Awards and the gender gap in knowledge contributions in STEM

By JANA GALLUS* AND EMMA HEIKENSTEN†

* Gallus (corresponding author): UCLA Anderson, 110 Westwood Plaza, Los Angeles, CA 90095 (e-mail: jana.gallus@anderson.ucla.edu). † Heikensten: SEB, Stjärntorget 4 16979 Solna, Sweden (e-mail: emma.heikensten@seb.se). We acknowledge financial support by the Harvard Kennedy School's Women and Public Policy Program, the Lara Warner Scholarship, Jan Wallanders och Tom Hedelius stiftelse, the Swiss National Science Foundation and the Hellman Fellows Fund.

I. Introduction

Teams and organizations miss out if their smartest members do not bring their ideas to the table. Previous work finds that there are gender differences in the tendency to speak up, and that self-stereotyping provokes such differences even when fear of discrimination is ruled out (Baldiga Coffman 2014).

Gallus and Heikensten (2019) show that gender differences in the tendency to speak up play an important role for female participation in STEM (science, technology, engineering and math) and, specifically, in math. Moreover, their lab experiment shows that these differences can be eliminated by providing suitable forms of social recognition via awards.

In this article, we focus on the role that self-stereotyping plays for the gender difference in speaking up in math, and we explore whether the effect of recognition on the gender gap operates through the mechanism of self-stereotyping.

This research contributes to both the literature on gender differences and related policy interventions (Bohnet 2016), as well as the research on awards as incentives (Frey and Gallus 2017). Gender research shows persistent differences in educational and occupational representation and participation, notably in math-intensive STEM fields (e.g., Kahn and Ginther 2017, Carter et al. 2018). Meanwhile, policy makers and practitioners often use awards to address gender imbalances (e.g., the EU Prize for Women Innovators). Although awards are widely-used, their causal effects are still poorly understood. While some studies find positive effects of awards (e.g., Kosfeld and Neckermann 2011, Ashraf et al. 2014, Bradler et al. 2016, Gallus 2017, Gallus, Jung and Lakhani 2019), others document unintended consequences, including for the very behavior they were designed to reward and improve (Borjas and Doran 2015, Malmendier and Tate 2009, Gubler, Larkin and Pierce 2016, Robinson et al. 2019). Our line of work contributes to an effort to establish a generalizable framework of awards (Gallus,

Campbell and Gneezy 2019): it sheds light on the mechanisms through which awards impact behavior by focusing on beliefs and self-stereotyping, and it cleanly studies the effects of different forms of awards by exogenously varying one important dimension along which awards differ: their public visibility (see Gallus and Heikensten 2019). This builds on previous evidence from a large-scale field experiment (Gallus 2017) showing that award recipients' beliefs in their ability to make valuable contributions to a public goods community may be an important mechanism behind strong and long-lasting effects of symbolic awards. In the present line of work, we use a lab experiment that allows us to isolate beliefs and study self-stereotyping.

II. Experimental Design and Econometric Strategy

Our outcome variable is participants' tendency to speak up and have their answer count as the group's answer. We build on the paradigm used in Baldiga-Coffman (2014), where participants are randomly paired in groups of two and where they do not see the identity of their partner, to rule out fear of discrimination. They individually answer multiple-choice questions. For each question, participants indicate where "in line" their answer (loosely called "idea") should stand. To do so, they select an integer between 1 and 4. This measure aims to mirror real-life decisions, such as when and how assertively to speak up or raise one's hand. The person in the group who chooses the lowest number has their answer submitted as the group's answer (we reverse-code the variable in the analysis to facilitate its interpretation as a higher confidence to speak up). If group members choose the same number for a particular question, one answer is randomly chosen to be the group's answer. Incentives are perfectly aligned as payoffs are tied to group performance. Each group member has an incentive to rank their answer according to their true belief in their answer's correctness, providing us with an opportunity to measure participants' confidence in their own ability compared to the ability of a randomly selected person in the room.

Comparing participants' tendency to speak up over domains that are associated with different gender stereotypes allows us to see whether a possible under-contribution by women is rooted in an inherently lower confidence in own ability, or whether it is indeed driven by the stereotype of the domain. Our male-typed domain of interest is STEM, and math in particular, as compared to more gender-neutral verbal tasks.

The experiment consists of three parts, each of which gives participants 10 minutes to answer up to 25 incentivized multiple-choice questions, 19 math and 6 verbal questions. The first part measures individual ability (how many questions participants answer correctly). The second and third parts measure ability as well as confidence to speak up.

Between parts two and three, the recognition treatment is administered: depending on which experimental condition the group is in, the person in the group who had the most correct math answers in part two receives either private recognition, a virtual award, or is lauded in an actual award ceremony. Importantly, the experiment holds the content of the award constant and only varies the publicness of recognition. See Gallus and Heikensten (2019) for further details.

We use a difference-in-differences model predicting confidence to contribute an answer from participants' gender, the gender-domain of the question (male-typed math vs. gender-neutral verbal questions), and the interaction between the two, contrasting the pre- and post-treatment periods (in column 3 of Table 1). We control for a range of ability proxies: average individual performance per domain and part for parts 1, 2 and 3, and question-specific ability (whether or not the answer to the respective question is correct) for parts 2 and 3, where we also measure confidence to speak up. Further controls include session size, share of women in the session, a dummy for whether the participant had attended high school in the U.S., a dummy for being a student at the host university (Harvard University), and race dummies. Standard errors are clustered at the participant level.

III. Results

A. Summary Statistics

The data was gathered at the CLER lab at Harvard Business School between April and December of 2016. There were 26 sessions in total, run with between 8 and 26 participants per session (mean 14.5, median 12, s.d. 6). Recruitment was open to accommodate up to 30 participants. Participants spent about 40-50 minutes at their computer terminals, excluding potential waiting times when participants needed to be seated and when they were getting paid at the end.

In total, 378 subjects participated, 212 women (56%) and 166 men (44%), which is representative for lab populations. 128 participants (34%) were in the Private Recognition treatment, 126 participants (33%) were in the Virtual Award treatment, and 124 participants (33%) were in the Award

Ceremony treatment. In this note, we do not analyze the effects of these different treatments.

The fraction of women per treatment varied between 52% and 60%. Almost 54% of the participants were recognized (this number is slightly higher than 50% since cases where both group members had the same part 2 score afforded both of them recognition). 109 women were recognized, that is 51% of the total number of women and 54% of all treated participants. 94 men were recognized, that is 57% of the total number of men and 46% of all treated participants. Gallus and Heikensten (2019) present summary statistics for all participants, split across the three treatment arms, showing that the sample is balanced across treatments.

B. Evolution of Self-Stereotyping

To analyze self-stereotyping patterns and how they evolve after recognition is provided, we study the extent to which domain (male-typed math tasks vs. gender-neutral verbal tasks) predicts men's versus women's confidence to contribute ideas, contrasting the pre-treatment period with the post-treatment period. The findings suggest that recognition reduces self-stereotyping.

Column 1 of Table 1 considers only the male-typed math tasks. It shows that there is a significant gender gap in participants' tendency to contribute their ideas at baseline, i.e., before the recognition intervention. Women are significantly less likely to speak up, controlling for ability ($p<0.001$). Column 2 shows that no such difference exists on the gender-neutral verbal tasks ($p=0.481$). This suggests that women's lower tendency to speak up is not rooted in an inherently lower confidence to contribute ideas, but that it instead depends on the domain. Column 3 reports results on the pooled math and verbal questions. It shows that the interaction between gender and domain is statistically significant ($p=0.004$). Thus, before the recognition intervention, the domain and its interaction with gender is a statistically significant predictor of participants' confidence to contribute their ideas.

Importantly, column 1 shows that women tend to become significantly more confident to contribute their ideas on math tasks in the post-treatment period ($p=0.026$), i.e., after the recognition intervention, while there is no significant pattern for verbal tasks ($p=0.434$) (column 2). The pooled regression in column 3 shows that the interaction between gender, domain and pre- vs. post-treatment period is statistically significant ($p=0.046$). This weakening of the extent to which domain and

gender predict a participant's confidence to speak up indicates that the recognition treatments may have reduced the extent of self-stereotyping among women. Gallus and Heikensten (2019) provide further analyses on the specific effects of the recognition treatments on both recipients' and non-recipients' confidence to speak up in the different domains.

[Insert Table 1 Here]

IV. Discussion

Our line of research studies belief-based reasons for shortfalls in knowledge transfer in STEM fields, and it evaluates policies designed to address such mis-calibrated beliefs. By shedding light on the mechanism of self-stereotyping, this note adds to Gallus and Heikensten (2019), who have established a significant gender gap in the tendency to speak up in STEM domains, and who show that this gender gap can be eliminated by providing suitable forms of recognition. Here, we study the evolution of self-stereotyping patterns from before the recognition intervention to after awards have been bestowed. The findings indicate that recognition may significantly reduce self-stereotyping among women in STEM domains.

REFERENCES

Ashraf, N., O. Bandiera, and S. Lee (2014). "Awards unbundled: Evidence from a natural field experiment." In: *Journal of Economic Behavior and Organization* 100, pp. 44–63.

Baldiga Coffman, K. (2014). "Evidence on self-stereotyping and the contribution of ideas." In: *Quarterly Journal of Economics* 129, pp. 1625–1660.

Bohnet, I. (2016). "What Works: Gender Equality By Design." In: *Cambridge, MA: Harvard University Press*.

Borjas, George J., and Kirk B. Doran. 2015. "Prizes and Productivity: How Winning the Fields Medal Affects Scientific Output." In: *Journal of Human Resources,* 50(3): 728-758.

Bradler, C., R. Dur, S. Neckermann, and A. Non (2016). "Employee recognition and performance: A field experiment." In: *Management Science* 62, pp. 3085– 3099.

Carter, A.J., A. Croft, D. Lukas, and G.M. Sandstrom (2019). "Women's visibility in academic seminars: Women ask fewer questions than men." In: *PLoS ONE* 14(2): e0212146.

Frey, B. S. and J. Gallus (2017a). "Honours Versus Money: The Economics Of Awards." In: *Oxford: Oxford University Press*.

Gallus, J. (2017). "Fostering public good contributions with symbolic awards: A large-scale natural field experiment at Wikipedia." In: *Management Science* 63, pp. 3999–4015.

Gallus, Jana, Sandy Campbell, and Uri Gneezy. 2019. "Awards: Tangibility, Self-Signaling and Signaling to Others." Unpublished working paper.

Gallus, Jana and Emma Heikensten, 2019. "Shine a light (on the bright): The effect of awards on confidence to speak up in gender-typed knowledge work." Unpublished working paper.

Gallus, Jana, Olivia Jung, and Karim Lakhani. 2019. "Managerial recognition as an incentive for innovation platform engagement: A field experiment and interview study at NASA." Harvard Business School Working Paper No. 20-059

Gubler, T., I. Larkin, and L. Pierce (2016). "Motivational spillovers from awards: Crowding out in a multi-tasking environment." In: *Organization Science* 27, pp. 286–303.

Kahn, S. and D. Ginther (2017). "Women and STEM." In: *NBER Working Paper* 23525.

Kosfeld, M. and S. Neckermann (2011). "Getting more work for nothing? Symbolic awards and worker performance." In: *American Economic Journal: Microeconomics* 3, pp. 86–99.

Malmendier, Ulrike, and Geoffrey Tate. 2009. "Superstar CEOs." In: *Quarterly Journal of Economics,* 124(4): 1593-1638.

Robinson, Carly D., Jana Gallus, Monica G. Lee, and Todd Rogers. 2019. "The Demotivating Effect (and Unintended Message) of Awards." *Organizational Behavior and Human Decision Processes,* in press.

TABLE 1—Regression results

	Math	Verbal	Pooled
Female	-0.179***	-0.0479	0.00955
	(0.0443)	(0.0679)	(0.0705)
Math			-0.446***
			(0.0922)
Female*Math			-0.205**
			(0.0708)
Post-treatment	-0.0659	0.0826	0.0740
	(0.0342)	(0.0444)	(0.0448)
Female*Post-treatment	0.0833*	-0.0451	-0.0482
	(0.0373)	(0.0576)	(0.0593)
Math*Post-treatment			-0.162**
			(0.0512)
Female*Math*Post-treatment			0.134*
			(0.0667)
Qn i correct	0.943***	0.688***	0.885***
	(0.0293)	(0.0342)	(0.0247)
Avg score	0.0582***	0.0731***	0.0700***
	(0.00798)	(0.0179)	(0.00729)
Constant	1.431***	1.857***	1.853***
	(0.161)	(0.204)	(0.160)
Controls	Yes	Yes	Yes
Observations	13766	4332	18098
N	378	378	378

Notes. OLS predicting confidence to speak up for question *i*. Avg score denotes individual *j*'s average score per domain and part (for parts 2 and 3). Controls include Part 1 average score per domain, session size, share of women in session, U.S. high school dummy, student at host university dummy, and race dummies. Standard errors in parentheses.

*p<0.05, **p<0.01, ***p<0.001