Evidence on Self-Stereotyping and the Contribution of Ideas

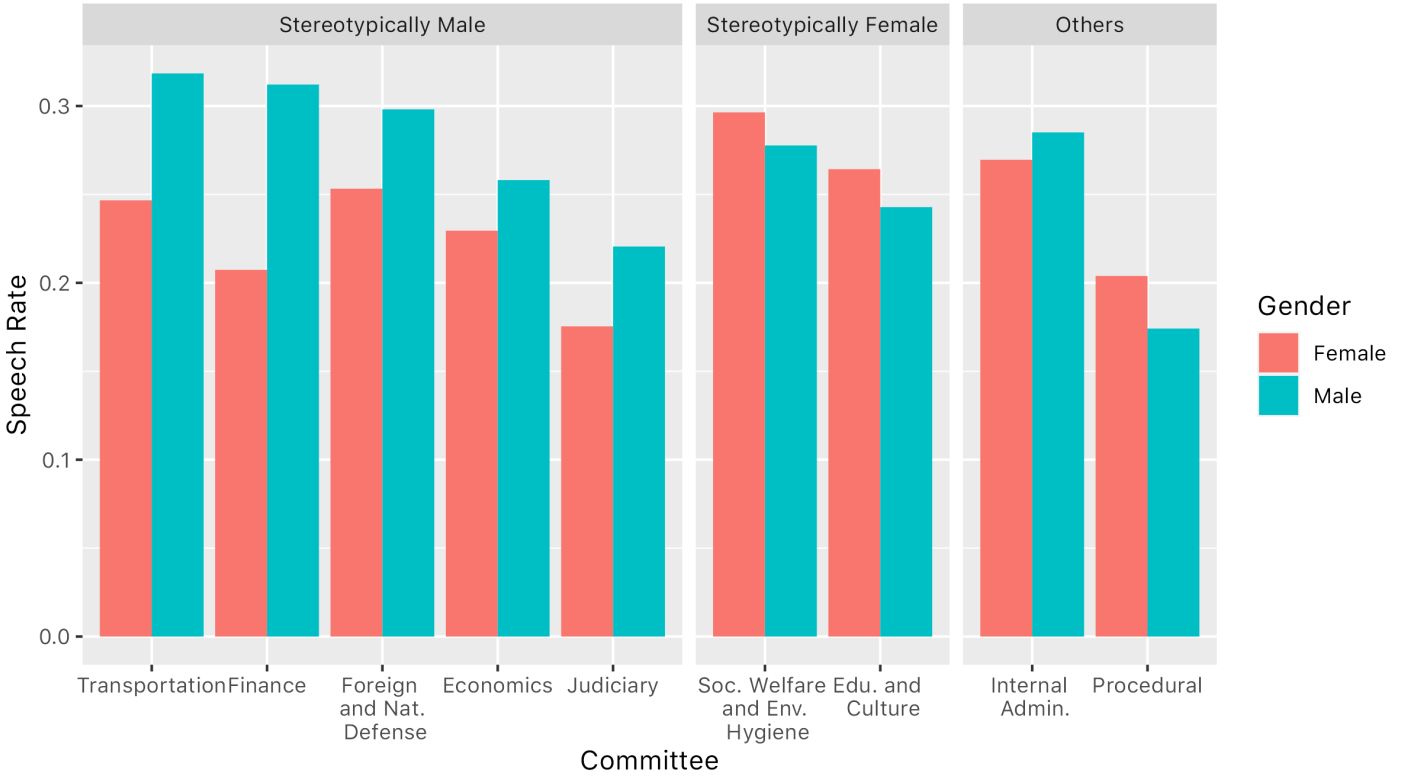
Katherine B. Coffman (2014, QJE)

Yu-Hsin Ho

Jul 18, 2023

Introduction

- Women are now jointly making important decisions in public domains
  - Politically: more females are becoming MPs
  - Economically: female share of board members is increasing
- However, participation does not mean having real influential power
- Some people talk more than others, and gender plays a significant role here
  - Imagine: seminars, meetings, classrooms, and committees.



Proportion of speaking by MP's gender in legislative committees in Taiwan

Problems

Willingness to contribute ideas varies between domains:

- Valuable knowledge may not be elicited in group discussion
- Valuation of individual ability is biased
  - Affect labor market outcomes: wage, promotion, leadership
- Systematically differed behaviors generate gender gaps in equilibrium outcome
  - willingness to compete
  - willingness to lead
  - willingness to contribute ideas

Goals of This Paper

Coffman (2014) conducted experiments to:

1. Identify the role of gender self-stereotyping on willingness to contribute ideas
2. Discuss the mechanism
3. Impute the efficiency loss in group discussion due to gender stereotypes

Experimental Design

Part B: Measuring objective ability

- Each participant answers 5 multiple choice questions (MCQs) from each of 6 domains of knowledge
- Variables collected as a measure of individual ability

Part C: Measuring willingness to contribute ideas

- Each participant is assigned to a group of 2
- Answers MCQs and then choose a relative position to stand in the group → key outcome variable

Part D: Measuring confidence and risk preference

Part B: Measuring objective ability

- Individuals answer 5 MCQs from each of 6 different domains of knowledge (30 in total)
  - Incentivized: Get ECUs if an answer is correct
- Domains of knowledge: purposefully generate variations in gender stereotypes
  1. Arts and literature (Art)
  2. Pop culture (Pop)
  3. Environmental science (Env)
  4. History (Hist)
  5. Geography (Geo)
  6. Sports and games (Sports)
- Target: objective ability of subjects in different domains of knowledge

Sample questions:

> SPORTS&GAMES: How many Major League Baseball teams today are named for four-footed animals?
>
> (A) one (B) two (C) three (D) four (E) five
>
> (B) Two: Tigers, Cubs.

> ART&LIT: What war do the girls in Little Women grow up during?
>
> (A) American Revolution (B) World War I (C) World War II (D) Civil War (E) Vietnam War
>
> (D) Civil War

Part C: Measuring willingness to contribute ideas

Each participant is assigned to a group of 2; plays 30 rounds of group tasks:

1. Two members both answer the MCQ
2. Decides which position they want in the group: 1, 2, 3, or 4. → willingness to contribute ideas
3. Group member with the smallest chosen number is selected as the group representative

> E.g. Eva picks 2, and Dexter picks 3. → Eva's answer becomes the group representative

Part C: Measuring willingness to contribute ideas

Each participant is assigned to a group of 2; plays 30 rounds of group tasks:

1. Two members both answer the MCQ
2. Decides which position they want in the group: 1, 2, 3, or 4. → willingness to contribute ideas
3. Group member with the smallest chosen number is selected as the group representative

Payoff scheme

1. Get ECUs if one answers correctly
2. Get extra ECUs if one group representative's answer is correct

Key: Subjects are incentivized to elicit answers from the most knowledgeable person

Part D: Measuring confidence, risk preference, and stereotypes

Self-confidence & confidence in the teammate

Incentivized elicitation mechanism:

- For question q, participant i states her minimum acceptable correct probability p_{iq}
- p_{iq}^c is randomly drawn from $[0.01, 1]$
  - if $p_{iq}^c \geq p_{iq}$, then the computer would help the participant select one answer with correct probability p_{iq}^c
  - if $p_{iq}^c < p_{iq}$, then the participant's answer is used
- Get ECUs if the finally chosen answer is correct

Part D: Measuring confidence, risk preference, and stereotypes

Risk preference

Measured by multiple price lists. Used as a control variable.

Stereotypes

- For each domain of question, subjects are asked to pick a number on a continuous slider scaled -1 to 1:
  - -1: women know more
  - 0: no gender difference
  - 1: men know more

Standardized z-scores are constructed as "maleness" scores for each domain of question

Treatment Intervention

Between subjects, 2×2 treatment intervention:

1. Feedback: Showing domains that one's Part B score dominates her teammate in before the beginning of Part C
2. Public: Taking a picture before Part C and having the picture revealed to the teammate in Part F
    - Still, subjects do not know the teammate's gender in Part C

Sessions information

- 460 participants
- 38 sessions at Ohio State University

Results

Summary statistics

TABLE I

SUMMARY STATISTICS

	Men	Women	Total	p value [H_0: M = W]
White	75.1%	57.5%	68.3%	<0.01
East Asian	12.8%	22.3%	16.5%	<0.01
Black or African American	4.0%	7.8%	5.4%	0.07
Asian Indian	3.2%	4.5%	3.7%	0.48
Attended high School in United States	91.5%	83.2%	88.3%	<0.01
Current Ohio State undergraduate	90.4%	95.0%	92.2%	0.08
Avg. # of correct Part B answers	13.6	12.6	13.2	<0.01
Arts and literature (Art)	2.26	2.45	2.34	0.10
Pop culture (Pop)	1.02	1.15	1.07	0.15
Environmental science (Env)	4.05	3.78	3.95	<0.01
History (Hist)	1.64	1.17	1.46	<0.01
Geography (Geo)	1.93	1.88	1.91	0.64
Sports and games (Sports)	2.70	2.13	2.49	<0.01
Avg. # of correct Part C answers	15.8	14.3	15.2	<0.01
Arts and literature	2.06	2.25	2.13	0.05
Pop culture	2.97	2.87	2.93	0.34
Environmental science	2.96	2.52	2.79	<0.01
History	2.25	2.07	2.18	0.10
Geography	2.44	2.24	2.36	0.05
Sports and games	3.14	2.31	2.82	<0.01
Totals	281	179	460	

Notes. p-values are from tests of proportions for binary variables and Fisher-Pitman permutation tests for nonbinary variables, with a null of equality of distributions between men and women.

The maleness of knowledge

Average maleness score (stereotypes)

TABLE II

PERCEIVED GENDEREDNESS OF CATEGORIES

	Avg. maleness given by men	Avg. maleness given by women	Overall average	Normalized z-score
Art	−0.317	−0.419	−0.356	−1.18
Pop	−0.263	−0.348	−0.297	−1.01
Env	0.142	0.057	0.109	0.13
Hist	0.196	0.061	0.144	0.23
Geo	0.215	0.065	0.157	0.27
Sports	0.643	0.571	0.615	1.56

Note. Elicited on [−1,1] scale where −1 is labeled "Women know more," 1 is labeled "Men know more," 0 is "no gender difference."

Gender gap in willingness to contribute ideas

Average positions in line

TABLE III

AVERAGE PLACE IN LINE BY CATEGORY AND GENDER

	Average place in line						
	Art	Pop	Env	Hist	Geo	Sports	Pooled
Men	2.60	2.24	2.11	2.22	1.96	2.11	2.21
	[0.62]	[0.58]	[0.71]	[0.71]	[0.65]	[0.68]	[0.51]
Women	2.36	2.20	2.39	2.48	2.31	2.69	2.40
	[0.66]	[0.60]	[0.67]	[0.71]	[0.71]	[0.67]	[0.51]
p-value	<0.01	0.42	<0.01	<0.01	<0.01	<0.01	<0.01
Observations	460	460	460	460	460	460	460

Notes. Lower positions in line indicate greater willingness to contribute. Unit of observation is a mean of an individual's five places in line within each category or across all categories in the pooled column. p-values derived from Fisher-Pitman permutation test for two independent samples, testing the null of equality of the two distributions using Monte Carlo method with 200,000 simulations. Standard deviations of mean place in line shown in brackets.

Gender gap persists after controlling individual ability

Outcome: position in line

TABLE IV

WILLINGNESS TO CONTRIBUTE

OLS predicting position in line for question i in Part C							
Category	Art	Pop	Env	Hist	Geo	Sports	Pooled
Maleness z-score	−1.18	−1.01	0.13	0.23	0.27	1.56	
Female dummy	−0.226****	−0.090*	0.217****	0.145**	0.296****	0.383****	0.145***
	(0.065)	(0.048)	(0.067)	(0.071)	(0.065)	(0.058)	(0.048)
Maleness of category							−0.120****
							(0.013)
Female × maleness							0.265****
							(0.019)
Answered qn. i correctly	−0.480****	−0.977****	−0.683****	−0.415****	−0.553****	−1.003****	−0.706****
	(0.053)	(0.058)	(0.040)	(0.046)	(0.048)	(0.046)	(0.023)
Part B score in category	−0.081****	−0.016	−0.103***	−0.119****	−0.081***	−0.067**	−0.024***
	(0.025)	(0.025)	(0.034)	(0.026)	(0.031)	(0.031)	(0.009)
Constant	3.53****	4.27****	2.26****	3.42****	2.80****	3.78****	3.33****
	(0.226)	(0.175)	(0.294)	(0.273)	(0.252)	(0.255)	(0.193)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clusters	460	460	460	460	460	460	460
Obs.	2,299	2,300	2,300	2,298	2,298	2,298	13,793
R^2	0.268	0.548	0.160	0.170	0.166	0.322	0.241

Mechanism

- People shy away from contributing ideas when a topic is **gender-incongruent**
  - Persists after conditional on ability
  - The degree of shying away: female > male
- One explanation: People **lack confidence** in gender-incongruent topics
  - Gender stereotypes play a significant role in shaping beliefs
  - Empirical question: Can the gender-incongruency predict self-reported confidence?

Gender-incongruency effect on self-confidence

Probit outcome: guessing myself has the highest Part B score in the group (conducted at the end of Part B)

TABLE V

PREDICTING PARTICIPANT BELIEFS ABOUT PART B PERFORMANCE

Probit predicting Pr(guessed she had highest Part B score in group)							
Category	Art	Pop	Env	Hist	Geo	Sports	Pooled
Maleness z-score	−1.18	−1.01	0.13	0.23	0.27	1.56	
Female dummy	0.211****	0.207****	−0.021	−0.132***	−0.254****	−0.416****	−0.088****
	(0.052)	(0.051)	(0.051)	(0.051)	(0.050)	(0.045)	(0.023)
Maleness of category							0.121****
							(0.015)
Female × maleness							−0.249****
							(0.021)
Part B score in category	0.123****	0.075***	0.100****	0.125****	0.089****	0.125****	0.092****
	(0.023)	(0.026)	(0.026)	(0.022)	(0.025)	(0.028)	(0.008)
Constant	0.403****	0.389****	0.614****	0.390****	0.511****	0.438****	0.457****
	(0.021)	(0.022)	(0.022)	(0.021)	(0.022)	(0.020)	(0.009)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clusters	460	460	460	460	460	460	460
Obs.	460	460	460	460	460	460	2,760
Pseudo R^2	0.114	0.071	0.054	0.121	0.067	0.231	0.113

People display significantly lower confidence in stereotypically mismatch areas.

Controlling self- and other-confidence...

TABLE VI

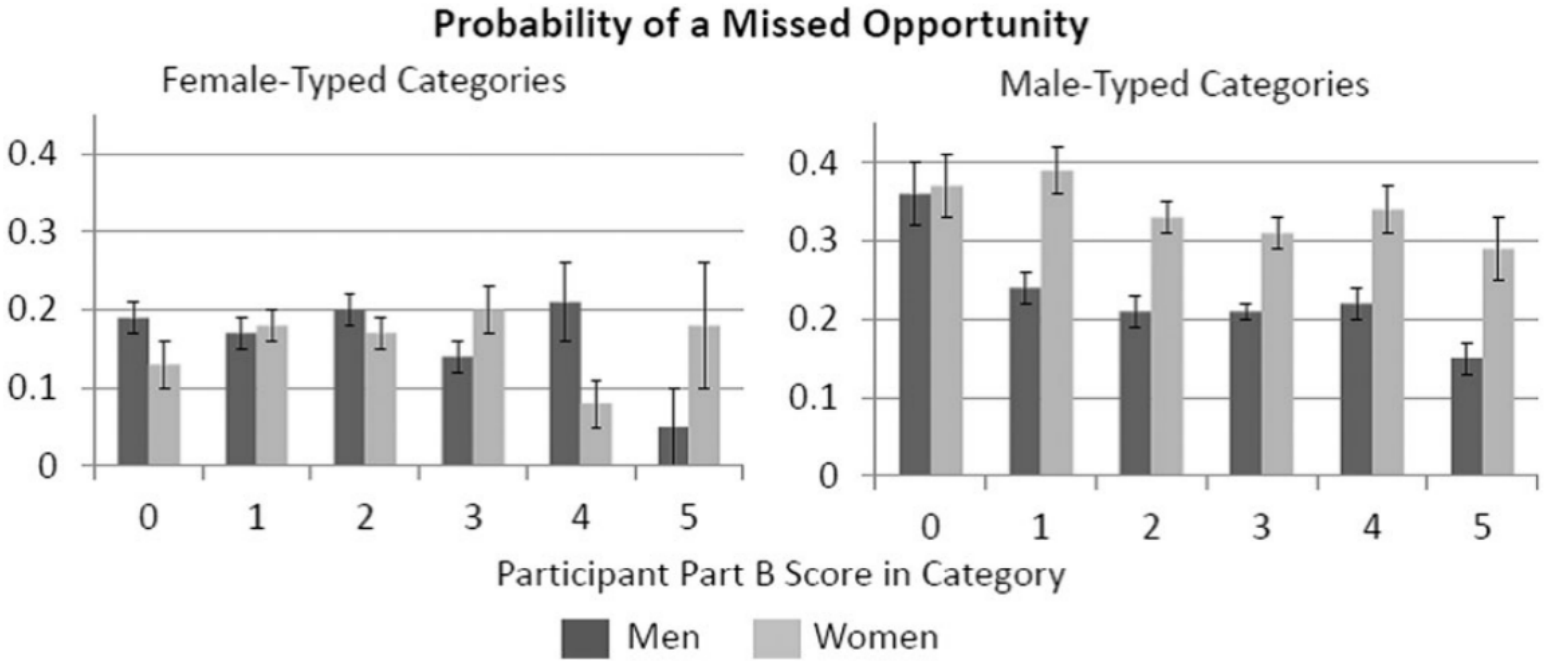
THE ROLE OF CONFIDENCE IN PREDICTING WILLINGNESS TO CONTRIBUTE

OLS predicting position in line for question *i* in Part C							
Category	Art	Pop	Env	Hist	Geo	Sports	Pooled
Maleness *z*-score	−1.18	−1.01	0.13	0.23	0.27	1.56	
Female dummy	−0.011	−0.023	0.144***	0.104*	0.090*	0.099**	0.079*
	(0.054)	(0.040)	(0.052)	(0.056)	(0.048)	(0.051)	(0.040)
Maleness of category							−0.036****
							(0.010)
Female × maleness							0.071****
							(0.015)
Qn. *I* correct	−0.171****	−0.232****	−0.243****	−0.147****	−0.166****	−0.244****	−0.193****
	(0.041)	(0.045)	(0.035)	(0.037)	(0.036)	(0.040)	(0.016)
Part B score	−0.011	0.023	−0.023	−0.029	−0.007	0.016	0.010
	(0.022)	(0.020)	(0.027)	(0.022)	(0.023)	(0.025)	(0.007)
Pr(qn. *I* correct)	−0.026****	−0.026****	−0.025****	−0.025****	−0.026****	−0.027****	−0.027****
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Pr(partner qn. *i* correct)	0.003*	0.002**	0.005****	0.003***	0.002**	0.005****	0.003****
	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Ranked self first	−0.106**	−0.134****	−0.061	−0.130**	−0.147****	−0.198****	−0.126****
	(0.052)	(0.038)	(0.052)	(0.054)	(0.044)	(0.050)	(0.021)
Constant	4.24****	4.35****	3.68****	4.18****	4.12****	3.93****	4.14****
	(0.192)	(0.166)	(0.225)	(0.219)	(0.170)	(0.202)	(0.157)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clusters	460	460	460	460	460	460	460
Obs.	2,299	2,300	2,300	2,298	2,296	2,298	13,791
R^2	0.578	0.744	0.495	0.476	0.558	0.632	0.593

Gender-incongruent effects still exist after conditional on ability and confidence

Welfare: Missed opportunity

- **Missed opportunity**: One answers the question correctly but not being chosen as the group's answer.
- Stereotypes may cause efficiency loss if the actually chosen representative answer is wrong.



Treatment effects

TABLE VII

THE IMPACT OF TREATMENTS ON POSITION IN LINE

OLS predicting position in line for question *i* in Part C for group member with best Part B score in category

	Female-typed categories	Male-typed dategories	Pooled
Female dummy	−0.112	0.361****	0.224***
	(0.091)	(0.092)	(0.082)
Maleness of category			−0.094****
			(0.024)
Female × maleness			0.248****
			(0.043)
Feedback treatment	−0.079	−0.124*	−0.089
	(0.068)	(0.065)	(0.058)
Public treatment	0.068	0.057	0.073
	(0.067)	(0.062)	(0.057)
Female × feedback	0.019	−0.020	−0.027
	(0.113)	(0.106)	(0.096)
Female × feedback × maleness			−0.038
			(0.049)
Female × public	−0.132	−0.095	−0.120
	(0.110)	(0.104)	(0.095)
Female × public × maleness			0.081
			(0.050)
Constant	3.574****	2.807****	3.157****
	(0.217)	(0.247)	(0.214)
Controls	Yes	Yes	Yes
Clusters	391	452	459
Obs.	2,914	5,845	8,759
R^2	0.395	0.168	0.242

No effect of feedback on willingness to contribute

- Telling people *"you're good"* does not help

Wrap up

1. People shy away from contributing ideas when a topic is **gender-incongruent**
    - Gender stereotypes of tasks are important. Not identified via gender-neutral real-effort games.
2. This could be mostly (but not fully) explained by low confidence
3. Efficiency costs incurred by stereotypes:
    1. Average individual ability: 51% correct rate
    2. Group answer correct rate: 58% correct rate
    3. At least one member is correct: 71% correct rate
4. Providing positive information on self-ability does not help

Further extensions

- We already knew quotas help mitigate the gender gap in willingness to compete and overcome some norms in mixed-gender environments.
- Can we fix this problem through policy intervention without efficiency loss?

Proposal

- Preferential treatment in Part C for women:
  - Having women's answers moved forward by one position
- Coffman (2014) manipulated data to do a simple counterfactual:
  - Moving all women forward by one position reduces missed opportunity: 23.8% → 21.2%
  - But, this does not capture the **subject's response** & **general equilibrium** effect.
- Under a well-designed lab experiment, we aim to evaluate the effectiveness of preferential treatment on willingness to contribute ideas

Research questions

1. **Baseline**: Does the gender-incongruence effect on willingness to contribute ideas exist in Taiwan?
2. **Behaviors**: How would subjects respond to preferential treatment (PT)?
   - Moderate-confidence women may take the PT, while low-confidence women may actively try to offset PT.
   - This self-selection mechanism may generate Pareto-improvement in male-typed domains: All group members benefit from high-quality answers.
   - Would male participants move themselves forward?
3. **Attitudes**:
   - Backfire? How would subjects treat each other in the dictator game?
   - What's the self-efficacy of preferentially treated women?

Experimental design

- 2 people in a group: one man and one woman
  - Teammate's gender is explicitly revealed to participants through roll-calling
- Treatment in Part C: Move females forward by one position
- Dictator game following Part C: Measuring treatment effect on attitudes toward teammate
- Survey questions on self- and other-efficacy
  - *"From 1 to 10, how would you grade your/your teammate's performance?"*
  - *"To what extent do you consider your team's outcome could be attributed to your performance?"*