JUDGING OTHERS IN THE ACADEMY: IMPLICATIONS OF UNCERTAINTY AND BIAS

Abigail J. Stewart
Joint meeting of the Philosophy of Social Science Roundtable and European Network for the Philosophy of the Social Sciences
Seattle, May 9, 2015
Judgments of others in the academy

- Ubiquitous
  - Many informal judgments
  - Formal, consequential judgments
    - Of particular work (papers, books, grant proposals)
    - Of scholars over course of lifetime

- Confident
  - Educated
  - Experienced
  - Informed
Reasons for less confidence

- Academic judgments of others often include an element of forecast
  - Admission of students
  - Hiring, tenure and promotion of colleagues

- Always uncertainty in forecasts
  - Scientific predictions
    - Where particle or planet “is”
    - Weather
    - Earthquakes
  - Social phenomena
    - Elections
    - Outcomes of athletic competitions
    - What people will do under various circumstances (e.g., obedience)
  - Our own feelings ("affective forecasting")
    - Many regularities in our “errors”
    - Can reduce them through improved observation of detail
    - Gaps remain
Reasons for less confidence

Threats to the validity of our judgments of others

1. Our judgments of “merit” are subject to a variety of influences that are not intrinsic to assessing merit.

2. Conditions can reduce reliance on these extrinsic factors, but

3. Difficult to reduce them to zero and

4. Particularly difficult in areas in which we are experts especially confident.
What do we do about this?

- Suggestions
Threats to the validity of our judgments of others
Fundamental attribution error

- Observed by Jones and Harris in 1967 study of attitudes about Castro

<table>
<thead>
<tr>
<th></th>
<th>Assigned Position</th>
<th>Chose Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pro-Castro</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anti-Castro</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Attributed opinion to personal attitude more when *chose* position than when *assigned* BUT

*Attributed to personal attitudes even when position was assigned*

- Over-attribute of causes of outcomes to person or disposition and under-attribute to situation

  “It seems that behavior in particular has such salient properties it tends to engulf the total field rather than be confined to its proper position as a local stimulus whose interpretation requires the additional data of a surrounding field.” (Heider, 1958)

Is the problem limited information?

- Is behavior salience really the problem?
- Judges were given little information; maybe lack of information is cause
- Recent study of impact of GPA on admissions judgments with information on leniency of grading at institution
  - GPA has strong positive impact on admissions decision
  - Leniency of grading at institution has some, but little effect

Many derivative observations

- **Just world theory**
  - People deserve the outcomes they get

- **Blaming the victim**
  - Victims cause their own victimization


Relevance to academic judgments?

Demonstrated tendency in wide range of circumstances to underestimate the impact of situational factors on individuals’ behavior and outcomes

- attribute causes of success to person
- also attribute causes of failure to person
Reliance on group-based schemas
Schemas: Unconscious Hypotheses

- Schemas (expectations or stereotypes) influence our judgments of others regardless of our own group.
- Schemas also influence group members’ expectations about how they will be judged.

The Implicit Association Test: A simple and convincing way to explore the effects of your own schemas.

Take it at: implicit.harvard.edu
What do we know about schemas?

- Schemas allow rapid, but sometimes inaccurate, processing of information.
- Schemas often conflict with consciously held or “explicit” attitudes.
- Schemas change based on experience and exposure.

Who uses schemas? When?

- **Widely culturally shared**
  - Both men and women hold them about gender.
  - Both whites and minorities hold them about race.
  - People usually do not consciously apply them.

- **Applied more** under circumstances of:
  - Stress from competing tasks
  - Time pressure
  - Lack of critical mass
  - Ambiguity/including lack of information


Present in many academic judgment contexts: admissions, hiring, promotion reviews, etc.
Schemas affect evaluation: Resume studies

Numerous studies show that schemas affect evaluation

A common method: paired resume studies

- Two resumes are prepared – identical except for one item (name, etc.)
- Each evaluator receives one resume to evaluate at random
- Average evaluations compared
Evaluation of identical CVs

For a faculty position:
- Male and female psychology professors recommended for hire “Brian” over “Karen” as an assistant professor (2:1).

For an undergraduate lab manager position:
- Male and female science professors rated male applicants more competent, more hireable, more suitable for mentoring, and offered higher salaries.

For sales, administrative support, clerical and customer services positions:
- Similar findings for “Jamal” and “Greg”

Evaluation of identical resumes: Sexual identity

- Pairs of matched resumes sent for 5 different occupations in 7 different states
  - Overall, 40% fewer call backs for gay applicants
  - Largest difference in Ohio, Texas, Florida (as compared to California, New York, Nevada and Pennsylvania)


Additional Resource:
When evaluating equally qualified same-gender job applicants,

Mothers...

- were rated as **less** competent and **less** committed to paid work than nonmothers.
- were **less** likely to be recommended for hire, promotion, and management, and were offered **lower** starting salaries than nonmothers.

Fathers...

- were rated as **more** committed to paid work than nonfathers.
- were offered **higher** starting salaries than nonfathers.

Evaluation of fellowship applications: Gender

“...the success rate of female scientists applying for postdoctoral fellowships at the [Swedish Medical Research Council] during the 1990s has been less than half that of male applicants.”

Women applying for a post-doctoral fellowship had to be 2.5 times more productive to receive the same reviewer rating as the average male applicant.

Similar findings:
- USA/GAO report on Peer Review in Federal Agency Grant Selection (1994)
- European Molecular Biology Organization Reports (2001)

Race penalties in grant success

- 83,188 NIH grant applications from 40,069 individuals from 2000-2006.
- Differences in funding rate persists even after controlling for education and training, previous NIH experience, research productivity, and other factors.

Leadership outcomes

Positions of Leadership for Asians/Asian Americans

Stereotypes about East Asians show that competence is accepted, but warmth and dominance are not.


Being “white” is positively associated with leadership

What happens when the influence of schemas is truly eliminated?

During the period from 1970-1996 many US symphony orchestras began conducting screened auditions.

Outcomes changed!

- Data from 14,000 applicants over time show that the use of a screen increased the probability that a woman would advance from preliminary rounds by 50%.

Can’t easily remove influence of schemas

- Screens help with musical auditions
- “Blind review” may help (but so many internal cues make this difficult)
  - Evidence people search for cues even when blind
- More information does help
  - “individuating” information reduces reliance on schemas
  - schemas still have an effect
SeparaDng impact of labels from individuaDng information

- Participants rated transcripts of 3 online conversations with 6 people described variously in terms of race and gender
  - Person A had 6 different race-gender identities, etc.
- Also collected data on average ratings of “typical person of a race-gender group” (no transcript)
- Ratings of the personality characteristics of the 6 persons reflected both
  - Category labels for gender (but not race)
  - IndividuaDng information

Many other irrelevant factors influence judgments
Halo effects

“suffusing ratings of special features with a halo belonging to the individual as a whole” (Thorndike, 1920, p. 25).

- Influence of appearance/likeability etc. on ratings of other domains
- Influence of one performance on judgments of other performances
- Influence of own mood on judgments
Impact of positive mood on essay ratings

- Autobiographical mood induction (happy, neutral or sad memory)
- Social judgment task
  - Evaluate essay by “Robin Taylor”
    - Rated middle-aged male
    - Young female
    - Rated as equally likable
    - Male much more likely to be a philosopher than female; also more competent and more intelligent
- Evaluations of essay
  - More positive if positive mood
  - Expected halo for older male expanded by positive mood

Mood exaggerates halo effects on evaluation of the essay.

Figure 1. Mood moderates halo effects on the evaluation of an essay: positive mood increased and negative mood eliminated the halo effect.
Reliance on prestige

- Special case of halo effects
- Prestige of institution predicted fate of resubmitted papers

- Structure of faculty hiring networks
  - 461 doctoral departments in North America (history, computer science, and business)
  - 19,000 faculty in those departments
    - 86% had received doctorate at one of the sample departments
    - 25% of the institutions produced 71-86% of the faculty
    - Top 10 institutions produce 1.6 to 3 times more faculty than the next 10, etc.
    - Only 9-14% of faculty are placed at institutions more prestigious than their degree


Influence of institutional prestige on licensing of technology

- All measures of previous performance predict licensing
- University's general prestige increases its licensing, controlling for:
  - Volume of invention disclosures
  - Regional amount of high-technology activity
  - Number of universities within a county
  - Source of university funding
  - Presence or absence of a medical school
  - Year
  - University's past licensing performance

Reliance on networks

- Networks play an important role in scholarly circles
- Homophily: “similarity breeds connection”
  - Affects nature of networks
    - Women have more women in their networks
    - Men have more men

**Reliance on networks**

![Bar chart showing gender homophily in network ties](image)

**Fig. 1. Gender Homophily in Network Ties**

Evaluation of fellowship applications: Gender

“...the success rate of female scientists applying for postdoctoral fellowships at the [Swedish Medical Research Council] during the 1990s has been less than half that of male applicants.”

Women applying for a post-doctoral fellowship had to be 2.5 times more productive to receive the same reviewer rating as the average male applicant.

“Friendship bonus”: applicants affiliated with a committee member rated higher. Friendship and Gender had similar size effects.

Similar findings:
- USA/GAO report on Peer Review in Federal Agency Grant Selection (1994)
- European Molecular Biology Organization Reports (2001)

Persistent impact of “nepotism” in peer review

- 10 years later estimated impact of gender and reviewer affiliation on research grants in Medicine
  - Gender no longer significant after practices altered
  - Reviewer affiliation equally significant

Belief in our own expertise at decision-making is a problem

- Tetlock study of expert political pundits’ predictions
  - Worse than chance
  - Specialists worse than non-specialists
  - Resisted admitting wrong

- Kahneman
  - Experts rely on intuition more than deliberate process and have high confidence
  - Intuition is reliable when based on
    - An environment that is regular and predictable (a “high-validity” environment)
    - Opportunity for prolonged practice
  - Fields differ
    - Compare Anesthesiology vs Radiology


Belief in “objectivity”

- Some participants (adult men) primed by answering questions about their objectivity (over 88% rated selves as above average in objectivity)
- Resume study approach (rate file of Gary or Lisa)
- No difference in hiring recommendation in control condition;
  men favored over women in the “objectivity” condition

Beliefs about requirements for success in the field may be a problem

- Survey of faculty, postdocs, graduate students
- 30 disciplines (9 SS, 9 Hum and 12 STEM)
- 9 research universities (5 public, 4 private)
- 1820 usable responses
- Representation of women and African-Americans in the field

Beliefs about our fields are a problem

*Field-specific Ability Beliefs* (4 “I believe”; 4 “others in my field believe”)

- Being a top scholar of [discipline] requires a special aptitude that just can’t be taught.
- If you want to succeed in [discipline], hard work alone just won’t cut it; you need to have an innate gift or talent.
- With the right amount of effort and dedication, anyone can become a top scholar in [discipline]. (R)
- When it comes to [discipline], the most important factors for success are motivation and sustained effort; raw ability is secondary. (R)
How fields develop gender gaps

- **Field-specific Ability Beliefs**: Belief that success in a field requires raw aptitude or talent.
- **Cultural Stereotyping**: Belief that men are more likely than women to possess raw intellectual ability.
- **Academic Gender Gaps**: Women are underrepresented in fields that emphasize the need for innate talent rather than effort.
Percent US PhDs awarded to women in 2011 in STEM fields
Percent US PhDs awarded to women in 2011 in social sciences and humanities

Fig. S2. Percentage of US PhDs awarded to women in 2011 (NSF Survey of Earned Doctorates) in STEM (A) and Social Science and Humanities (B) disciplines.
Field-specific ability beliefs and female percentage of 2011 US PhDs
Alternative explanations ruled out

- STEM vs. SS/Hum
- Total hours worked
- Selectivity (% applicants admitted and department ranking)
- Strength of applicant pools (GRE composite)
- Systemizing vs. Empathizing discipline ratings
  - Identifying the abstract principles, structures, or rules that underlie the relevant subject matter (Systemizing)
  - Analyzing the relevant subject matter and constructing a systematic understanding of it (Systemizing)
  - Having a refined understanding of human thoughts and feelings (Empathizing)
  - Recognizing and responding appropriately to people’s mental states (Empathizing)

Anchoring heuristic affects judgments over time

- Start from a previous judgment in making a new judgment
  - Make insufficient adjustment, given new information

Is it impossible for academics to judge merit?
Our judgments can be improved

- Some situational factors help
  - Comparative pool (haloes very likely when no comparisons)
  - Enough information
  - Avoidance of “evidence” that is likely to include bias

- Some situational constraints help
  - Establishment a priori of detailed criteria that can be observed
  - Reliance on specific evidence in judging each criterion
  - Avoidance of halos, intuition, and global judgments
  - Disciplined evaluation of criteria across individuals
  - If a group process
    - Diversity in group membership
    - Enough time
    - Explicit process for correcting inaccuracies

Some internal factors help

- Recognition of uncertainty
- Commitment to consider role of situation in producing outcomes
  - Access to resources (human, time, and material)
  - Accumulation of advantage and disadvantage
- Examination of own process of developing opinions
  - Acknowledgment of irrelevant reactions/interpretations
  - Explicit consideration of alternative interpretations
  - Search for evidence to help decide
- Lower confidence in judgments

Observations on the Education of the Judgment (1867)—Michael Faraday

- “...it is because our procedure was hasty, our data too few, and our judgment untaught, that we fell into mistake; not because the data were wrong.” (p. 193)

- “This education has as its first and its last step *humility.*” (p. 215)