Beliefs about Gender and Task Performance: Replicating Bordalo et al. (2019) in an Indian Context

ECO-3610: Experimental Economics

Akriti Verma, Elizabeth Jose, Madhumitha GI,
Neha Maniar, Shashvathi Hariharan

Research Question

Context about Paper:
Bordalo et al. (2019) show how stereotypes shape beliefs through:

- **Stereotype exaggeration**: self-beliefs exceed actual ability
- **Difficulty-based distortion**: larger overestimation for hard questions

Research Question:

- How do gender stereotypes distort beliefs about task performance in male-typed vs. female-typed domains in an Indian university context?

Hypotheses

H1: Stereotype Exaggeration

- Participants' self-beliefs align with gender stereotypes independent of actual ability.

H2: Gender Salience Amplification

- Revealing partner's gender amplifies belief distortions.

H3: Contextual Attenuation

- Smaller distortions at Ashoka (60% female) due to gender balance.

Experimental Design Flow (Part 1)

1. **Consent:** Participants read study info and agree
2. **Instructions:** Overview of categories, rules, belief elicitation
3. **Comprehension Check:** MCQs on incentives
4. **Ability Task — Cars:** 10 Qs + Self-belief
5. **Ability Task — Beauty:** 10 Qs + Self-belief

Experimental Design Flow (Part 2)

6. Partner Assignment (TREATMENT): Random assignment + pronoun reveal + partner score predictions

7. Demographics: Age, gender, major

8. Payment + Thank You

Linking Hypotheses to Survey Design

H1: Stereotype Exaggeration
Two gender-typed domains: **Cars (male)** and **Beauty (female)**.
Respondents:

- Answer 20 factual questions (actual ability)
- Predict their own score (self-belief)

If self-beliefs favour one's gender even after controlling for actual ability $\rightarrow$ stereotype distortion.

H2: Gender Salience Amplification
Partner introduced through a gendered pronoun $\rightarrow$ triggers gender salience.
Respondents:

- Predict partner's score in each domain
- Rate confidence in predictions

If revealing gender increases stereotype-aligned predictions $\rightarrow$ amplification effect.

Outcome Variables, Treatment, Measurement

Outcome Variables

1. **Self-belief:** Expected score for self
2. **Partner Confidence** Belief about partner's ability

Treatment: Gender Matching using Partner Gender Reveal

- 1 = indicates female
- 2 = indicates male
- 0 = control

Why & How We Generated Data Using GPT

Why synthetic (artificial) data?

- Time + recruitment constraints before exam period
- Needed a larger, balanced sample to test hypotheses
- Pilot full analysis pipeline before running real study
- Stress-test design (power, robustness, coding checks)

How we generated it

- Used a large language model (GPT) as a data generator
- Long, structured prompt specifying:
  - Experimental context + treatment structure
  - Variable names and distributions from Qualtrics
  - Behavioural patterns from Bordalo et al. (2019)
- Model instructed to output a clean CSV only

Synthetic Data Generation Prompt: Summary

Task to GPT Generate a **synthetic dataset of 400 participants** for our adaptation of *Bordalo et al. (2019)* at Ashoka University, with:

- Internally coherent observations and realistic heterogeneity
- Explicit stereotype-driven patterns in beliefs
- Added noise so that regressions and tests are non-trivial

Final output: **CSV only**, no extra text.

Sample Design & Treatments

- Total $N = 400$ Ashoka-style undergraduates
- Treatment: `treatment_pronoun` $= 0/1$
  - 200 control (no pronoun / gender-anonymous partner)
  - 200 treatment (partner gender made salient)
- Within treatment: `partner_gender_code` $= 0$ (control)/1(male pronoun)/2(female pronoun)
  - 100 male partners, 100 female partners
  - Indicators for `gender_matching` + `dyad_type`

What the Prompt Asked GPT to Simulate

Survey / decision structure encoded

- Two 10-question domains:
  - Cars (male-typed domain)
  - Beauty & Skincare (female-typed domain)
- After each domain:
  - Self-belief about own score in that domain
  - Randomly assigned partner + beliefs about partner's performance

Actual performance variables ("true ability")

- `cars_correct`, `beauty_correct` $\in \{0, \ldots, 10\}$
- Distributions with SD $\approx$ 1.5–2 and 5–10% "weird" types

Self-beliefs and distortions

- `selfbelief_cars`, `selfbelief_beauty` in [0,10]
- Added noise (SD $\approx$ 1–1.5), truncated to [0,10]

Partner Beliefs, Derived Variables & Targets

Partner beliefs (by treatment and gender-salience)

- Latent "true" partner ability drawn from same gendered distributions
- Variables:
  - `partner_cars_belief`, `partner_beauty_belief` $\in [0, 10]$
  - `partner_score_belief` $\in [0, 20]$ (aggregate score out of 20)
  - `partner_confidence` $\in [0, 100]$ (granular measure)
- Noise added (SD $\approx$ 2–3), truncated to feasible ranges

Derived variables and behavioural indicators

- Distortions:
  - `distort_self_cars = selfbelief_cars - cars_correct`
  - `distort_self_beauty = selfbelief_beauty - beauty_correct`
- Overconfidence dummies:
  - `overconfident_cars = 1(distort_self_cars > 0)`
  - `overconfident_beauty = 1(distort_self_beauty > 0)`
- Additional variables:
  - `pref_domain_self` (cars vs beauty vs equal)
  - `time_dummy_fast` (top 25% fastest responses)

Sample Profile

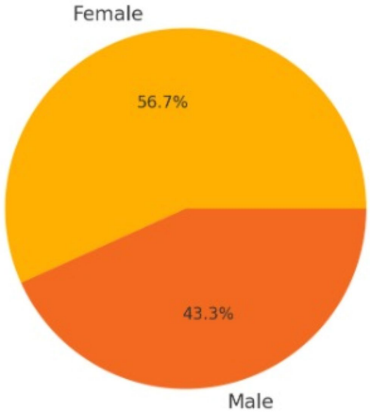


Figure A: Gender Split

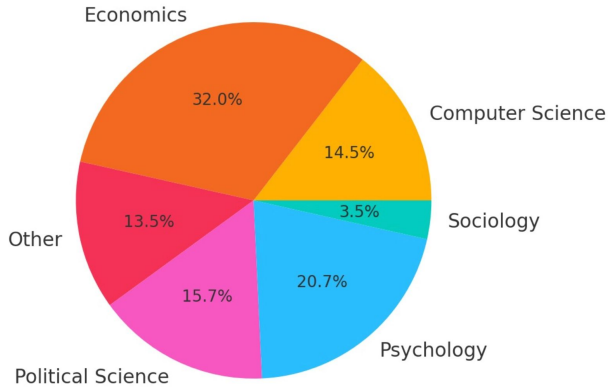


Figure B: Major Composition

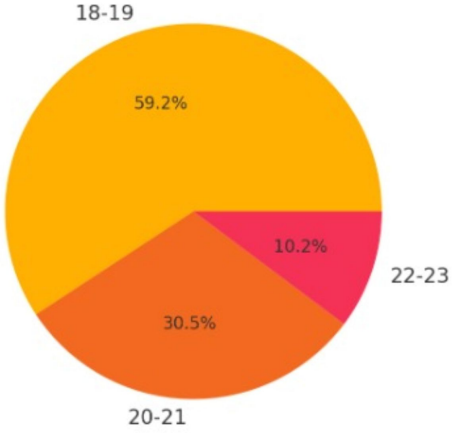


Figure C: Age Split

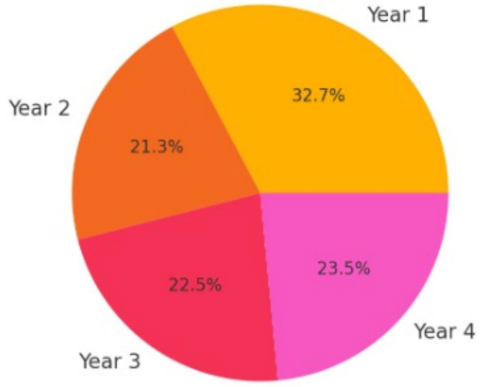


Figure D: Year of Study

Power Analysis: Parameters and Rationale

- The original paper reports stereotype effects in belief formation, using bank-level regressions.
- We base our power calculation on the stereotype coefficient ($\theta\sigma$) reported in:
  - **Table 4, Column 3**: beliefs about men $\rightarrow \theta\sigma = 0.45$
  - **Table 4, Column 4**: beliefs about women $\rightarrow \theta\sigma = 0.14$
- These are belief outcomes measured on a 0–1 scale, so we assume a conservative standard deviation of 0.23 which we estimated using simulated data.
- The power test evaluates whether an effect of this magnitude is detectable using a two-sample mean comparison (t-test framework), which maps cleanly onto the treatment versus control contrast in our design.

Power Result and Sample Choice

- Using the two published stereotype coefficients:
  - Large effect: $\Delta = 0.45$
  - Moderate effect: $\Delta = 0.14$
- Stata estimates the required sample sizes for 80% power (two-sided test, $\alpha = 0.05$) to detect each effect.
- Both required sample sizes fall well below 400 observations, meaning that any effect within the range observed in Bordalo et al. (2019) would be identifiable with substantially fewer subjects.
- Since we simulated **400 participants**, our experiment is **well powered** to detect stereotype-induced distortions in belief reporting, including smaller effects than those documented in the original study.
- Final note: our use of $N = 400$ reflects feasibility and design planning, rather than minimum detectable power.

Descriptive Analysis: Summary Statistics

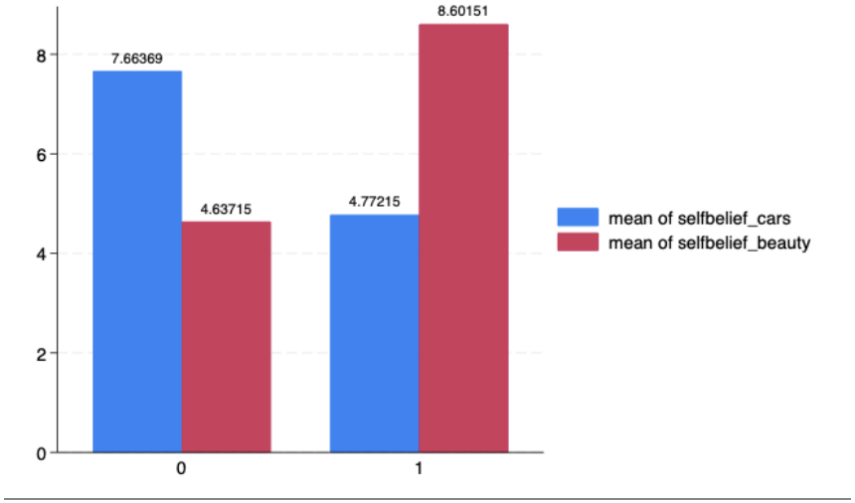


Figure 1: Self-Belief Index

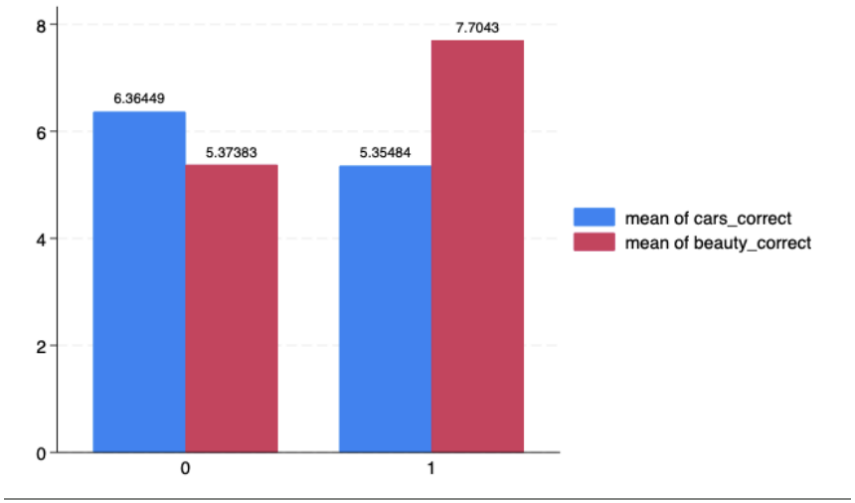


Figure 2: Score

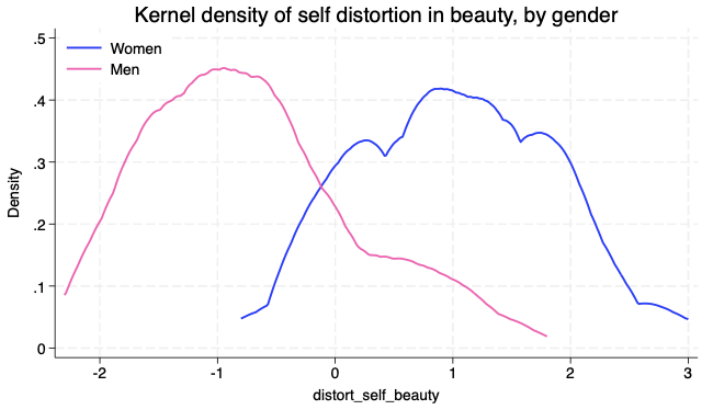
Gender coding: 0 = male, 1 = female.

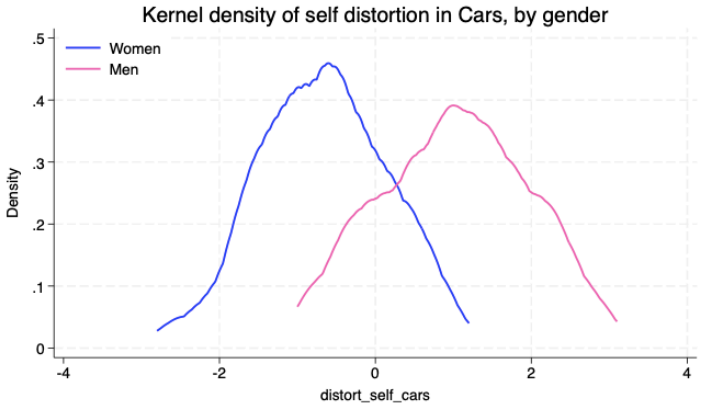
Self-beliefs differ more by gender than actual performance: men are more confident in cars, while women show higher belief in beauty.

Score differences are small, yet belief differences are large, indicating stereotype-driven confidence rather than ability.

Overall, the data suggests that confidence is gender-typed and amplified relative to real performance outcomes.

Descriptive Analysis: Graphs





Self-distortion in Beauty

Self-distortion in Cars

In Beauty, women's distortion distribution lies to the right of men's, indicating higher overconfidence relative to actual performance.

In Cars, the pattern reverses: men display stronger positive distortion, while women are more calibrated or underconfident.

This highlights confidence asymmetry: small score differences generate large, systematic belief gaps by gender.

Descriptive Analysis: Graphs

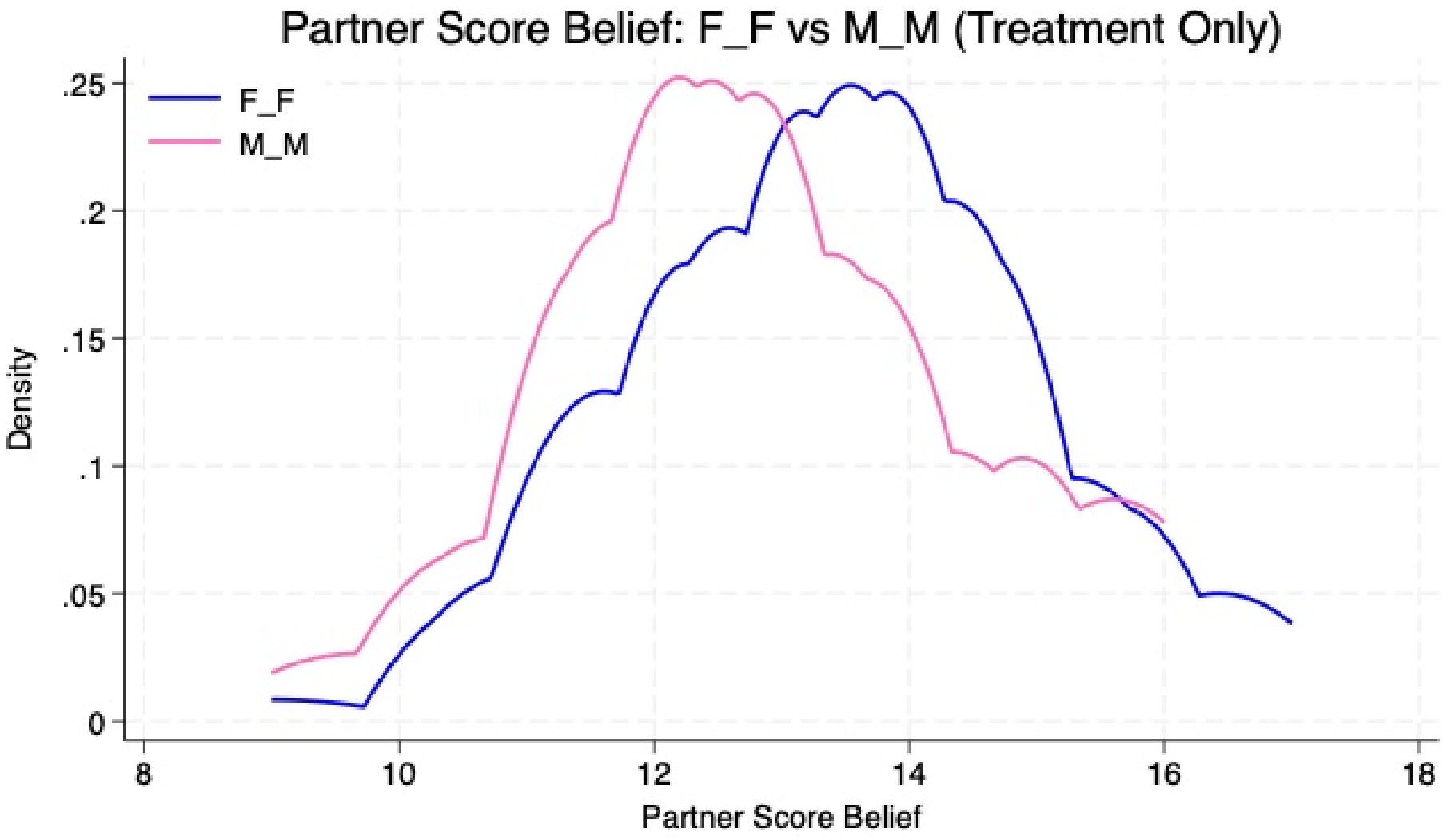


Figure 3: Graph

Descriptive Analysis: Graphs

- We restrict attention to the **treatment group** and study gender-matched vs. gender-mismatched dyads.
- For the treatment sample, we construct a categorical dyad variable capturing partner pairings, where

$$\texttt{dyad\_type} \in \{M\_F,\ F\_M,\ F\_F,\ M\_M\},$$

  with, for example, M_F denoting a *male subject matched with a female partner*.
- Kernel densities for all dyad combinations are reported in the log file; here we focus on the most informative case of gender-matched dyads.
- Male–male (M_M) dyads exhibit slightly lower partner-belief densities, suggesting moderate beliefs and competitive ingroup deflation.
- Female–female (F_F) dyads display more confidence projection consistent with altruistic evaluation.
- Women express greater performance optimism, while men remain less inflationary.

Statistical Test for Differences across genders

Significant gaps between men and women in both domains → confidence patterns align with stereotypes rather than performance. Distortion is measured by the difference in belief about performance and actual score received.

```
. ttest distort_self_cars, by(gender)

Two-sample t test with equal variances
```

Group	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
Female	227	-.6674009	.0545106	.8212842	-.7748148	-.559987
Male	173	1.016185	.0716771	.9427652	.8747049	1.157665
Combined	400	.06075	.0604698	1.209397	-.0581293	.1796293
diff		-1.683586	.0883945		-1.857364	-1.509807

```
    diff = mean(Female) - mean(Male)                              t = -19.0463
H0: diff = 0                                     Degrees of freedom =      398

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 1.0000
```

Figure (a): Distortion of beliefs (cars)

```
. ttest distort_self_beauty, by(gender)

Two-sample t test with equal variances
```

Group	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
Female	227	1.020705	.0560191	.8440125	.9103184	1.131091
Male	173	-.7236994	.0667885	.878466	-.8555301	-.5918687
Combined	400	.26625	.0609282	1.218564	.1464696	.3860304
diff		1.744404	.0867008		1.573955	1.914853

```
    diff = mean(Female) - mean(Male)                              t =  20.1198
H0: diff = 0                                     Degrees of freedom =      398

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

Figure (b): Distortion of beliefs (beauty)

Statistical Test for Differences Within Treatments

When looking at male respondents within the treatment, gender revelation leads to partner beliefs shifting toward stereotype-consistent expectations (MF = 1 and MM = 0)

Two-sample t test with equal variances

Group	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
0	56	7.839286	.1502742	1.124549	7.53813	8.140442
1	32	4.71875	.1864047	1.054464	4.338575	5.098925
Combined	88	6.704545	.1987152	1.864114	6.309577	7.099514
diff		3.120536	.2437173		2.636042	3.60503

diff = mean(0) - mean(1) t = 12.8039
H0: diff = 0 Degrees of freedom = 86

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
Pr(T < t) = 1.0000 Pr(|T| > |t|) = 0.0000 Pr(T > t) = 0.0000

Figure 1: Second-order beliefs about car score

```
. ttest partner_beauty_belief, by(dyad_MF_MM)
```

Two-sample t test with equal variances

Group	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
0	56	5.089286	.159701	1.195093	4.769238	5.409334
1	32	9.0625	.2196841	1.242721	8.614451	9.510549
Combined	88	6.534091	.2418736	2.268975	6.053341	7.014841
diff		-3.973214	.2686864		-4.507345	-3.439083

diff = mean(0) - mean(1) t = -14.7876
H0: diff = 0 Degrees of freedom = 86

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
Pr(T < t) = 0.0000 Pr(|T| > |t|) = 0.0000 Pr(T > t) = 1.0000

Figure 2: Second-order beliefs about beauty score

Statistical Test for Differences Within Treatments

Female participants report higher partner confidence when paired with men in the male-typed domain, and lower partner confidence when paired with women, even though this belief need not reflect true ability. (FF = 0 and FM = 1)

```
. ttest partner_cars_belief, by(dyad_FM_FF)

Two-sample t test with equal variances
```

Group	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
0	70	4.871429	.1301718	1.089095	4.611743	5.131114
1	61	7	.1258848	.9831921	6.748193	7.251807
Combined	131	5.862595	.1299367	1.487193	5.605531	6.11966
diff		-2.128571	.1823673		-2.48939	-1.767753

```
    diff = mean(0) - mean(1)                                  t = -11.6719
H0: diff = 0                                     Degrees of freedom =      129

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 1.0000
```

Figure 3: Second-order beliefs about car score

```
. ttest partner_beauty_belief, by(dyad_FM_FF)

Two-sample t test with equal variances
```

Group	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
0	70	8.585714	.1402347	1.173288	8.305954	8.865475
1	61	5.442623	.1594468	1.245319	5.123682	5.761564
Combined	131	7.122137	.1730602	1.980765	6.779758	7.464516
diff		3.143091	.2114687		2.724695	3.561487

```
    diff = mean(0) - mean(1)                                  t =  14.8632
H0: diff = 0                                     Degrees of freedom =      129

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

Figure 4: Second-order beliefs about car score

Statistical Tests of Belief Distortion: First- and Second-Order Effects

First-Order Distortions (Self-Belief vs Actual Score)

- Men significantly overestimate performance in Cars, women do so in Beauty.
- Distortion gaps persist even when performance does not differ equivalently.
- Women remain relatively cautious even when skilled, while men project confidence across domains.
- This reflects stereotype-consistent belief formation rather than pure ability assessment.

Statistical Tests of Belief Distortion: First- and Second-Order Effects

Second-Order Distortions (Partner Belief)

- Under pronoun treatment, partner-confidence shifts toward stereotyped expectations.
- In Cars: Female participants report higher confidence when matched with men (FM) than women (FF).
- In Beauty: Female–female dyads (FF) show higher partner-belief than mixed dyads.
- Suggests identity cues trigger representativeness amplification — small score differences inflate belief gaps.
- May also reflect gender differences in competitiveness vs cooperative evaluation.

Conclusion: Across self- and partner-beliefs, confidence is not an unbiased reflection of performance: it is *gender-filtered, stereotype-aligned, and treatment-responsive.*

Regression Equation

Baseline Interaction

$$BeliefOnPartnerConfidence_i = \beta_0 + \beta_1 Female_i + \beta_2 GenderMatching_i + \beta_3(Female_i \times GenderMatching_i) + \epsilon_i$$

Interaction with Controls

$$BeliefOnPartnerConfidence_i = \beta_0 + \beta_1 Female_i + \beta_2 GenderMatching_i + \beta_3(Female_i \times GenderMatching_i) + \beta_4 Age_i + \beta_5 Major_i + \beta_6 YearOfStudy_i + \epsilon_i$$

- We construct a categorical variable for gender-matching, defined as:

$$\texttt{gender\_matching} \in \begin{cases} 0 & = \text{Control (no gender revealed)} \\ 1 & = \text{Gender-matched pairs } (F\_F,\ M\_M) \\ 2 & = \text{Gender-mismatched pairs } (M\_F,\ F\_M) \end{cases}$$

- Interpretation of **coefficient of interest (β_3)**: The female-male difference in matching relative to the female-male difference in control.

OLS

```
. reg partner_confidence i.female##i.gender_matching
```

Source	SS	df	MS		
Model	12242.5181	5	2448.50362	Prob > F	0.0000
Residual	39745.2794	394	100.876344	R-squared	0.2355
Total	51987.7975	399	130.295232	Adj R-squared	0.2258
				Root MSE	10.044

partner_confidence	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
1.female	-.7140931	1.495853	-0.48	0.633	-3.654945	2.226758
gender_matching						
1	13.62668	1.728625	7.88	0.000	10.2282	17.02516
2	6.077574	2.083067	2.92	0.004	1.982258	10.17289
female#gender_matching						
1 1	-3.525193	2.340946	-1.51	0.133	-8.127501	1.077115
1 2	4.57731	2.653992	1.72	0.085	-.6404459	9.795067
_cons	60.14118	1.089395	55.21	0.000	57.99942	62.28293

Figure: OLS Regression Output 1

OLS Regression: Partner Confidence (Baseline)

- We regress second-order beliefs (partner confidence) on gender of subject, gender-matching, and their interaction.
- Female coefficient is near zero and insignificant $\Rightarrow$ no baseline gender gap in partner expectations.
- GenderMatching is large and significant ($\beta_2 \approx 13.6$, $p < 0.001$) $\Rightarrow$ matched partners are rated higher.
- Female $\times$ GenderMatching positive but weak $\Rightarrow$ women do not inflate partner-belief as strongly as men.

Takeaway: Matching drives confidence, and not gender alone. In-group expectations boost perceived partner ability.

OLS

```
. reg partner_confidence i.female##i.gender_matching i.agenumeric i.major_num i.year_of_study
```

Source	SS	df	MS		
Model	13230.1447	15	882.009645	Prob > F	0.0000
Residual	38757.6528	384	100.931388	R-squared	0.2545
				Adj R-squared	0.2254
Total	51987.7975	399	130.295232	Root MSE	10.046

Number of obs	=	400				
F(15, 384)	=	8.74				

partner_confidence	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
1.female	-1.453279	1.529537	-0.95	0.343	-4.460595	1.554037
gender_matching						
1	13.99977	1.786404	7.84	0.000	10.48741	17.51212
2	6.112524	2.13285	2.87	0.004	1.918998	10.30605
female#gender_matching						
1 1	-3.408218	2.38048	-1.43	0.153	-8.088625	1.272188
1 2	4.913878	2.680378	1.83	0.068	-.3561766	10.18393

Figure: OLS Regression Output 2

OLS w/ Controls: Age, Major, Year

- Adding controls barely changes estimates. Gender-matching remains highly predictive.
- $\beta_2 \approx 14$ with $p < 0.001 \Rightarrow$ partner-belief increases sharply when genders align.
- Interaction term now positive and marginally significant: female-matching amplifies belief slightly in pooled tasks.
- Controls increase model fit ($R^2 \approx 0.25$) but do not change core inference.

Inference: Partner-belief differences are driven by identity cues, not demographic structure or ability signals.

OLS

```
. reg partner_beauty_belief i.female##i.gender_matching i.agenumeric i.major_num i.year_of_study
```

Source	SS	df	MS		
Model	679.901136	15	45.3267424	Number of obs	= 400
Residual	520.008864	384	1.35418975	F(15, 384)	= 33.47
				Prob > F	= 0.0000
				R-squared	= 0.5666
				Adj R-squared	= 0.5497
Total	1199.91	399	3.00729323	Root MSE	= 1.1637

partner_beauty_belief	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
1.female	-.0492488	.1771685	-0.28	0.781	-.3975907	.299093
gender_matching						
1	-2.147954	.2069219	-10.38	0.000	-2.554796	-1.741112
2	1.857201	.2470511	7.52	0.000	1.371458	2.342943
female#gender_matching						
1 1	3.57362	.2757345	12.96	0.000	3.031481	4.115758
1 2	-3.545534	.3104722	-11.42	0.000	-4.155973	-2.935096

Figure: OLS Regression Output 3

OLS: Partner Belief in Beauty Domain

- Strong effects appear when looking only at beauty (a female-typed task.)
- Female dummy remains insignificant, but GenderMatching is large and highly significant ($p < 0.001$).
- Female $\times$ GenderMatching is strongly positive (≈ 3.58, $p < 0.001$) $\Rightarrow$ females trust female partners more in Beauty.
- In mixed dyads, confidence is lower, confirming stereotype congruence mechanics.

Conclusion: When stereotypes align (Beauty–Female), matching boosts belief formation dramatically, implying identity amplifies expectation.

OLS

```
. reg partner_cars_belief i.female##i.gender_matching i.agenumeric i.major_num i.year_of_study
```

Source	SS	df	MS		
Model	425.088314	15	28.339221	Prob > F	0.0000
Residual	483.289186	384	1.25856559	R-squared	0.4680
Total	908.3775	399	2.27663534	Root MSE	1.1219

	Number of obs =	400
	F(15, 384) =	22.52
	Prob > F =	0.0000
	R-squared =	0.4680
	Adj R-squared =	0.4472
	Root MSE =	1.1219

partner_cars_belief	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
1.female	-.6781079	.1707988	-3.97	0.000	-1.013926	-.34229
gender_matching						
1	1.221488	.1994824	6.12	0.000	.8292735	1.613703
2	-1.905547	.2381689	-8.00	0.000	-2.373825	-1.437268
female#gender_matching						
1 1	-2.375269	.265821	-8.94	0.000	-2.897915	-1.852622
1 2	2.931442	.2993098	9.79	0.000	2.34295	3.519933

Figure: OLS Regression Output 4

OLS Regression: Partner Beliefs in Cars

- Outcome: second-order beliefs about partner ability in the car domain.
- Female dummy is negative and highly significant ($\beta_1 \approx -0.67, p < 0.001$), indicating lower partner-confidence among women overall.
- Gender matching increases confidence ($\beta_2 \approx +1.22, p < 0.001$) i.e. same-gender pairs are rated higher.
- Interaction splits sharply by stereotype:
  - FF pairs $\Rightarrow$ strong down-weighting of partner ability ($-2.38, p < 0.001$)
  - FM pairs $\Rightarrow$ large confidence uplift for male partners ($+2.93, p < 0.001$)

Conclusion: In Cars, confidence is allocated along gender-stereotype lines: female–female teams are penalized, male partners are rewarded.

Caveats & Interpretation

Limitations

- Reflects model priors, not real cognition
- Lower behavioural noise
- Stereotype effects may appear cleaner
- limited to 2 categories
- Did not incorporate DIM
- Ashoka not representative of broader Indian University

How to use the data

- Pilot for coding and design
- Not a substitute for real data

Conclusion and Policy Takeaways

- Our major contribution: proving the gender salience hypothesis. In Bordalo's study, the results on this were insignificant.
- Broadly, the expectations that we had for beliefs based on gender and gender type of questions hold.
- This has important implications for policy:
  - Designing environments that are conducive to gender-neutral contributions so as to control for "perceptions" or "stereotype" led discourse and outcomes.
  - Designing curriculum that starts accounting for stereotype-led belief formation, particularly with regards to certain topics that may be associated with genders and lead to the emphasising of gender norms.
  - Both can be done by framing narratives differently.