

A (DYNAMIC) INVESTIGATION OF STEREOTYPES, BELIEF-UPDATING, AND BEHAVIOR

Katherine Coffman¹ Paola Ugalde A.² Basit Zafar³

¹Harvard Business School

²Arizona State University

³University of Michigan

December 5, 2020

MOTIVATION

Large gender differences in human capital choices. For decision-making under uncertainty, both preferences and beliefs matter.

- ▶ Large lab literature on (gender differences in) beliefs, but less work on the link with decision-making
- ▶ Large field literature on (gender differences in) decisions, but little work on beliefs (primarily due to data limitations)
- ▶ Even less work on the dynamics of belief updating and choices
(Zimmermann, 2020)

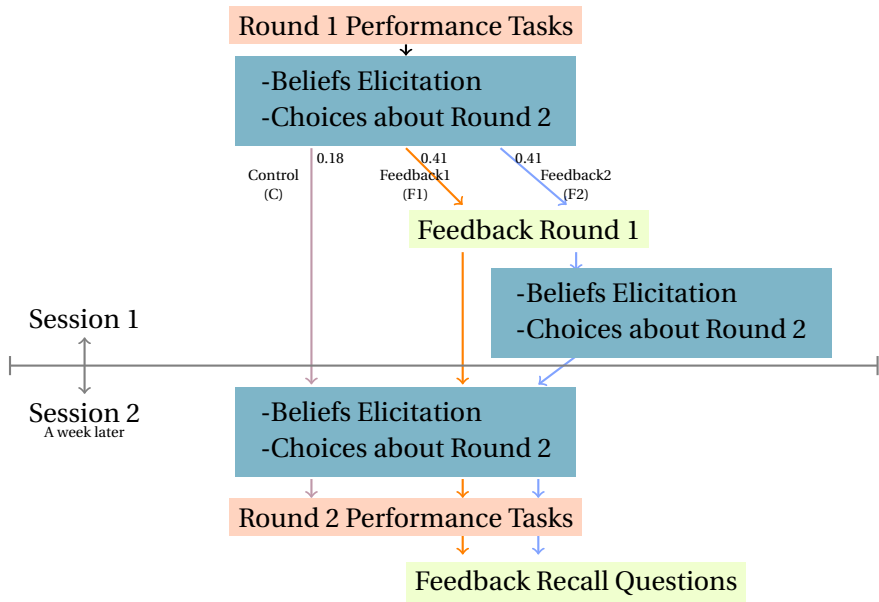
THIS PAPER

We designed a 2-stage lab experiment that explores how individuals update beliefs over time, and how these beliefs map into decisions

- ▶ Explore men and women across male and female-typed domains
 - Idea: Separate gender differences from impact of stereotyping
- ▶ Collect not only detailed beliefs data but also choices about willingness to compete in each domain at **multiple points in time** before and after feedback

Design mimics real-world settings. We have full control over the environment and observe performance/beliefs/signals.

EXPERIMENTAL DESIGN



CHOICES ABOUT ROUND 2

EXPERIMENTAL DESIGN

- ▶ Participants make choices about how they want to be compensated in Round 2

1. Under a fixed payment scheme

\$1 per problem solved correctly on the Math test in Round 2	\$1 per problem solved correctly on the Verbal test in Round 2
---	---

2. On a price list for willingness to compete given different per-problem reward:

\$4 per problem solved correctly on the Math test in Round 2 if in top-4, \$0 otherwise	\$1 per problem solved correctly on the Verbal test in Round 2
\$3.5 per problem solved correctly on the Math test in Round 2 if in top-4, \$0 otherwise	\$1 per problem solved correctly on the Verbal test in Round 2
\$3 per problem solved correctly on the Math test in Round 2 if in top-4, \$0 otherwise	\$1 per problem solved correctly on the Verbal test in Round 2
\$2.5 per problem solved correctly on the Math test in Round 2 if in top-4, \$0 otherwise	\$1 per problem solved correctly on the Verbal test in Round 2
\$2 per problem solved correctly on the Math test in Round 2 if in top-4, \$0 otherwise	\$1 per problem solved correctly on the Verbal test in Round 2
\$1.5 per problem solved correctly on the Math test in Round 2 if in top-4, \$0 otherwise	\$1 per problem solved correctly on the Verbal test in Round 2

Our key outcome is the lowest reward at which they were willing to compete in that domain (WTA)

FEEDBACK PROVISION

EXPERIMENTAL DESIGN

- ▶ Participants get feedback on performance relative to a reference group
 - 9 participants that took exact same quizzes before the main experiment, from same subject population
- ▶ For each quiz, participant learns how they performed relative to randomly chosen person from reference group: better or worse
 - Ties are randomly broken

		Math	
		GG	GB
		F:0.13, M:0.21	F:0.22, M:0.18
Verbal		BG	BB
		F:0.16, M:0.23	F:0.50, M:0.38

DATA

- ▶ Online experiment at Arizona State University
 - Average compensation \$19.5
- ▶ 2046 students completed week 1
 - Sample sizes were specified in the AEA registry (AEARCTR-0005712)
- ▶ 1816 students completed week 2 ▶
 - No differential attrition by gender, treatment, feedback or performance ▶

	Female	Male	Total
F1	461	303	764
F2	411	278	689
C	222	141	363
Total	1,094	722	1,816

REGRESSION MODEL

$$\text{Outcome}_{iD} = \beta_0 + \beta_1 \text{Female}_i + \beta_2 \text{Congruent}_{iD} + \mathbf{Y}_i + \mathbf{X}_i + \epsilon_{iD}$$

- $D \in \{\text{Verbal, Math}\}$
- $\text{Outcome}_{iD} \in \{\text{Beliefs, WTA}\}$
- $\text{Female}_i = \mathbb{1}(\text{Female})$
- $\text{Congruent}_{iD} = \mathbb{1}(D \text{ congruent with } i\text{'s gender})$
- \mathbf{Y}_i :
 - Number of correct answers in D and $-D$, respectively
 - Position 1-10 given performance in D and $-D$, respectively
- \mathbf{X}_i : family income, ACT, minority, parents education, honors, risk aversion, F2

REGRESSION MODEL

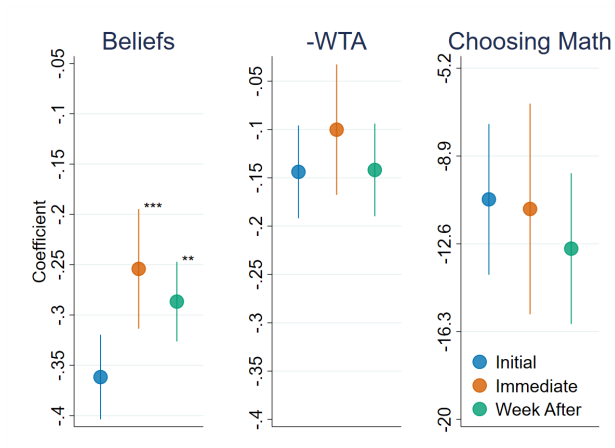
$$Outcome_{iD} = \beta_0 + \beta_1 Female_i + \beta_2 Congruent_{iD} + \mathbf{Y}_i + \mathbf{X}_i + \epsilon_{iD}$$

$$\mathbb{1}(\text{Math})_i = \beta_0 + \beta_1 Female_i + \mathbf{Y}_i + \mathbf{X}_i + \epsilon_i$$

- $D \in \{\text{Verbal, Math}\}$
- $Outcome_{iD} \in \{\text{Beliefs, WTA}\}$
- $Female_i = \mathbb{1}(\text{Female})$
- $Congruent_{iD} = \mathbb{1}(D \text{ congruent with } i\text{'s gender})$
- \mathbf{Y}_i :
 - Number of correct answers in D and $-D$, respectively
 - Position 1-10 given performance in D and $-D$, respectively
- \mathbf{X}_i : family income, ACT, minority, parents education, honors, risk aversion, F2

β_1 : GENDER GAPS

OVER TIME

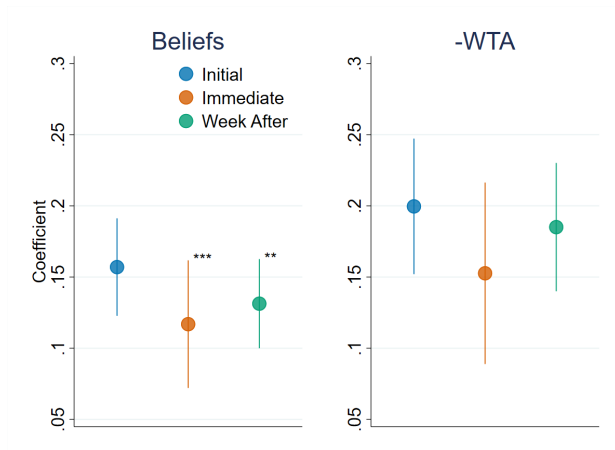


Note: 90% CI

- ▶ Beliefs respond strongly to feedback, though a week later some of the impact fades
- ▶ Impact on choices is much more muted

β_2 : STEREOTYPE DIFFERENCES

OVER TIME



Note: 90% CI

- ▶ More optimistic beliefs in gender-congruent domains. The difference closes after feedback
- ▶ Participants more willing to compete in gender-congruent domains. The gender congruency effect gets weaker post-feedback.

UNPACKING THE GAPS

Recall

- ▶ High accuracy in recall of feedback one week later ▶
 - Women significantly more likely to accurately recall feedback than men ▶
- ▶ Both women and men significantly more likely to recall bad feedback than good

Good News vs Bad News

- ▶ Bad news has a stronger effect than good news ▶ ▶ ▶
- ▶ Feedback does not bring participants closer to their maximal achievable earnings ▶

CONCLUSION

We find:

- ▶ Significant gender and gender stereotype gaps in prior beliefs about own ability.
- ▶ Choices are strongly linked with prior beliefs
- ▶ Feedback:
 - Modest impact on beliefs, but most of the impact dissipates a week later
 - Limited impact on choices. Any impacts are short-lived.

Implications:

- ▶ Gender differences and impact of stereotypes quite sticky.
- ▶ Patterns are not due to selective recall.
 - Suggests need for more work on importance/drivers of motivated beliefs
- ▶ One-time feedback interventions are not likely to be sufficient for creating lasting impact on dynamic decisions

EXTRA SLIDES

QUESTIONS EXAMPLES

► Verbal

Choose the word that best completes the following sentence.

The employee was never the most efficient worker, but his many years of unwavering _____ to the firm made him one of the most trusted advisers even to the chief executive.

- servitude
- penury
- drudgery
- fidelity

► Math

Jack is three times as old as Jill, but in three years he will only be twice as old. How old is Jack?

- 9
- 12
- 6
- 3



EXPERIMENT INSTRUCTIONS

Beliefs about Round 1 Performance: Verbal

Please provide a guess of how many problems out of 12 you solved correctly on the **Verbal** test. If this is your decision-that-counts and your answer to this question is correct, you will earn \$1 in additional payment.

Next



EXPERIMENT INSTRUCTIONS

Beliefs about Round 1 Performance: Verbal

You just told us that you believe you had a score of 3 on the **Verbal** test. What do you think the chances are that you had exactly that score? If this is your decision-that-counts, you can earn an additional \$1 based upon your answer. You have an incentive to tell us your best, honest guess (click [here](#) to learn more about your incentives).

Move the slider below to indicate how you'd fill in the following: "I believe I have a ___% chance of having exactly a score of 3 on the **Verbal** test."



Next



EXPERIMENT INSTRUCTIONS

Beliefs about Round 1 Performance: Verbal

Now we are curious how you think your **Verbal** performance compares to the performance of the **reference group** (that is, the 9 ASU students who also completed the **Verbal** test).

If we compared your **Verbal** test score to the scores of the **reference group**, how do you think you would rank? If this is your decision-that-counts and your answer to this question is correct, you will earn \$1 in additional payment.

- First: better than all other 9 participants.
- Second: worse than 1, better than 8 other participants.
- Third: worse than 2, better than 7 other participants.
- Fourth: worse than 3, better than 6 other participants.
- Fifth: worse than 4, better than 5 other participants.
- Sixth: worse than 5, better than 4 other participants.
- Seventh: worse than 6, better than 3 other participants.
- Eighth: worse than 7, better than 2 other participants
- Ninth: worse than 8, better than 1 other participant.
- Last: worse than all other 9 participants.



EXPERIMENT INSTRUCTIONS

Beliefs about Round 1 Performance: Verbal

The purpose of this question is to ask you about the chance that you rank in each of the following positions when your performance on the **Verbal** test is compared to the **reference group** (that is, the 9 ASU students who also completed the **Verbal** test). If this is your "decision-that-counts", you can earn an additional \$1 based upon your answer. You have an incentive to tell us your best, honest guess (click [here](#) to learn more about your incentives).

What do you think the chances are that you rank in each of the following positions relative to the **reference group**? The sum of the probabilities should add up to 100.

Ranking	Probability
First: better than all other 9 participants	<input type="text"/>
Second: worse than 1, better than 8 other participants	<input type="text"/>
Third: worse than 2, better than 7 other participants	<input type="text"/>
Fourth: worse than 3, better than 6 other participants	<input type="text"/>
Fifth: worse than 4, better than 5 other participants	<input type="text"/>
Sixth: worse than 5, better than 4 other participants	<input type="text"/>
Seventh: worse than 6, better than 3 other participants	<input type="text"/>
Eighth: worse than 7, better than 2 other participants	<input type="text"/>
Ninth: worse than 8, better than 1 other participant	<input type="text"/>
Last: worse than all other 9 participants	<input type="text"/>
Total:	0



DATA

SUMMARY STATISTICS

	Female	Male	P-value
Score Math R1	5.74	6.49	0.00
Score Verbal R1	6.94	7.17	0.07
Score Math R2	4.18	4.77	0.00
Score Verbal R2	4.87	5.11	0.01
Academic Characteristics			
GPA	3.69	3.64	0.01
ACT	22.75	25.31	0.00
% Honors	58.04	60.94	0.22
Beliefs Before Feedback			
Math score	6.36	7.73	0.00
Verbal score	7.18	7.62	0.00
Math rank	5.26	6.78	0.00
Verbal rank	5.93	6.57	0.00
Choices Before Feedback			
WTA Math (¢)	337.51	285.14	0.00
WTA Verbal (¢)	297.09	296.51	0.91
% Chooses Math	0.40	0.56	0.00
<i>Total</i>	1,094	722	



ATTRITION RATE

	All	F1+F2
	(1)	(2)
Female (F)	-0.005 (0.029)	0.003 (0.025)
F1	-0.005 (0.027)	
F1*F	0.031 (0.036)	
F2	0.002 (0.027)	
F2*F	0.020 (0.036)	
Bad News _{Verbal}		0.005 (0.026)
Bad News _{Verbal} *F		0.031 (0.030)
Bad News _{Math}		0.042 (0.026)
Bad News _{Math} *F		-0.003 (0.030)
Score _{Math}	-0.003 (0.004)	0.003 (0.006)
Score _{Verbal}	0.002 (0.003)	0.003 (0.004)
Mean	0.108	0.111
R2	0.054	0.049
Obs.	2,046	1,642



WEEK AFTER

	Before		Immediately After		Week After	
	Female	Male	Female	Male	Female	Male
Ranking Belief	5.60	6.67	5.28	6.30	5.40	6.45
Ranking Belief Good News	6.58	7.44	7.19	7.76	7.02	7.59
Ranking Belief Bad News	5.27	6.30	4.38	5.36	4.86	5.88
WTA (¢)	317.28	290.87	319.93	298.66	327.97	298.65
WTA Good News (¢)	279.30	263.29	248.60	241.71	270.02	257.80
WTA Bad News (¢)	331.06	304.60	356.33	335.49	348.58	319.42
% Chooses Math	39.67	56.23	39.90	57.91	38.48	56.93
% Chooses Math Good News	54.80	62.20	61.02	72.27	54.80	64.96
% Chooses Math Bad News	35.19	52.99	31.40	47.17	33.65	52.56



RANK BELIEFS

EFFECT OF PRIORS

	Prior		Immediately After		Week After	
	(1)	(2)	(3)	(4)	(5)	(6)
Female (F)	-0.33*** (0.03)	-0.24*** (0.04)	-0.04 (0.03)	-0.27*** (0.03)	-0.02 (0.02)	
Congruent	0.15*** (0.02)	0.11*** (0.03)	0.01 (0.02)	0.12*** (0.02)	0.02 (0.01)	
Prior			✓		✓	
Mean	0.00	0.00	0.00	0.00	0.00	
R2	0.42	0.49	0.81	0.49	0.80	
Clusters	1,453	689	689	1,453	1,453	
Obs.	2,906	1,378	1,378	2,906	2,906	

Conditional on performance:

- ▶ Prior beliefs explain gender and stereotype gaps after feedback



CHOICES: 1 (MATH)

EFFECT OF PRIORS

	Prior	Immediately After	Week After		
	(1)	(2)	(3)	(4)	(5)
Female (F)	-10.73*** (2.45)	-11.13*** (3.46)	-5.36** (2.18)	-12.80*** (2.48)	-5.53*** (1.86)
Prior			✓		✓
Mean	45.91	47.17	47.17	44.87	44.87
R2	0.24	0.31	0.73	0.24	0.59
N	1,453	689	689	1,453	1,453

Conditional on performance:

- ▶ There is a gender gap even conditional on prior choice.



CHOICES: WTA

THE EFFECT OF PRIORS

	Initial	Immediately After	Week After		
	(1)	(2)	(3)	(4)	(5)
Female (F)	15.36*** (4.00)	11.06* (5.81)	-1.21 (3.96)	15.05*** (3.97)	6.21* (3.26)
Congruent	-21.33*** (3.97)	-16.85*** (5.48)	-2.82 (3.46)	-19.64*** (3.73)	-7.00** (2.76)
Prior WTA			✓		✓
Mean	305.44	311.13	311.13	316.56	316.56
R2	0.16	0.23	0.68	0.20	0.52
Clusters	1,361	653	653	1,361	1,361
Obs.	2,603	1,262	1,262	2,627	2,627

CHOICES: WTA

THE EFFECT OF PRIORS

	Initial	Immediately After	Week After		
	(1)	(2)	(3)	(4)	(5)
Female (F)	15.36*** (4.00)	11.06* (5.81)	-1.21 (3.96)	15.05*** (3.97)	6.21* (3.26)
Congruent	-21.33*** (3.97)	-16.85*** (5.48)	-2.82 (3.46)	-19.64*** (3.73)	-7.00** (2.76)
Prior WTA			✓		✓
Mean	305.44	311.13	311.13	316.56	316.56
R2	0.16	0.23	0.68	0.20	0.52
Clusters	1,361	653	653	1,361	1,361
Obs.	2,603	1,262	1,262	2,627	2,627

Conditional on performance:

- ▶ Prior WTA completely explains gender gaps immediately after feedback
 - The gap reappears a week after

CHOICES: WTA

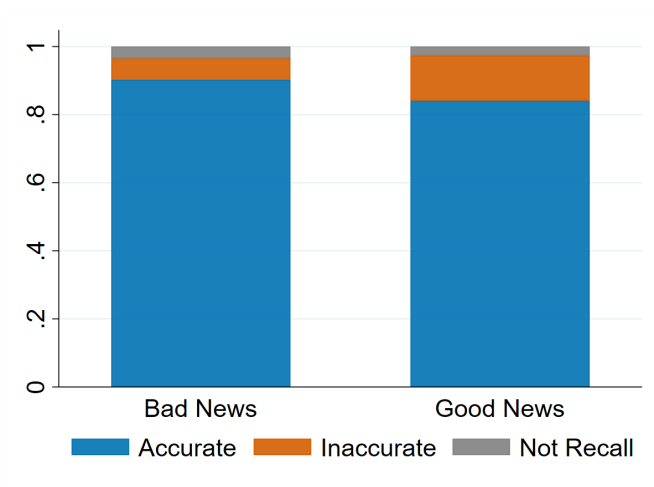
THE EFFECT OF PRIORS

	Initial	Immediately After	Week After		
	(1)	(2)	(3)	(4)	(5)
Female (F)	15.36*** (4.00)	11.06* (5.81)	-1.21 (3.96)	15.05*** (3.97)	6.21* (3.26)
Congruent	-21.33*** (3.97)	-16.85*** (5.48)	-2.82 (3.46)	-19.64*** (3.73)	-7.00** (2.76)
Prior WTA			✓		✓
Mean	305.44	311.13	311.13	316.56	316.56
R2	0.16	0.23	0.68	0.20	0.52
Clusters	1,361	653	653	1,361	1,361
Obs.	2,603	1,262	1,262	2,627	2,627

Conditional on performance:

- ▶ Prior WTA completely explains gender gaps immediately after feedback
 - The gap reappears a week after
- ▶ Prior WTA completely explains stereotype differences immediately after feedback
 - Differences reappear a week after

FEEDBACK RECALL



- ▶ Participants more likely to recall negative feedback



FEEDBACK RECALL

OUTCOME: 1 (RECALL FB)

	(1)	(2)	(3)
Female (F)	0.06***	0.06***	0.07***
	(0.02)	(0.02)	(0.02)
Congruent	0.01	0.01	-0.00
	(0.01)	(0.01)	(0.01)
Good News _D		-0.09***	-0.08***
		(0.02)	(0.03)
Good News _D * Congruent			0.03
			(0.03)
Good News _D *F			-0.04
			(0.03)
Good News _{-D}		-0.01	-0.01
		(0.02)	(0.02)
Group F2	-0.01	-0.01	-0.01
	(0.01)	(0.01)	(0.01)
Mean	0.88	0.88	0.88
R2	0.03	0.04	0.04
Clusters	1,453	1,453	1,453
Obs.	2,906	2,906	2,906



GOOD VS BAD NEWS

BELIEFS ABOUT RANKING

	(1)	(2)	(3)
Female (F)	-0.16*** (0.04)	-0.15*** (0.04)	-0.20** (0.08)
Congruent	0.09*** (0.03)	0.09*** (0.03)	0.20*** (0.06)
Prior	0.67*** (0.01)	0.67*** (0.01)	0.67*** (0.01)
Good News _D		0.34*** (0.06)	0.28*** (0.09)
Good News _D *F			0.27** (0.10)
Good News _D * Congruent			-0.17** (0.08)
Bad News _D		-0.74*** (0.06)	-0.63*** (0.09)
Bad News _D *F			-0.07 (0.10)
Bad News _D * Congruent			-0.13* (0.08)
Good News _D = Bad News _D		0.02	
Mean	5.82	5.82	5.82
R2	0.72	0.75	0.75
Clusters	1,816	1,816	1,816
Obs.	3,632	3,632	3,632

Conditional on performance and prior beliefs:

- ▶ Bad news have a stronger effect than good news
- ▶ Good news effect is stronger for women and weaker on gender congruent domains
- ▶ Bad news effect is stronger in gender congruent domains



GOOD VS BAD NEWS

CHOICE: 1 (MATH) WEEK AFTER

	(1)	(2)	(3)
Female (F)	-5.26*** (1.64)	-5.21*** (1.63)	-4.29 (3.33)
Chose Math in Prior	0.69*** (0.02)	0.69*** (0.02)	0.69*** (0.02)
Good News _{Math}		0.32 (2.40)	0.74 (3.24)
Good News _{Math} *F			-0.66 (4.31)
Bad News _{Math}		-6.24*** (2.22)	-5.32 (3.30)
Bad News _{Math} *F			-1.45 (3.96)
Good News = Bad News		0.14	
Mean	45.81	45.81	45.81
R2	0.60	0.60	0.60
Obs.	1,816	1,816	1,816

Conditional on performance and priors;

- ▶ Bad news in math has larger effect than good news in math
- ▶ There is no differential effect by gender



GOOD VS BAD NEWS

CHOICE: WTA WEEK AFTER

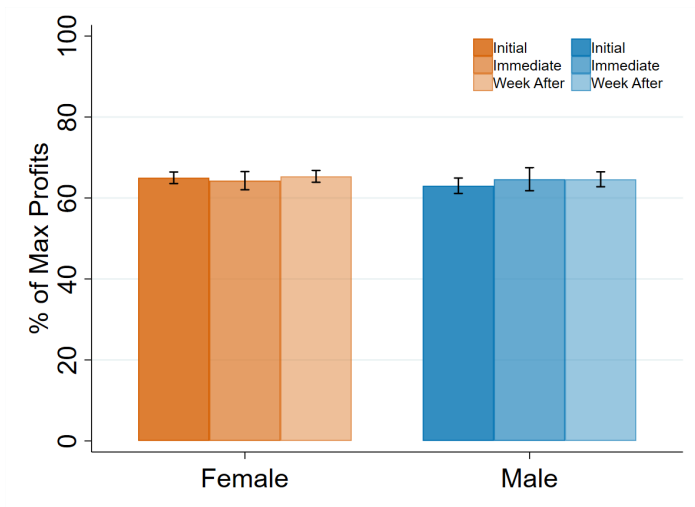
	(1)	(2)	(3)
Female (F)	7.23** (2.84)	7.02** (2.81)	14.22** (5.64)
Congruent	-5.52** (2.42)	-5.58** (2.38)	-1.17 (4.86)
Prior WTA	0.64*** (0.02)	0.64*** (0.02)	0.64*** (0.02)
Good News _D		-16.81*** (4.14)	-10.08 (6.22)
Good News _D * Congruent			1.24 (6.21)
Good News _D *F			-13.07* (7.19)
Bad News _D		17.14*** (3.96)	25.57*** (5.95)
Bad News _D * Congruent			-9.41 (5.99)
Bad News _D *F			-6.36 (7.01)
Good News _D = Bad News _D		0.96	
Mean	316.08	316.08	316.08
R2	0.53	0.54	0.54
Clusters	1,708	1,708	1,708
Obs.	3,284	3,284	3,284

Conditional on performance and prior beliefs:

- ▶ Bad news have a stronger effect than good news
- ▶ Good news effect is stronger for women

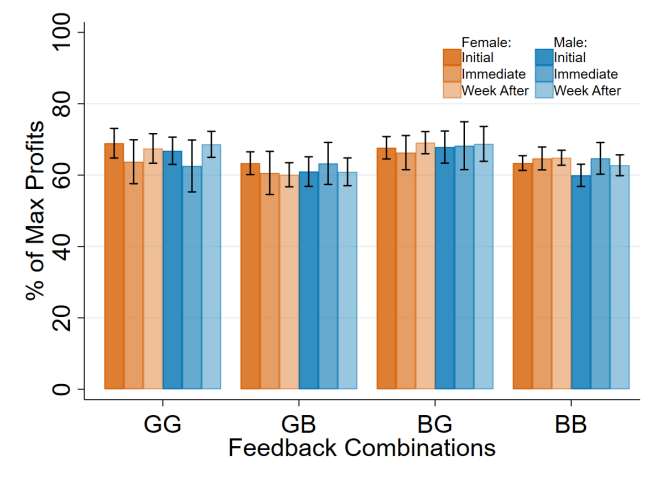


EXPECTED EARNINGS



EXPECTED EARNINGS

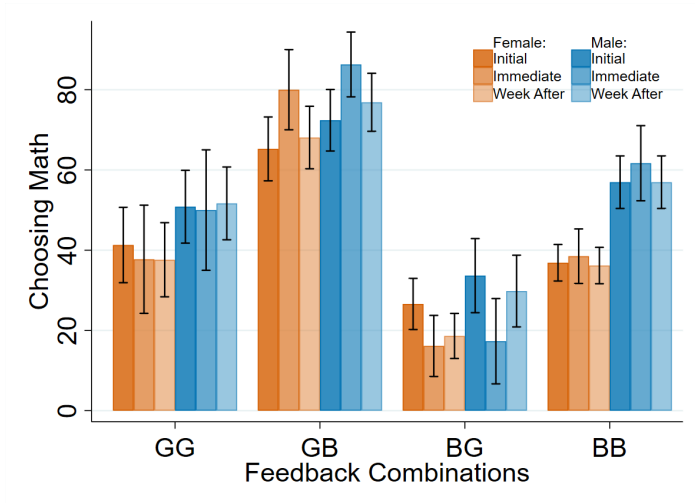
BY FEEDBACK TYPE



Note: GG: good news in both domains. BB: bad news in both domains. GB: good news in math, bad news in verbal. BG: bad news in math, good news in verbal.

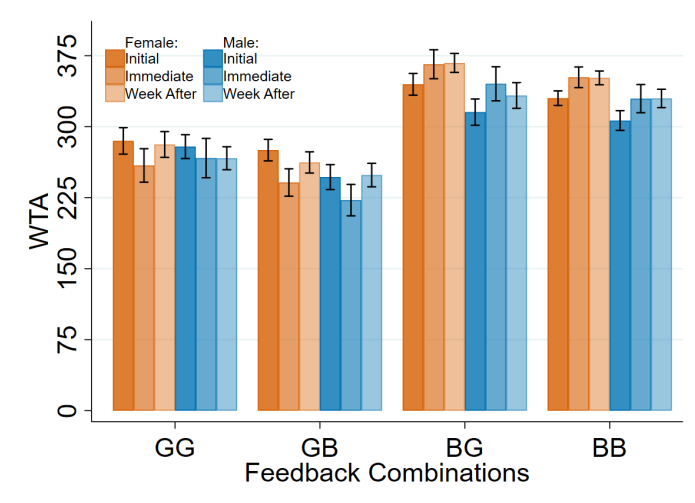


CHOICES: 1 (MATH)



Note: GG: good news in both domains. BB: bad news in both domains. GB: good news in math, bad news in verbal. BG: bad news in math, good news in verbal.

CHOICES: WTA



Note: GG: good news in both domains. BB: bad news in both domains. GB: good news in D , bad news in $-D$. BG: bad news in D , good news in $-D$.