EXPERIMENTAL ECONOMICS

GENDER

Ernesto Reuben
Large differences in the presence and compensation of women in many (top) jobs

- There is a persistent gender gap in wages accompanied with important gender differences in labor-market trajectories (Blau & Kahn 2013, Goldin 2014, Goldin et al. 2017)
- Female MBAs from U of Chicago have 30% lower salaries 5 years after graduation and 60% lower salaries 10 years out (Bertrand et al., 2010)
DECOMPOSING THE GENDER GAP

Blau & Kahn (2017)

1980

38%

1% Other differences

9% Differences in education and experience

18% Differences within the same occupation

10% Differences across occupations

2010

21%

10%

8% 2%

1%
Why the Persistent Gender Gaps?

Supply side

- Differences in preferences between career and family (Bertrand et al. 2010, Goldin 2014)
- Differences in risk aversion
- Differences in competitiveness
- Differences in bargaining

Demand side

- Taste-based discrimination
- Belief-based or statistical discrimination
  - Correctly inferring differences in performance
  - Biased beliefs about women’s relative performance
Croson & Gneezy (2009)

- Considerable evidence that women are significantly more averse to taking risks
- Example with 550 MBA students and a multiple price list

Risk averse
- Men: 58%
- Women: 80%

Risk neutral
- Men: 37%
- Women: 16%

Risk loving
- Men: 5%
- Women: 4%
Considerable evidence that women are significantly more averse to taking risks

Example with 550 MBA students and a survey question

Are you generally fully prepared to take risks?

- Yes
- No

Females
Males
Risk Aversion

Croson & Gneezy (2009)

- Considerable evidence that women are significantly more averse to taking risks

Why?

- **Emotions**: Women report more nervousness and fear in anticipation of negative outcomes
- **Overconfidence**: Men being more overconfident in their success in uncertain situations
- **Perception of risk**: Men are more likely to see a risky situation as a challenge while females women risky situations as threats
- **Domain**: Most experiments use monetary lotteries to elicit risk preferences
- And remember, what does small-stake risk aversion really mean?
Maestripieri et al. (2009)

- Elicit risk aversion with monetary lotteries, testosterone (salivary and 2D:4D digit ratio), and empathy with the Baron-Cohen eye test from 320 men and 140 women.

<table>
<thead>
<tr>
<th>Association of risk aversion with ...</th>
<th>Both</th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salivary testosterone</td>
<td>−0.082***</td>
<td>−0.020</td>
<td>−0.137**</td>
</tr>
<tr>
<td></td>
<td>(0.022)</td>
<td>(0.029)</td>
<td>(0.057)</td>
</tr>
<tr>
<td>2D:4D digit ratio</td>
<td>41.77</td>
<td>−37.16</td>
<td>79.03</td>
</tr>
<tr>
<td></td>
<td>(32.22)</td>
<td>(40.96)</td>
<td>(50.19)</td>
</tr>
<tr>
<td>Baron-Cohen eye test</td>
<td>0.595***</td>
<td>0.219</td>
<td>1.319***</td>
</tr>
<tr>
<td></td>
<td>(0.222)</td>
<td>(0.254)</td>
<td>(0.422)</td>
</tr>
</tbody>
</table>
COMPETITIVENESS

[adjective] a desire and/or ability to perform in competitions
Selection into tournaments

- 550 subjects
- Task: add sums of four two-digit numbers (e.g., 11 + 42 + 86 + 70) for four minutes
- Choice of payment scheme
  - Piece-rate: $4 per correct answer irrespective of the performance of others
  - Tournament: if you answer the most sums in a group of four then $16 per correct answer, otherwise $0
Why do women compete less?
- Differences in performance
  - Not the case in arithmetic
- Differences in beliefs
  - Men are overconfident
- Preferences for risk
  - Women are more risk averse
- Differences in altruism
  - Women do not want to hurt others
- Aversion to competition
  - Women dislike performing in competitions

Identifying competitiveness with an experiment (Niederle & Vesterlund 2007)
1. Everyone plays under piece-rate
2. Everyone plays under tournament
3. Choose between piece-rate and tournament but compete against performance in 2
4. Choose between piece-rate and tournament but do not play again, just submit performance in 1
5. Elicit expected rank in 2
Niederle & Vesterlund (2007)

No difference in performance

Large difference in tournament entry
Evidence of overconfidence, particularly by men

<table>
<thead>
<tr>
<th></th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean expected rank</td>
<td>1.40</td>
<td>1.83</td>
</tr>
<tr>
<td>% expecting to be 1st</td>
<td>75%</td>
<td>43%</td>
</tr>
<tr>
<td>% ranked 1st conditional on expecting to be 1st</td>
<td>27%</td>
<td>47%</td>
</tr>
</tbody>
</table>

For the same expected rank in task 3, women do not enter as much as men!
**COMPETITIVENESS AND TOURNAMENT ENTRY**

Niederle & Vesterlund (2007)

**Smaller difference in task 4 that disappears with controls**

For the same expected rank in task 4, no difference in tournament entry!
COMPETITIVENESS AND TOURNAMENT ENTRY

Niederle & Vesterlund (2011)

- Women avoid performing in competitive environments → replicated many times

Why a gap in tournament entry?

- Beliefs
  - Gap weakens in tasks where women are expected to perform better (e.g. verbal tasks) and when feedback about relative performance is given

- “Culture”
  - Gap weakens when competition is for teams and not individuals, in matrilineal societies, among young children, and for girls who attend same-sex schools

- Measurement error
Gender differences in competitiveness are commonly identified through the residual in regression analysis

If one has not controlled for the right variables or the measures of expectations and risk aversion are very noisy, the effect of competitiveness is easily overestimated (Gillen et al. 2017, van Veldhuizen 2017)

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Man</td>
<td>0.27***</td>
<td>0.19***</td>
<td>0.17***</td>
<td>0.13**</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Prob. of rank 1</td>
<td>0.06**</td>
<td>0.06**</td>
<td>0.06*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td></td>
</tr>
<tr>
<td>Expected rank</td>
<td>-0.13***</td>
<td>-0.13***</td>
<td>-0.11***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td></td>
</tr>
<tr>
<td>Risk aversion</td>
<td>-0.07***</td>
<td>-0.06**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other controls</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>R²</td>
<td>0.06</td>
<td>0.18</td>
<td>0.20</td>
<td>0.26</td>
</tr>
</tbody>
</table>
Competitiveness and education

- Competitive Dutch students are more likely to pick the most prestigious high school track (Buser et al. 2014) and competitive US undergraduates have higher earnings expectations (Reuben et al. 2017)

Competitiveness and entrepreneurship

- Competitive Tanzanian entrepreneurs invest more in their firms, hire and fire more, and use more performance-based compensation, but do not earn more (Berge et al. 2015)
Competitiveness and business (Reuben et al. 2016)

- Competitive MBA graduates earn more at graduation, and are more likely to start and stay in consulting or finance.
COMPETITIVENESS AND PERFORMANCE

Who would you bet on to win a race? (Gneezy & Rustichini 2004)

- 140 children aged 9 to 10 racing alone
Who would you bet on to win a race? (Gneezy & Rustichini 2004)

- 140 children aged 9 to 10 racing again (time difference compared to first race)

Similar qualitative findings with students and monetary incentives to solve mazes (Gneezy et al. 2003)
Do women avoid bargaining? (Small et al. 2007)

- 81 men and 72 women complete a word task for which they had been told they will be compensated between $3 and $10. When completed, the experimenter gives them $3 and says “Here’s $3. Is $3 OK?”

- **No negotiation cues vs. negotiation cue** (“payment is not fixed, you can negotiate for more”) vs. **asking cue** (“payment is not fixed, you can ask for more”)

![Graph showing bargaining rates by gender and condition]
Do women avoid bargaining? (Leibbrandt & List 2015)

- Posted 18 adds in major US cities for jobs paying around $18/hour for which 2382 job applicants (67% female) signaled interest
- **Control** (the position pays $18 per hour) vs. **negotiation** ("but the applicant can negotiate a higher wage")
Do women avoid bargaining? (Exley et al. 2018)

- 72 men and 74 women perform a real-effort task to determine their output as workers. Firms offer a random wage equal to their worker’s output \(-$4, -$2, +$0, \text{ or } +$2\).
- Workers negotiate the final wage either *always* or when workers *choose* to do so.
  - Negotiation occurs with a free-form chat with the firm and if there is no agreement both lose $5.
Do women avoid bargaining? (Exley et al. 2018)

- Should women negotiate more? → No, they would incur loses
Taste-based discrimination

- Not hiring women because of:
  - Misogyny
  - Threat to identity
  - Consumer expectations
  - etc.

Statistical discrimination

- Correctly inferring differences in performance
- Biased beliefs about women’s relative performance
- **Blind auditions (Goldin & Rouse 2000)**

- How was the gender composition of orchestras affected by the introduction of blind auditions?

- Exploiting differences in the adoption rate of this hiring practice, it is estimated that **30% of the increase of women is due solely to blind auditions**.

![Graph showing the fraction of candidates hired in blind and non-blind auditions](image.png)
**Labor-Market Discrimination**

**Moss-Racusin et al (2012)**

- Ask faculty to give feedback about the application materials of a student who will apply to lab manager positions

- Competence
- Likelihood of hiring
- Starting salary
- Likeability
CONSEQUENCES OF STEREOTYPES

Moss-Racusin et al (2012)

<table>
<thead>
<tr>
<th>John</th>
<th>Competence</th>
<th>Jennifer</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

26%

<table>
<thead>
<tr>
<th>John</th>
<th>Likelihood of hiring</th>
<th>Jennifer</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

27%
Moss-Racusin et al (2012)

**Evidence of discrimination**

- **But why?**
  - Women rated as more likable → unlikely to be taste
  - Responses to the modern sexism scale negatively correlate with competence, hireability, and mentoring of female students ($p < 0.01$) and positively correlate for male students ($p > 0.09$)
Gender earnings gap among lawyers (Goldin 2014)

- After 5 years: 6% (increase)
- After 15 years: 9% (increase)

Gender earnings gap among MBAs (Bertrand et al. 2010)

- After 3 years: 6% (increase)
- After 10 years: 13% (increase)

Legend:
- Red: Same education
- Blue: Same number of hours worked
- Green: No career interruptions
**DISCRIMINATION IN THE LAB**

**Why study discrimination in the lab?**
- Accurately measure performance
  - Measure whether there are differences in performance between men and women
- Separate belief-based from taste-based discrimination
- Observe how beliefs are updated with new information
- (Partly) identify the source of bias in beliefs (implicit and/or explicit)

**Reuben et al. (2014)**
1. Everyone adds sums and are paid per correctly-answered sum
2. Candidate picking task
3. Everyone adds sums again and are paid per correctly-answered sum

\[
\begin{align*}
47 + 98 + 35 + 22 &= 202 \\
15 + 57 + 46 + 45 &= 163 \\
25 + 16 + 46 + 92 &= 179 \\
35 + 23 + 83 + 36 &= 177 \\
58 + 86 + 82 + 55 &= 281 \\
47 + 98 + 35 + 22 &= 202
\end{align*}
\]
Discrimination in the Lab

Reuben et al. (2014)

- Candidate picking task
- Guess number of sums
  - Paid for accuracy (between $0 and $9)
- Pick one of the candidates
  - Paid according to candidate’s performance in part 3
  - Picking a candidate increases his/her earnings by $4
Reuben et al. (2014)

- Is there a gender gap in performance?

No statistical difference in performance!

Men: 11.52 sums

Women: 11.76 sums

- Is there a gender gap in performance evaluation?

Men are expected to perform better

Men: 13.04 sums

Women: 11.41 sums
Reuben et al. (2014)

- Is there a gender gap in picking candidates?

If one hires a man → 48% chance of hiring the low performer

If one hires a woman → 41% chance of hiring the low performer
Implicit Association Test (Greenwald et al. 1998)

- After part 3, between Male or Female pictures and liberal arts or science/math words
Reuben et al. (2014)

- IAT score correlates with the expected performance of candidates
Reuben et al. (2014)

- Candidate picking task + **Cheap Talk**
  - Guess number of sums and pick a candidate
  - Candidates state their expected future performance
  - Guess number of sums and pick a candidate again

- Candidate picking task + **Past Information**
  - Guess number of sums and pick a candidate
  - Candidates performance in part 1 is revealed
  - Guess number of sums and pick a candidate again

I will do 25 sums!

Part 1 performance: 12 sums

predicts better candidate 95% of the time
Reuben et al. (2014)

- Cheap talk improves performance but not the gender balance
- Past performance improves performance and the gender balance

Fraction hiring a candidate who is a ...

<table>
<thead>
<tr>
<th></th>
<th>Female</th>
<th>Low performer</th>
</tr>
</thead>
<tbody>
<tr>
<td>No information</td>
<td>40%</td>
<td>50%</td>
</tr>
<tr>
<td>Cheap talk</td>
<td>56%</td>
<td>62%</td>
</tr>
<tr>
<td>Past information</td>
<td>60%</td>
<td>67%</td>
</tr>
</tbody>
</table>

Chance of the low performer if one hires a ...

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>No information</td>
<td>40%</td>
<td>50%</td>
</tr>
<tr>
<td>Cheap talk</td>
<td>56%</td>
<td>62%</td>
</tr>
<tr>
<td>Past information</td>
<td>60%</td>
<td>67%</td>
</tr>
</tbody>
</table>
Reuben et al. (2014)

- Do implicit stereotypes affect explicit belief updating?
- Measuring the degree of belief updating

He said 16, so will probably do 12 sums

Reuben et al. (2014) investigated whether implicit stereotypes influence explicit belief updating. They defined belief updating as the process of adjusting one's beliefs in light of new information. The formula for measuring the degree of belief updating is given by:

\[ \varphi = \frac{12 - 8}{16 - 8} \]

He calculated the value of \( \varphi \) to be 0.5, indicating a moderate belief update.

The image illustrates the decision-making process: if the signal is 1, the new information is completely believed; if the signal is 0, the new information is completely discarded. The decision is determined by the difference between the posterior and the prior belief, normalized by the signal strength.

In this scenario, it is decided to do 16 sums, which is slightly more than the originally speculated amount of 12 sums.
Reuben et al. (2014)

- Do implicit stereotypes affect explicit belief updating?

**Past Performance**

<table>
<thead>
<tr>
<th></th>
<th>Observed $\varphi$</th>
<th>Optimal $\varphi$</th>
<th>Observed $\varphi$ for low IAT</th>
<th>Observed $\varphi$ for high IAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female candidate</td>
<td>0.70</td>
<td>0.90</td>
<td>0.71</td>
<td>0.67</td>
</tr>
<tr>
<td>Male candidate</td>
<td>0.73</td>
<td>0.96</td>
<td>0.74</td>
<td>0.73</td>
</tr>
</tbody>
</table>

Not enough updating (conservatism) $\rightarrow$ too much weight on a bad stereotype

Implicit stereotypes have no effect on explicit updating
Reuben et al. (2014)

- Do implicit stereotypes affect explicit belief updating?

**Cheap talk**

<table>
<thead>
<tr>
<th></th>
<th>Observed $\phi$</th>
<th>Optimal $\phi$</th>
<th>Observed $\phi$ for low IAT</th>
<th>Observed $\phi$ for high IAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female candidate</td>
<td>0.62</td>
<td>1.09</td>
<td>0.62</td>
<td>0.61</td>
</tr>
<tr>
<td>Male candidate</td>
<td>0.48</td>
<td>0.88</td>
<td>0.38</td>
<td>0.56</td>
</tr>
</tbody>
</table>

Even less updating (more conservatism)

Women are believed more (for good reasons)

Implicit stereotypes impact explicit updating

High IAT $\Rightarrow$ do not account for male overconfidence
Performance evaluation **is** susceptible to bias due to implicit and explicit stereotypes?

- Stereotypes can be partly overcome with information, but only when it is considered objective and accurate.
- Too much weight is given to very uninformative stereotypes.
- Inaccurate and/or subjective information can be both useful and susceptible to implicit biases.


• Gillen, Ben, Erik Snowberg, and Leeat Yariv. 2017. Experimenting with Measurement Error: Techniques with Applications to the Caltech Cohort Study.

References

REFERENCES

- van Veldhuizen, Roel. 2017. *Gender Differences in Tournament Choices: Risk Preferences, Overconfidence or Competitiveness?*