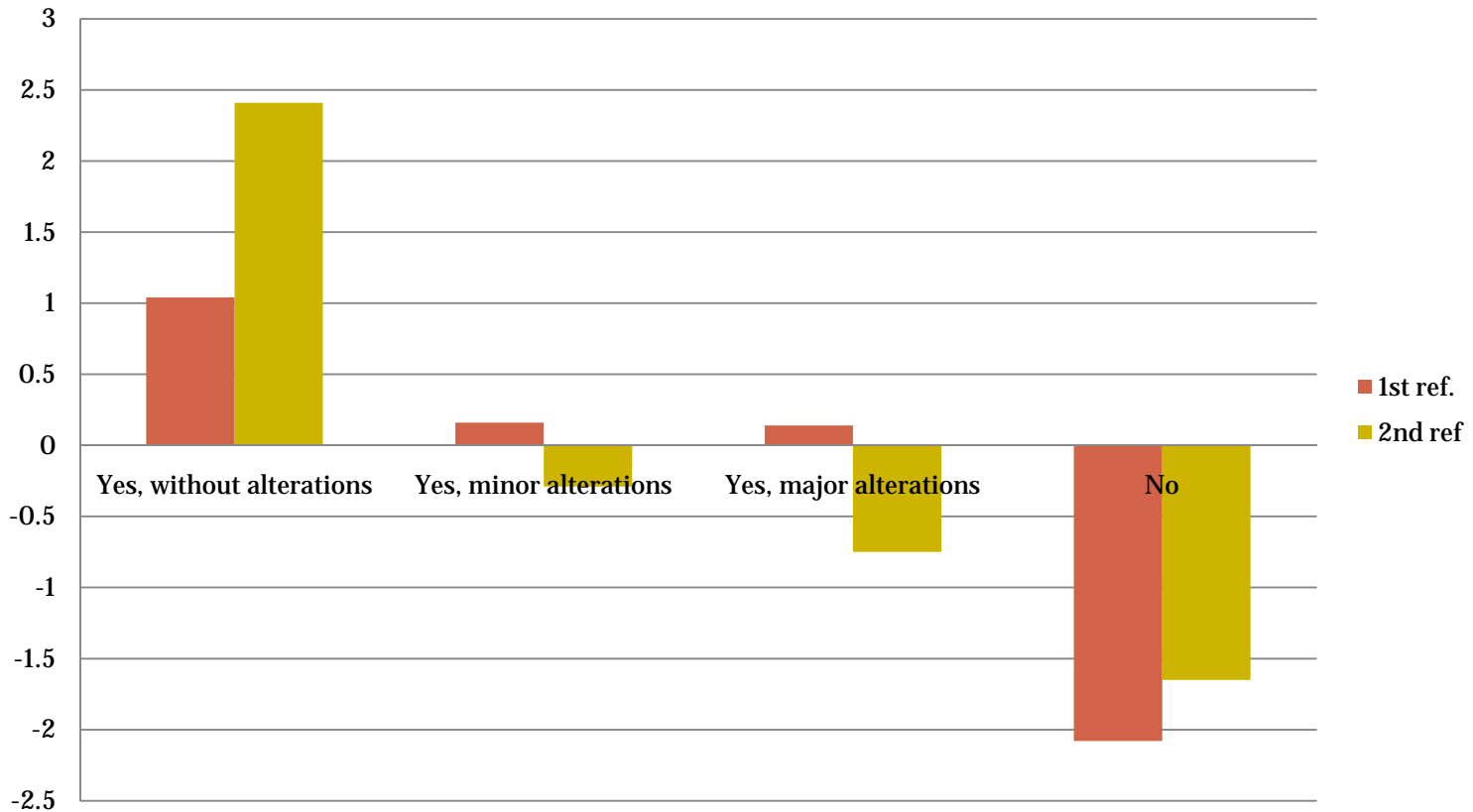
- Studies of grant submissions and article reviews
- Inter-rater reliability: agreement, but not perfect in reject/accept decisions (r between .2 and .4)
- Error detection: variability across reviewers, none perfect, and agreement not strong.

# Empirical Evidence on Quality (b)



- Predictive validity – citation counts
- Comparisons of accepted articles in journal A with those rejected by journal A, but later published elsewhere
- Good set of controls including length of exposure
- Those accepted originally had higher citation counts
- Example of results from Daniel's (1993) study of *Angewandte Chemie*

### Mean citation rate (log transformed), adjusting for time since publication, by reviewers' recommendations



# Empirical Evidence Suggests Peer Review Enhances Quality



- Reviewers agree more than they disagree
- Articles accepted more readily have larger impact on discipline (as shown by citation counts)

# Empirical Evidence re. Innovation



- Originality/innovation is highly valued
- But data indicate strong resistance to results that counter beliefs and views – confirmatory bias –
- George Akerlof's article on uncertainty: Rejected by three journals before publication, but won 2001 Nobel Prize in Economics 34 years later

# Empirical Evidence re. Fairness of PR



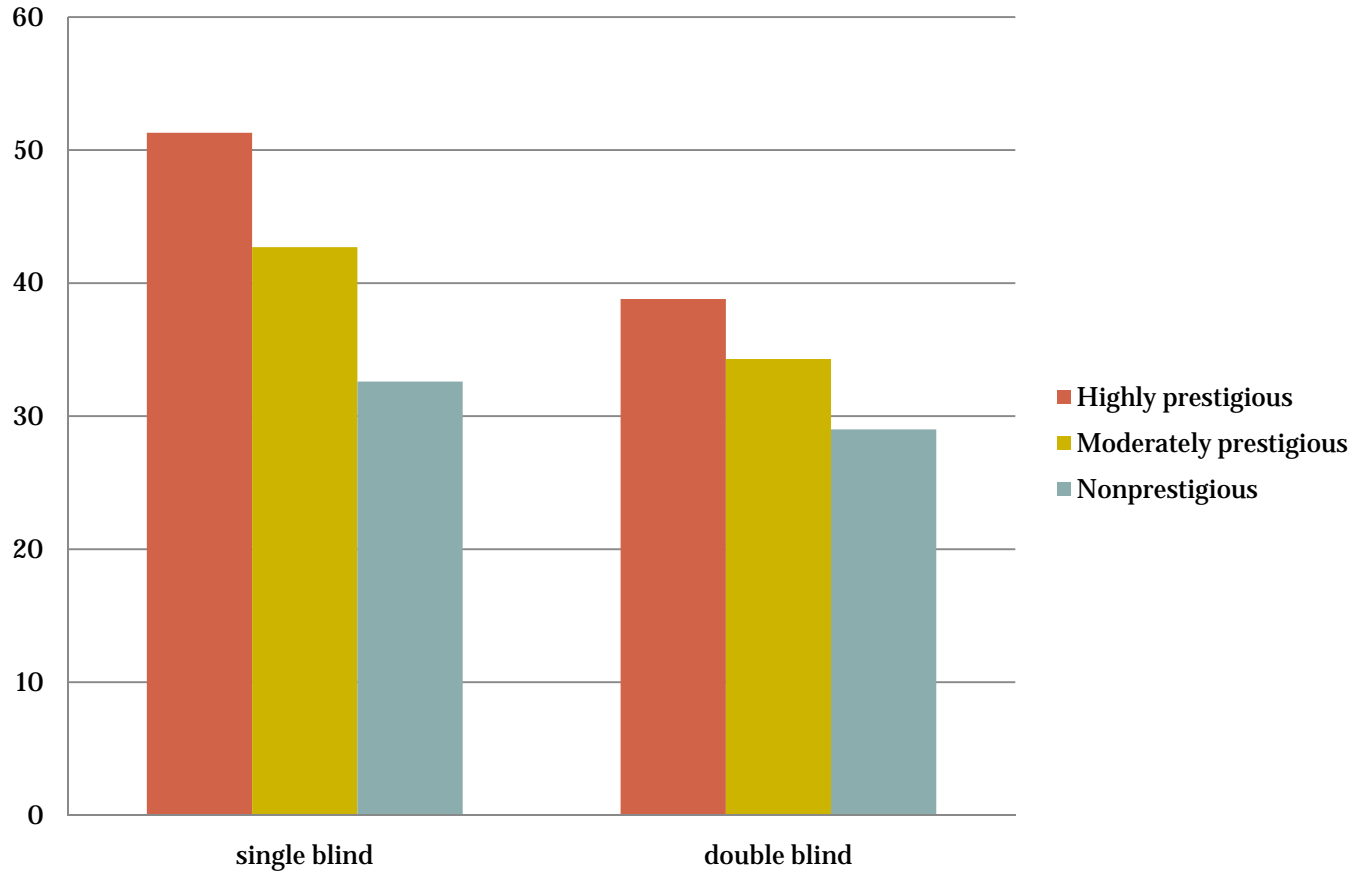
- To what extent do peer review systems meet the ideal norm of universalism?
- Studies usually compare single blind and double blind review processes
- Most evidence relates to three dimensions:
  - Prestige of individual/institution
  - National origin/language
  - Gender

# Evidence of Bias Related to Prestige



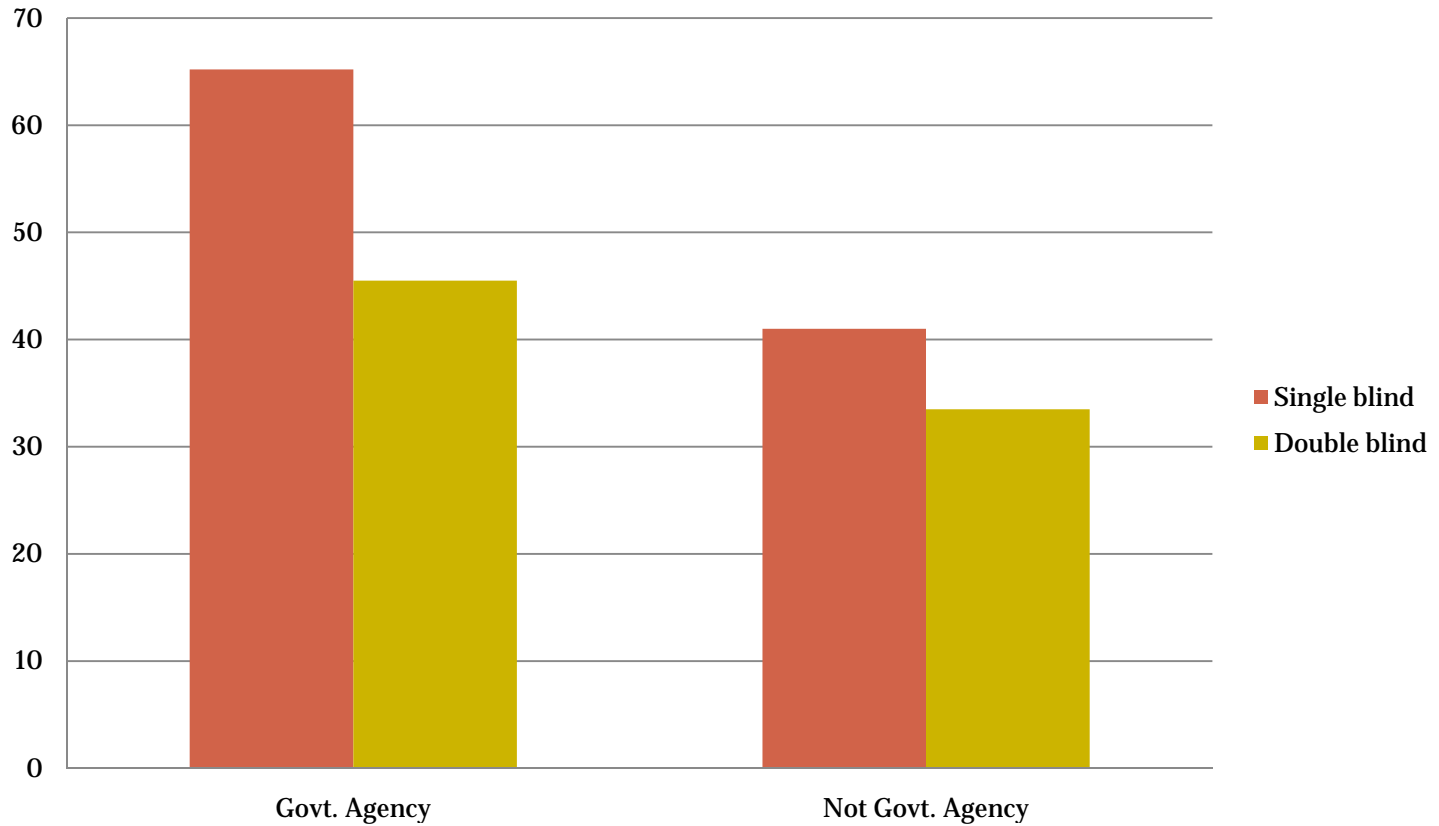
- **Variety of methods lead to similar results**
- **Resubmission of 13 already published articles using less prestigious names and affiliations** (Ceci and Peters, 1982)
- **Comparing fate of research abstracts submitted to professional meetings in years with different procedures** (Ross et al, 2006)
- **Random assignment of articles submitted to one journal to either double or single blind** (Blank, 1991)

### Acceptance Rate of Abstracts at AHA Meetings by Prestige of U.S. Institutions and Review Type





### Acceptance Rate of Abstracts at AHA Meetings by Government Agency Affiliation (U.S.) and Review Type

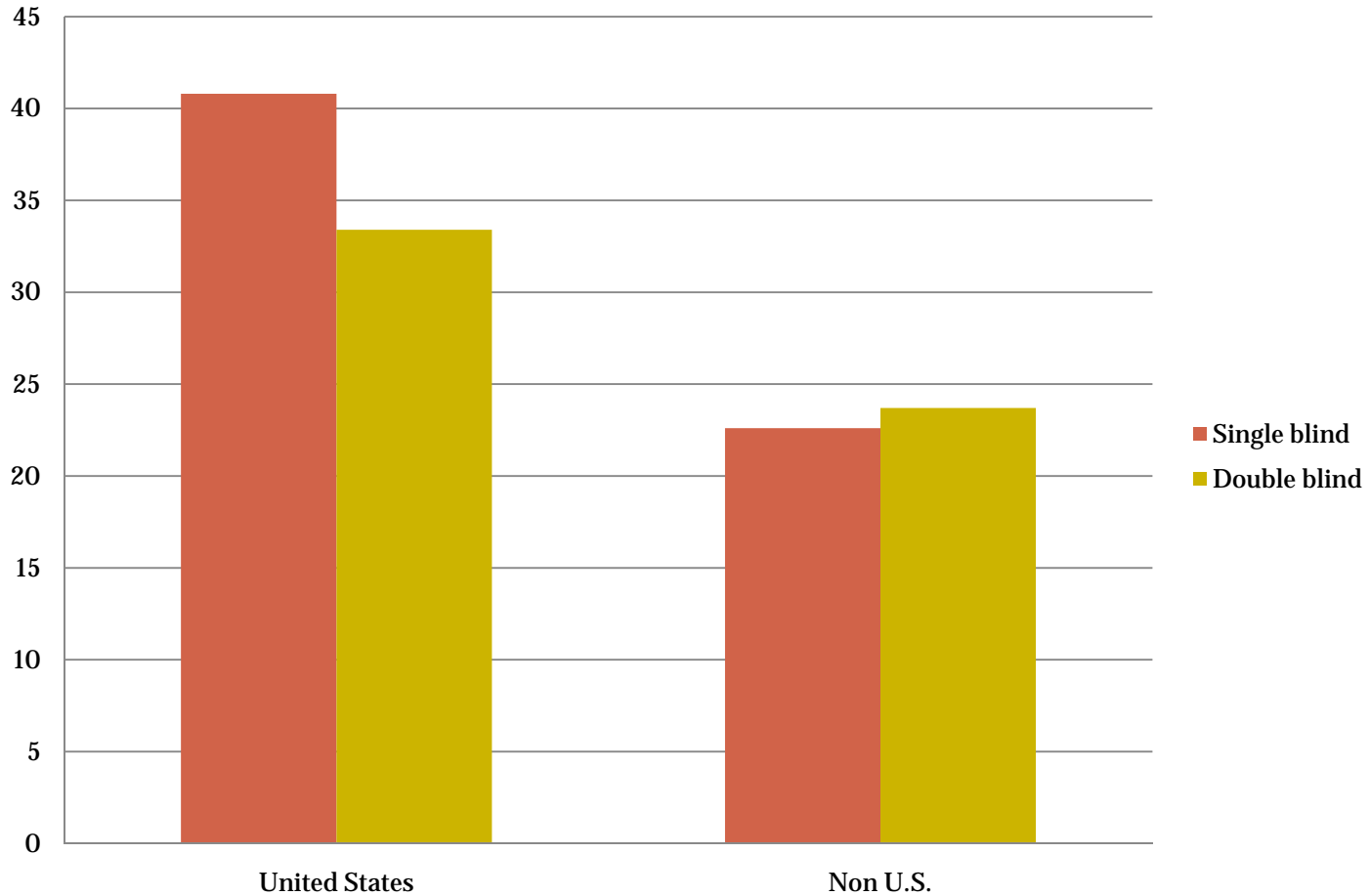


# Evidence of Bias Related to Country of Origin/Language

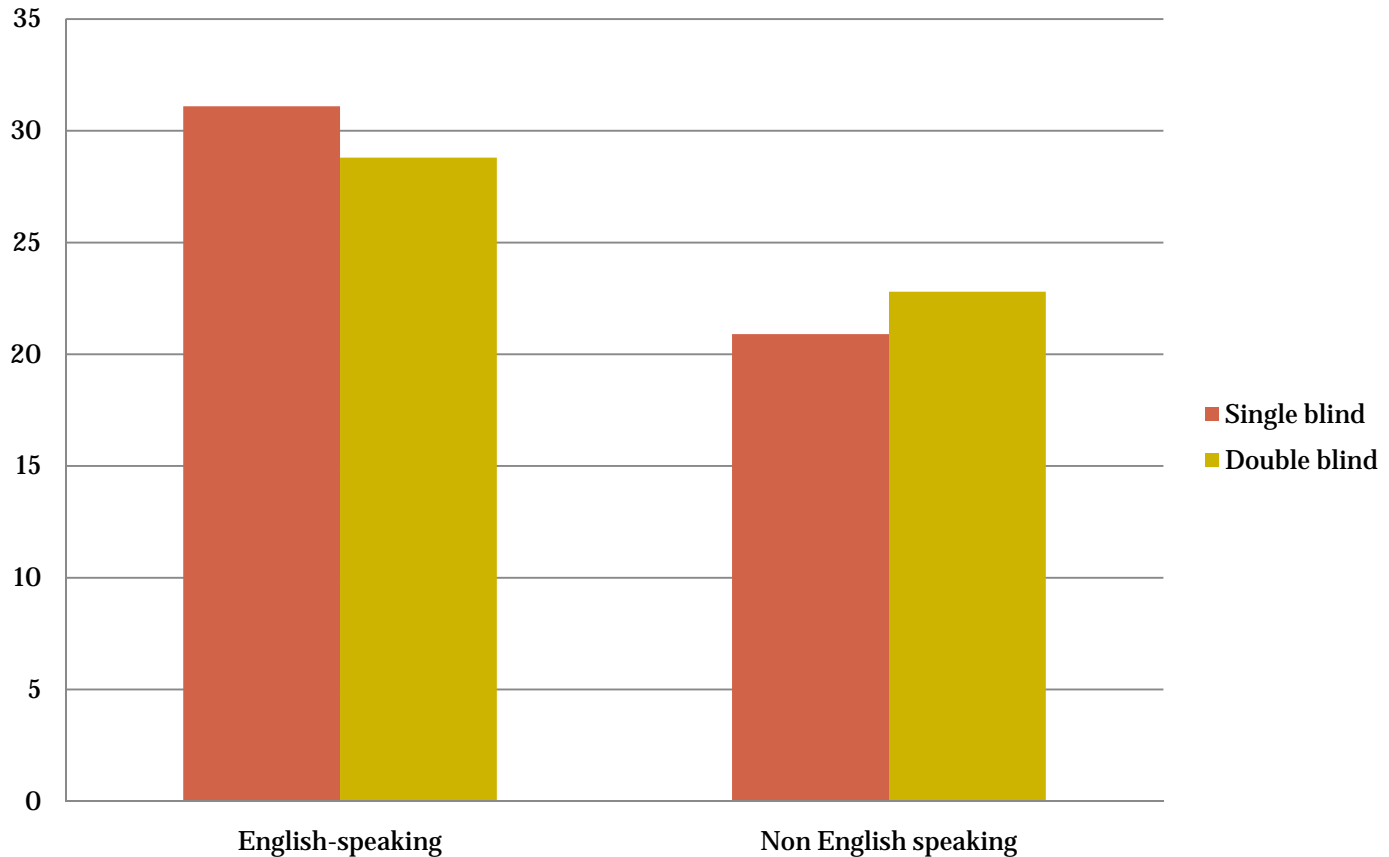


- **Studies mainly from sciences and medicine**
- **Preference for country in which journal is based (for Daniel's 1993, AC study)**
- **Preference for those from higher income countries**
- **Among those from higher income countries, preference for those from English speaking countries**

### Acceptance Rate of Abstracts at AHA meetings by Country of Author and Review Type



### Acceptance Rate of Abstracts at AHA Meetings by English Speaking Status of Countries (non-U.S.) and Review Type



# Evidence of Bias Related to Gender



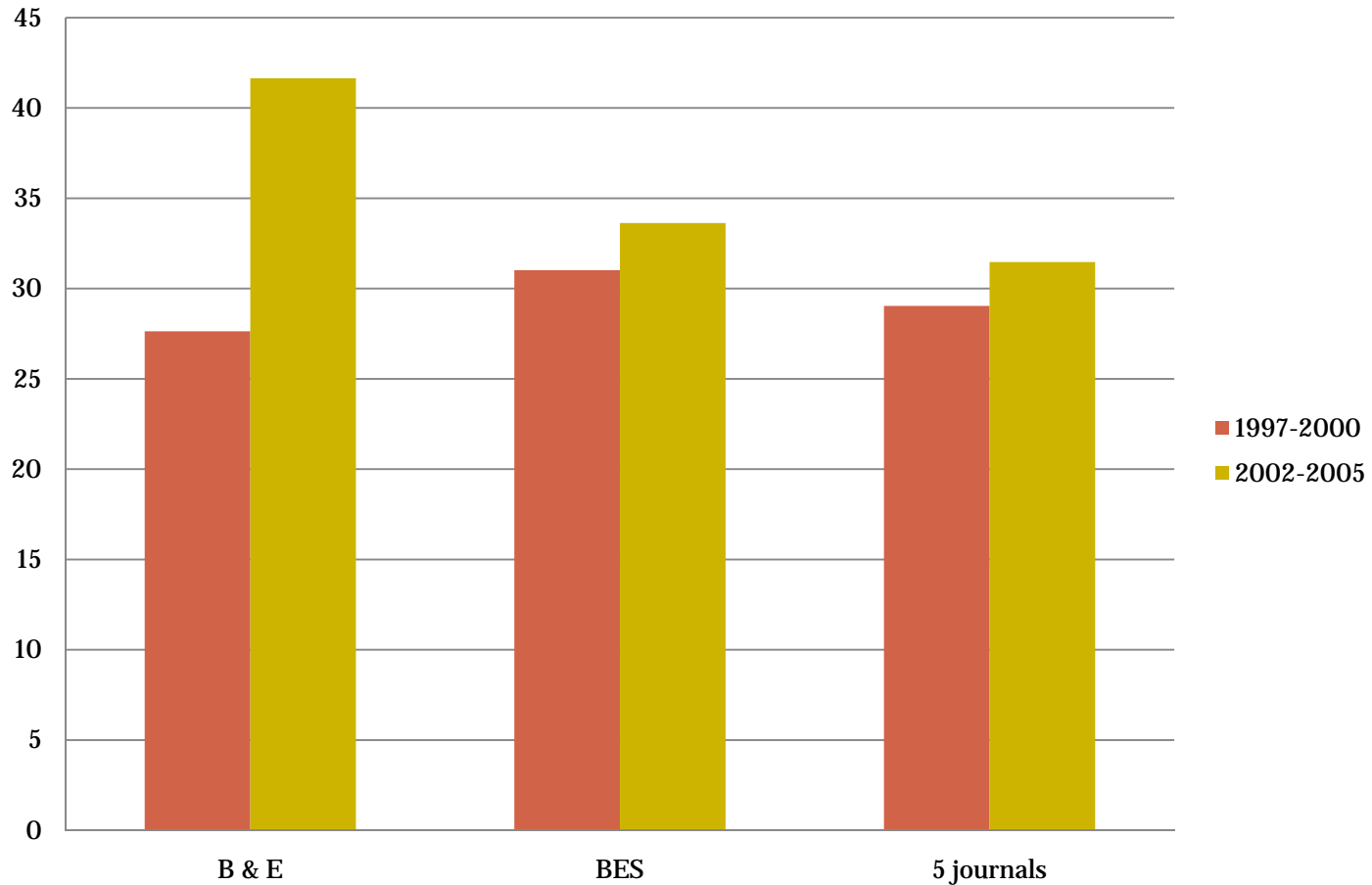
- **When differences appear, females are disadvantaged**
- **Appears in studies of both journal reviews and grant applications**
- **Appears with different methodological approaches**
- **Examples**

# Example 1: Changes in Female Authorship after Double Blinding Introduced



- Behavioral Ecology introduced db policy in 2001
- Compared representation of female authors in BE before and after the change with journals using s.b.
  - A journal with similar subject matter and impact factor (Behavioral Ecology and Sociobiology)
  - Four additional journals in the field (Budden et al, 2008)

**Percentage of Articles by Women (first authors) pre and post double blind policy, B & E and comparable journals**



## Example 2: Multiple Studies Document Differences in Grant Rewards



- **Meta-Analysis of gender differences in grant peer review**
- **21 studies, 66 analyses**
- **Men's advantage over women averaged about 7%; (with overall approval rate of 50% expected male rate = 52%, female rate = 48%)**
- **Accumulation of effects over time**



## Example 3 – Gender Differences in Grant Funding Not Explained by Differences in Productivity



- Used freedom of press legislation to get access to peer review scores on Swedish Medical Research Council funding decisions
- Strong controls for scientific productivity of applicants
- Strong impact of gender
- Strong impact of knowing someone on review committee (even though didn't participate in decision) (Wenneras & Wold, 1997)

# Summary of Evidence on Bias



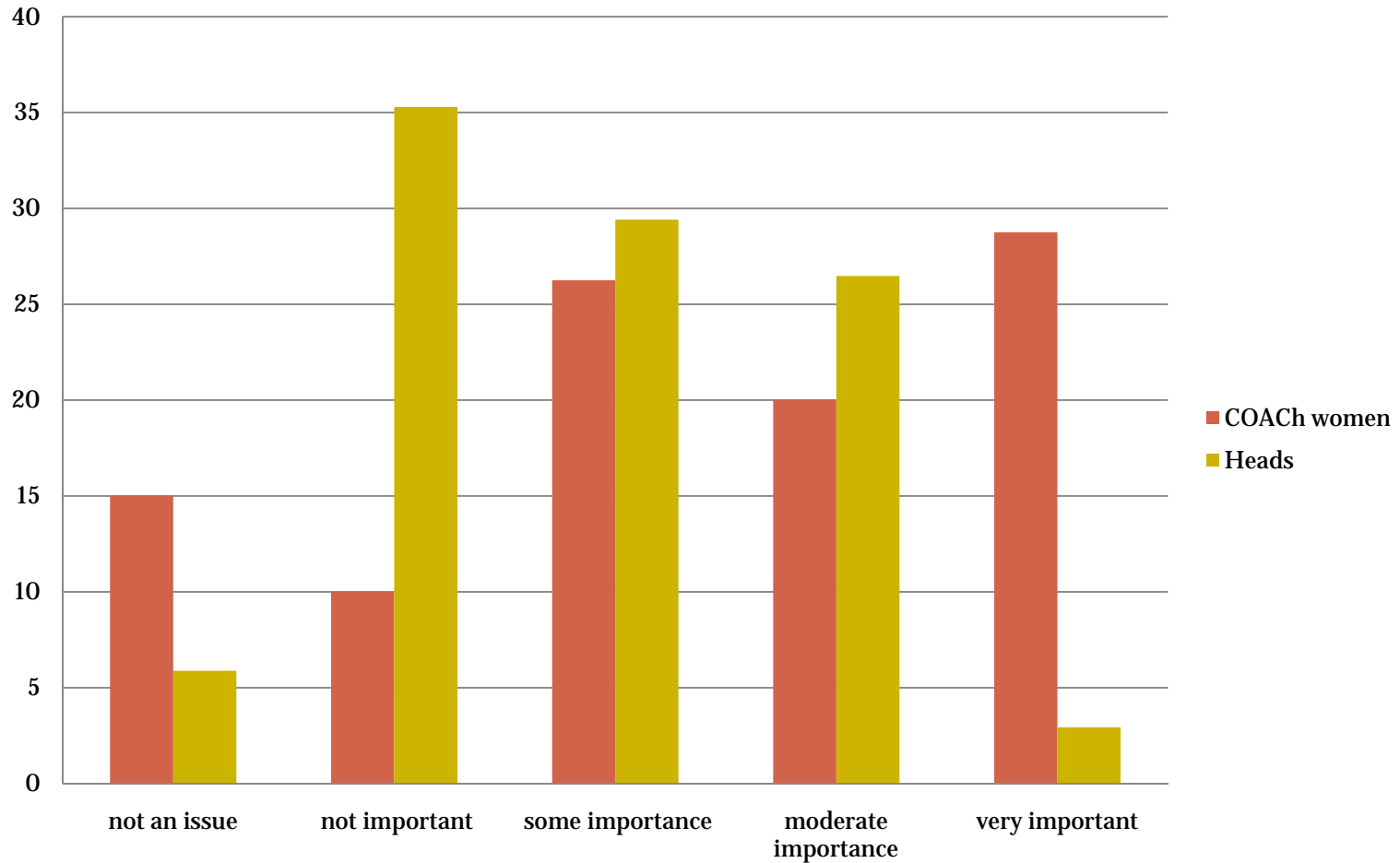
- When differences appear indicate single blind reviewing practices less universalistic than double blind
- Most evidence related to national origin, institutional affiliation, and gender
- Less evidence (or none that I know of) regarding other characteristics

# Legitimacy – What Do People Think: An Example from Surveys

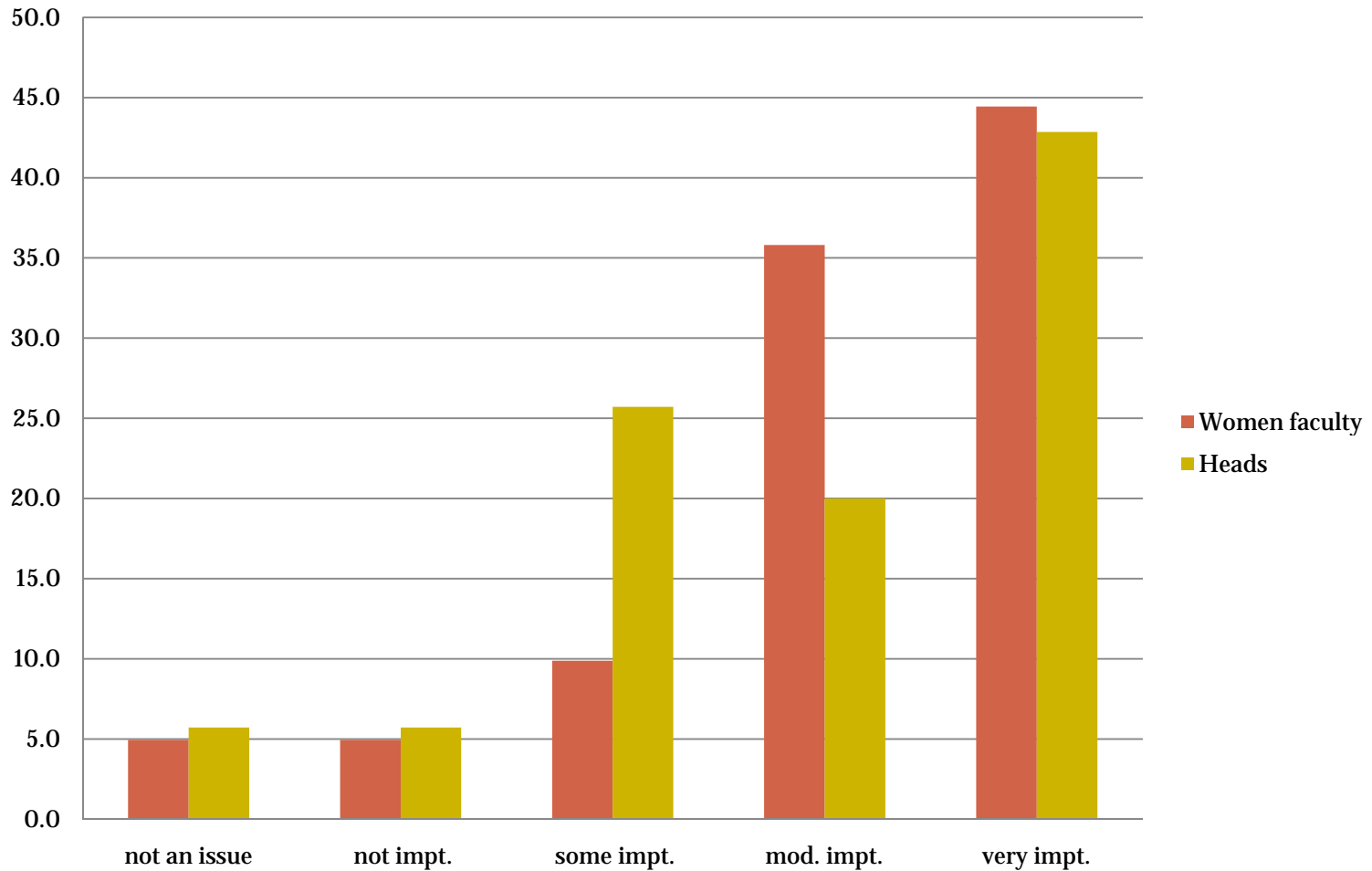


- Two sources of data on chemists
- Women, from all parts of the discipline who attended 4 different COACh workshops
- Chemistry department heads, who participated in equity workshops a few years ago
- Views regarding factors that affect women's career progress

### **% of COACH Women and Heads Saying Gender Discrimination in Peer Review Process of Papers and Grants Hinders Career Progress**



### **% of COACH Women and Heads Saying Subtle Biases that Accumulate Hinder Women's Career Progress**



# Is there a legitimacy problem?



- Over half of both the women and heads said that gender discrimination in peer review was at least somewhat important in hindering women's career progress (75% of women and 59% of heads)
- Almost all (90% of women and 89% of heads) say that the “Matthew effect” is at least somewhat important in hindering women's career progress
- Similar results with race-ethnic heads workshop: Over one-third (38%) said race-ethnic disc in peer review very or somewhat important in hindering career progress of minorities and all cited subtle biases that build up over time.

## Example 2: Statement of World Academy of Young Scientists



- 2005 statement from 2004 general assembly meeting: “We reached a consensus that slight modifications to the current review process would help in getting more objective reviews based on the quality of the research rather than the age, affiliation, gender or pedigree of the authors.”
- Explicitly called for altering the single-blind peer review system

# Continuation of WAYS Statement



- “Maintenance of trust within the international scientific community is crucial, not only for future scientific development, but also to continue the dialogue of civilizations....We propose here that DBPR is a better system because, in addition to being a reasonably fair process, it also bears symbolic power that will go a long way to quell fears and frustrations, thereby generating a better perception of fairness and equality in global scientific funding and publishing. This will, in turn help to keep research more accessible for future generations.” (Maingay, et al, 2005)



# Possible Objections to DBPR



- **Can't really blind**
- **Editors will object because of cost (time and \$)**
- **Quality will be hurt**
- **Reviewers and authors will object**
- **Probably others.....**

# Thinking like a social scientist and not a philosopher



- All of these are empirical questions
- Other disciplines – after conducting research have changed practices.
- E.g. only one economics journal is now sbpr, while in 1991 the majority were. Almost all social science journals have switched, as have many medical journals – and, of course, symphony orchestras.

# Questions – Comments???



**THE FLOOR IS YOURS.**