Shine a light (on the bright):

The effect of awards on confidence to speak up in gender-typed knowledge work

Jana Gallus
UCLA Anderson School of Management

Emma Heikensten
Stockholm School of Economics

November 2019

Abstract

Collaborative knowledge work may suffer if high-ability individuals do not feel confident to speak up and advance their ideas (e.g., due to self-stereotyping). We test whether recognition through awards increases high-ability group members' confidence to speak up when working on male-typed knowledge tasks. We use a lab experiment to study performance-based recognition with different degrees of publicness: private recognition, semi-public award, ceremony. We thus focus on managerial policies that are widely used in practice but have received limited scholarly attention. First, we show that self-stereotyping affects women's contribution of ideas in mathematics. Second, awards significantly increase recipients' and hence high-ability subjects' confidence to speak up. Third, the awards' visibility does not matter much, except when interacted with gender. The gender gap in confidence to speak up disappears among high-ability participants when awards are celebrated in a ceremony with face-to-face recognition. Losers remain unaffected.

.

Introduction

Managers and educators regularly encounter situations where high-ability employees or students do not feel confident to raise their hand and advance an idea or influence group members (Instone et al. 1983, Thomas-Hunt and Phillips 2004, Babcock and Laschever 2009) – often simply because they do not fit the stereotype attached to the respective field (e.g., Eagly 1987, Heilman 2010). This is of particular concern in the fields of science, technology, engineering and math (STEM), which are economically and societally important, are increasingly being driven by group work (Wuchty et al. 2007), and have been strongly associated with a male gender stereotype (e.g., Cech et al. 2013, Cardador 2017). Recent lab experimental research isolates the possible explanations for such undercontribution patterns observed among women working on male-typed tasks (e.g., history, sports) and shows that the behavior is largely driven by women's self-assessments; i.e., self-stereotyping and the associated lack of confidence (Baldiga Coffman 2014). In this paper, we investigate self-stereotyping in STEM fields, and we test if recognition through awards increases the confidence of high-ability individuals to contribute ideas to their group. We test whether the way in which recognition is provided (privately or with face-to-face contact in front of an audience) matters if our goal is to close the gender gap in speaking up in teams.

Awards are one of the most common nonmonetary rewards employed in firms (Gubler et al. 2016, WorldatWork 2017), yet scholars in management and economics have only recently begun to study their effects and the mechanisms through which they operate (Frey 2007, Frey and Gallus 2017a,b). While most studies have focused on the *ex ante* incentive effects of awards from a competitive perspective, we aim to redirect the focus and contribute to the emerging stream of research looking at collaborative contexts and awards that come to recognize individuals by surprise.[1] Such awards provide an important complement to standard organizational incentive schemes.[2] We distinguish between private feedback

[1] See Gallus and Frey (2016) for a discussion and overview of the literature on *ex ante* versus *ex post* awards.
[2] There are well-developed strands of research investigating, e.g., piece rate schemes (e.g., Gneezy and Rustichini 2000; Lazear 2000), rank-order tournaments (Lazear and Rosen 1981), as well as the problems they may give rise to,

which only the recipient sees, virtual awards (akin to online recognition platforms implemented in firms), and public awards, which provide this recognition in front of an audience.

We use a laboratory experiment to test the effectiveness of different types of performance-based recognition to increase high-ability individuals' confidence to speak up and have their own answer count as the group answer. This outcome measure aims to mirror real-life decisions, such as when and how assertively to speak up or raise one's hand, in contrast to letting others go first. It reflects individuals' confidence in their own ability to advance their group. Three experimental treatments vary the degree of publicness of the recognition. In the Private treatment the recognition is only displayed to the recipients, who are informed that their group partners are not aware of it. This is akin to privately getting positive feedback from management. In the Virtual Award treatment, the award is shown to the recipients while informing them that their group members are also made aware of the award, such as when being recognized in a company newsletter or in an online work context without face-to-face exposure to the audience (as in Gallus 2017). In the Award Ceremony treatment, the award is shown on the screen and everybody is asked to stand up and celebrate the award winners while they come to the front of the room and are publicly lauded. The performance information and language of the treatments are held constant across conditions and only the publicness is varied. This is important as it allows us to disentangle the effect of publicness, comparing private recognition to two forms of awards, which do or do not involve face-to-face contact between recipients and the audience. Our interest lies in studying the effects of these interventions on the gender gap in order to evaluate how to effectively use different forms of recognition in organizations.

Our analysis shows that the recognition intervention has a significant effect on subsequent speak-up patterns. It increases recipients' and, hence, the top performers' confidence to advance their ideas, while

such as multitasking (Holmström and Milgrom 1991) and ratchet effects (Gibbons 1987). See Prendergast (1999) for an overview of the economics literature on incentives in firms.

leaving the non-recipients' confidence to speak up unaffected.[3] The gender gap in speak-up patterns remains significant. The average effectiveness of the awards does not vary depending on the publicness. However, an examination of treatment effect heterogeneity by gender indicates a significant differential effect in the Award Ceremony treatment, where the gender gap is closed. This seems to be driven both by men's lower and women's higher responsiveness to the public ceremony involving face-to-face exposure. This finding is in line with Gerhards and Siemer (2016) who studied the *ex ante* incentive effects of public and private feedback and found that women provided more effort in the public treatment than did men. It is also consistent with Jalava, Schrøter Joensen, and Pellas (2015), who found that girls were more motivated by the prospect of receiving a symbolic reward (certificate) than were boys in a school context.[4] Our exploratory analysis of the mechanisms behind this effect of face-to-face recognition suggests that both perceived legitimacy to lead the team and trust in the award's signal may explain the closing of the gender gap in the Award Ceremony treatment.

Our study makes three main contributions to existing research. First, we examine the effects of recognition through surprise awards, which have so far received only scant scholarly attention but are an important component of organizational reward practices (e.g., Cranston and Keller 2013).[5] In light of the robust literature finding a gender gap in contests due to differences in competitiveness (Gneezy, Niederle,

[3] This result is in line with the literature on motivated beliefs (surveyed in Bénabou and Tirole 2016), which finds evidence for asymmetric updating, whereby subjects systematically under-update to negative signals and are closer to Bayesian updating when receiving positive signals (Möbius et al. 2010, Eil and Rao 2011). The result complements recent studies that found that the positive effects of recognition on performance were driven by the *non*-winners (Bradler et al. 2016; Hoogveld and Zubanov 2017; Neckermann and Yang 2017). The difference in results may be explained by our focus on confidence to speak up rather than performance, by the different nature of the underlying task (mathematics being more predictive of ability and less responsive to short-term effort provision than data-entry jobs), or by the differences in the audience and lower magnitude of peer effects (compared to Hoogveld and Zubanov's field experimental context of university tutorials).

[4] Note that these two studies both focused on *ex ante* incentives and their effects on effort provision, rather than surprise awards and their effects on the confidence to speak up in a non-competitive context.

[5] The field experiments by Bradler et al. (2016), Gallus (2017), Hoogveld and Zubanov (2017) and Robinson et al. (2018) and the lab experiments by Lepper, Greene, and Nisbett (1973) and Neckermann and Yang (2017) are the only experimental studies on surprise awards that we are aware of. They differ fundamentally from our study, as becomes clear from the two remaining points on how this study contributes to prior research. The potential downsides of surprise awards (e.g., managers exploiting the discretion they provide) and the measures award givers take to retain the element of surprise in different organizational conditions remain to be explored by future research.

and Rustichini 2003; Niederle and Vesterlund 2007, Croson and Gneezy 2009), surprise recognition is a promising managerial intervention since it tones down the competitive element. As is the case for an increasing number of organizational settings, our context is one of collaborative group work (Wuchty et al. 2007, Lazear and Shaw 2007). Second, our main outcome of interest is group members' confidence to speak up and advance their ideas (controlling for performance). We are not aware of previous studies that would have explored the effects of awards on confidence, yet it seems to be a prime mechanism through which awards work.[6] We use it as a basis from which to subsequently explore the implications for individual and group performance. Third, we explore whether the degree of public visibility of awards matters, again with an eye to detecting gender-specific differences among recipients and non-recipients. The degree of visibility is one important dimension along which corporate awards differ (Gallus and Frey 2016), and past research suggests that this difference may matter from a gender perspective. Specifically, the distinction between face-to-face versus computer-mediated and written communication has been shown to matter for the gender gaps in negotiation (e.g., Bhappu et al. 1997, Stuhlmacher et al. 2007) and trust (e.g., Eckel and Wilson 2002), suggesting that men and women may react differently to being observed and, vice versa, to observing others. More broadly, by integrating research on awards with the literature on workplace related gender differences, our study complements other solution-oriented streams of the gender literature (Bowles and Babcock 2013, Bohnet 2016, Bohnet, Van Geen, and Bazerman 2016, Bear and Babcock 2017).

Experimental design and treatment interventions

The experiment had four incentivized parts, which were programmed and conducted using z-Tree (Fischbacher 2007). The first part was designed to get an appropriate measure of ability on the tasks used

[6] Research on non-financial rewards in general has so far mostly focused on well-observable outcome measures, such as notably task performance (e.g., Ashraf, Bandiera, and Lee 2014; Chan et al. 2014; Gerhards and Siemer 2016; Gubler, Larkin, and Pierce 2016; Huberman, Loch, and Önçüler 2004; Kosfeld and Neckermann 2011; Lacetera and Macis 2010; Larkin 2011; Markham, Scott, and McKee 2002).

in the experiment. It started with instructions and two example questions, one math and one verbal question. To measure beliefs about absolute and relative ability, participants were asked to state the number of similar questions they expected to answer correctly, as well as what quartile they expected to end up in. This was followed by 10 minutes to answer up to 25 multiple choice questions, 6 verbal and 19 math questions. For each math question, there were four answer alternatives to choose from; the verbal tasks had six answer alternatives. Participants were informed that they were free to allocate their time between questions as they saw fit. The questions were presented in a random order that differed from person to person. Each question was incentivized with a payment of 10 cents per correct answer. While our interest lies in subjects' behavior when confronted with the male-typed math task, the inclusion of the more female-typed verbal tasks allowed us to see whether undercontribution patterns were driven by self-stereotyping (i.e., they only occurred on math questions), or whether they could have been driven by alternative factors, such as a generally lower confidence level among women. Moreover, this gave us the opportunity to explore whether receiving (or not receiving) recognition for performance in one field would have spillover effects on confidence in a non-targeted field. Given the limited contribution of the verbal tasks to the main research question (how recognition impacts confidence to speak up in the focal domain), we only included 6 verbal questions in each part and focused participants' time on the 19 math tasks per part.

After the first 10 minutes of questions had passed the participants were randomly paired into groups of two before the second part started. No communication between subjects was permitted, and nobody could identify the face or identity of their partner. Participants were informed that, from then on, they would cooperate with this other person and earn money as a group by submitting the correct answer for the group. As in Baldiga Coffman (2014), for each question, the participant would select an answer and then choose their willingness to contribute this answer as the group's answer by selecting an integer between 1 and 4 to indicate where in line the answer should stand. For each question, the person in the group who chose the lowest number would have their answer submitted as the group's answer. If the group members

chose the same number for where in line they wanted to be for a particular question, one answer was randomly chosen to be the group answer. In the analysis, we reverse the order of the selected options to make the interpretation more intuitive (e.g., if a participant chose to be first in line this will correspond to the highest “speak up” value, 4).

Since the individual payoffs were based on group performance, each group member had an incentive to rank their answer according to their belief in their answer's correctness, providing us with an opportunity to measure subjects' confidence in their own ability compared to the ability of a randomly selected person in the room. That is, we get a measure of willingness to contribute that can be thought of as a way of estimating relative confidence per question. Using four options for where in line subjects wanted to put their answer rather than, say, two, provided subjects more flexibility in their choice and allowed us to glean more information about their confidence (e.g., choosing a 2 gives room to a team member who is very certain to put his or her idea first, while a 4 minimizes the likelihood that one's answer gets submitted; see Baldiga Coffman 2014).

At the end of the second part, the group member who had answered the most math questions correctly was recognized. Thus, the recognition was independent of whether the answers were used as the group’s answer. Basing the recognition on performance allows us to test whether our treatment can be used to encourage high-ability subjects who would otherwise not have the confidence to speak up (for example in a gender-incongruent field). This structure is both plausible and relevant in real-world settings, where team managers can observe the performance of individual team members as well as their confidence to voice their opinions and ideas, irrespective of the quality of these ideas.

The recognition was provided in three different forms: Private Feedback, Virtual Award, and Award Ceremony. All the conditions used the eponymous "Math Master" title and symbol to address recipients and activate an identity of being a person who excels in mathematics. The text was modeled on the typical

language used to celebrate somebody's accomplishment (see Online Appendix Figure A.1): it congratulated the individual, used emotional language and symbols (a smiley), and referred to the individual's deservingness ("earned") and high performance:
"Congratulations! You have just earned our Math Master Award for having attained the highest math score in your group. Let us take a moment to recognize your performance. Great job! :)"

The publicness was the only difference between treatment arms; i.e., we deliberately held constant the information content and language. This is important because it allows us to compare the effectiveness of private recognition with that of two forms of public awards.

Subjects were informed that the recognition would only be provided once. They were moreover told that there would be no extra payment or certificate. This was done to prevent any anticipatory effects, whereby subjects could have exerted more effort in expectation of future recognition that would result in higher individual payoffs. The information that there would be no certificate was provided to reduce any expectations about possible material consequences that might otherwise arise from getting recognition (e.g., career-related implications from a university-sponsored certificate). We were interested in studying symbolic recognition.

Subjects in the Private Feedback condition were informed that their team partner would not be made aware of the recognition. In line with this, the non-recipients faced a waiting screen that they also faced at other times during the experiment, as they waited for their fellow participants to complete different rounds of the experiment. In the Virtual Award treatment arm, recipients faced the same screen as in the Private Feedback treatment, but one sentence was modified to tell them (truthfully) that their team member would be informed about the award. The non-recipients received a short message asking them to wait while we informed their respective team member that he/she had just earned our *Math Master Award* for having attained the highest math score in their group (see Online Appendix Figure A.2). Lastly, in the

Award Ceremony treatment, a new element of public recognition for the award was introduced compared to the Virtual Award treatment. On the screens, both the award recipients and the non-recipients were informed that there would be an award ceremony (see Online Appendix Figure A.3). All participants were asked to stand up and the award recipients were asked to come to the front of the room so that their performance could be recognized in an award ceremony. In the front of the room, a stage-like space had been prepared in the meantime, with the award printed out in large and for everyone to see. When the award recipients slowly walked across the stage they were applauded by their fellow subjects in the room. Participants were not informed about who was in their group and, thus, whom they had outperformed (or who had outperformed them). This was done to rule out any influence from physical appearance or from the gender composition of the group.

Part 3 of the experiment began immediately after the recognition had been provided. Participants were informed that they would go through the exact same exercise as in part 2. That is, they would continue to collaborate in the same group and indicate how willing they were to contribute each of their own answers as the group's answer. Just as in part two, the group's answers determined the participants' payoff. In the fourth and final part, subjects participated in a (slightly updated) version of the Holt and Laury lottery (2002) to elicit their degree of risk aversion. They also answered questions about their demographic characteristics as well as their thoughts about the experiment, how many questions they thought their partner had gotten correct, how often they thought their partner's answer had been submitted for the group, what they thought the experiment was about, and what strategy they had used when deciding how to prioritize (1–4) their answers.

Randomization

Randomization was done in three steps. Since the Award Ceremony treatment was public and took place in front of everybody else in the room, it could not be run jointly with the Private Feedback or Virtual Award treatments. The Award Ceremony treatment required an entire session, but the two other

treatments could be combined. We therefore first randomized the sessions into either Award Ceremony or Private Feedback and Virtual Award. Due to a lower than expected turnout, we had to add additional sessions and randomized them into Award Ceremony or Virtual Award/Private Feedback. For the sessions in which the Private Feedback and Virtual Award treatments were run, participants were randomized into either the Private Feedback or the Virtual Award treatment as they entered the room. For subjects in all the three treatment arms, randomization was used to determine where they would be seated and what participant in the room they would be paired with (the balance table can be found in the Online Appendix, Table A.1).

Econometric strategy

The experimental structure described above combines the setup in prior work on self-stereotyping (Baldiga Coffman 2014) with treatments that allow us to test the effectiveness of different forms of recognition. In the results section we start with an analysis to identify self-stereotyping (as in Baldiga Coffman 2014). We then test whether the recognition overall (all three treatment arms combined) has an effect on subjects' confidence, encouraging them to speak up (i.e., put their solution further to the front), and whether this effect of recognition differs by gender. We then analyze whether there are differences between the treatments, for men and women combined. In our main analysis of interest, we test whether the treatment effects in the three conditions are heterogeneous across genders. Finally, we study whether non-recipients are differentially affected by the intervention, depending on treatment and gender. In subsequent exploratory tests, we also look at individual and group performance effects of the recognition.

The experiment was designed to be tested with OLS regressions, with the main outcome variable being confidence to speak up on the math questions. Since the recognition is provided between parts 2 and 3, and that is also where the treatment differences occur, we use a difference-in-differences method to estimate the effects of the recognition. Below is the simplest version of the regression model to estimate

the main outcome of interest: (1) whether the recognition intervention encourages recipients to speak up, controlling for ability. We cluster on the subject level and hence get 378 unique clusters. The results section reports exact p-values (Wasserstein and Lazar 2016).

(1) Speakup = Recognition + Post-treatment + Recognition*Post-treatment + Correct + Part 2 Score

The variable *Speakup* refers to the choice of where in line the answer is placed. It is a proxy for the participant's confidence to speak up for each math question. The variable *Recognition* captures whether or not the subject received recognition (=1 if yes, 0 otherwise); *Post-treatment* refers to the third part of the experiment (=1 if after the intervention (part 3), 0 otherwise); *Recognition*Post-treatment* is the interaction term and main variable of interest, which allows us to estimate whether the recognition has an effect on the recipient's subsequent confidence to speak up (=1 if recognized and after the intervention, 0 otherwise). *Correct* is a proxy for whether or not the specific question is correct (=1 if correct, 0 otherwise) and hence reflects the subject's question-specific ability; *Part 2 Score* is the individual score in part 2, which the recognition is based on.

Our second specification tests whether the effect of the recognition is different for male and female subjects. We use a similar setup as in model (1) with 378 clusters and add a *Female* dummy indicating whether the subject is female (0 if male), an interaction between being female and in the Recognition group, i.e. *Female*Recognition*, and *Female*Recognition*Post-treatment* for female recipients in the post-treatment period.

(2) Speakup = Female + Recognition + Female*Recognition + Post-treatment + Recognition*Post-treatment+ Female*Recognition*Post-treatment + Correct + Part 2 Score

We next test for differential effects of the three types of recognition: Private Feedback, Virtual Award, and Award Ceremony. Since the main purpose of this study is to examine if recognition can increase the confidence of high-ability individuals to speak up, and since treatment receipt is non-random within groups, the analyses have to consider treatment recipients and non-recipients separately. We first conduct the analysis on the treatment recipients (203 clusters). It is only the randomized publicness of the recognition that differs between the experimental participants. The Private Feedback treatment serves as the baseline:

(3) Speakup = Virtual Award + Award Ceremony + Post-treatment + Virtual Award*Post-treatment + Award Ceremony*Post-treatment + Correct + Part 2 Score

where *Virtual Award* and *Award Ceremony* refer to the second and third treatments, respectively, and take the value of 1 if the subject was in the respective treatment and 0 otherwise. The variable *Virtual Award*Post-treatment* equals 1 if the subject was part of the Virtual Award treatment (treatment arm 2) and the observation comes from after the intervention (i.e., from part 3 of the experiment) and 0 otherwise. Similarly, *Award Ceremony*Post-treatment* equals 1 if the subject was in the Award Ceremony treatment and the observation comes from after the award bestowal (0 otherwise).

The subsequent and main part of the analysis interacts the different treatments with gender, again using the Private Feedback treatment as the baseline and including only treatment recipients in the regression (203 clusters):

(4) Speakup = Female + Virtual Award + Award Ceremony + Post-treatment + Virtual Award*Post-treatment + Award Ceremony*Post-treatment + Female*Private Feedback*Post-treatment + Female*Virtual Award*Post-treatment + Female*Award Ceremony*Post-treatment + Correct + Part 2 Score

where *Female*Private Feedback*Post-treatment* equals 1 if the subject is female and part of the Private Feedback treatment and the observation comes from after the intervention (i.e., from part 3), *Female*Virtual Award*Post-treatment* equals 1 if the subject is female and part of the Virtual Award treatment and the observation comes from after the award bestowal, and *Female*Award Ceremony*Post-treatment* equals 1 if the subject is female and part of the Award Ceremony treatment and the observation comes from after the award bestowal (0 otherwise).

The analyses of the intervention's effects on non-recipients (175 clusters) use model (3) to test for potential differential responses to the three recognition types and model (4) to interact the different treatments with gender, again taking the Private Feedback treatment as the baseline.

In the exploratory analyses we test individual performance effects of the recognition, the three treatments, and differences by gender. In addition, we look at the effect of the recognition on group performance. The analyses at the individual performance level use similar regression models as the analyses conducted for the main hypotheses. The dependent variable for the individual performance effects is the binary variable *Correct*, indicating whether the question is correctly answered (0 otherwise). For the analyses on group-level performance, the dependent variable is *Groupcorrect*, indicating whether the group submitted the correct answer (0 otherwise), which depends on the group members' speak-up values and, naturally, whether the individual answers are correct. Since the analysis is at the group level we cannot control for the time trend and test the treatment effect simultaneously. The results therefore have to be treated with caution. Model (5) examines if *Groupcorrect* is positively affected by the recognition, indicated by the variable *Post-treatment*. To understand if the group's answers are improved by the recognition, we control for whether the two group members' individual choices are correct (via *Correct* and *Partnercorrect*) and the group's total score in the pre-treatment period (*Group Score Pre-Treatment*). The

regression contains one observation per group and question. Standard errors are clustered at the group level (189 clusters).

(5) Groupcorrect = Post-treatment + Correct + Partnercorrect + Group Score Pre-Treatment

Results

Summary statistics

The data was gathered at the CLER lab at Harvard Business School starting on 14 April 2016. The data collection continued until 30 November 2016, but excluded the months of August and September. There were 26 sessions in total, run with between 8 and 26 subjects per session (mean 14.5, median 12, s.d. 6). Recruitment was open to accommodate up to 30 participants, but due to difficulties in estimating show-ups, the exact number of participants per session varied. Subjects spent about 40-50 minutes in front of their computers, excluding potential waiting times when subjects needed to be seated and when they were 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 Feedback treatment, 126 subjects (33%) in the Virtual Award treatment, and 124 subjects (33%) were in the Award Ceremony treatment. The fraction of women per treatment varied between 52% and 60%. Almost 54% of the participants were recognized (this number is larger 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 subjects. 94 men were recognized, that is 57% of the total number of men and 46% of all treated subjects. Table A.1 in the Online Appendix presents summary statistics for all participants, split across the three treatment arms. It shows that the sample is balanced across treatments.

Male gender stereotype of tasks predicts lower confidence to speak up among women

Before testing the main hypothesis, that recognition encourages high-ability group members to speak up, we analyze whether we do indeed see self-stereotyping in the male-typed field of mathematics (thus replicating Baldiga Coffman, who considered male-typed fields such as sports and games and history). As can be seen in Table 1, women exhibit a generally lower confidence to speak up and contribute their ideas to the group than men, in both parts 2 and 3 of the experiment. A *student's t-test* shows that these differences are significant (part 2: mean diff. = 0.215, std. err. = 0.057, p-value < 0.001; part 3: mean diff. = 0.183, std. err. = 0.059, p-value = 0.002). Thus, we do indeed find that there are gender differences in the overall confidence to contribute ideas to the group.

To test for self-stereotyping, which depends on the gender stereotype of the respective field, we next analyze whether women's confidence to speak up is lower only in the male-typed field of mathematics, or whether it is also lower in verbal tasks, which are not male-typed (the latter would indicate that self-stereotyping is not the mechanism driving the gender gap). When considering math and verbal questions separately, we find that there are indeed systematic gender differences depending on the subject matter (math or verbal) and hence gender-type of the field. There are no significant gender differences in confidence to speak up for the verbal tasks, but there are significant differences on both math parts.

[Insert Table 1 here]

These differences in subjects' confidence to speak up are not reflected in systematic gender differences in actual ability. Table 2 below shows that there are no gender differences in ability on math tasks in part 2, i.e., before the intervention. In part 3, men seem to be performing somewhat better than women. (We control for individual ability in the main analyses.)

[Insert Table 2 here]

The regressions in Tables A.2 and A.3 in the Online Appendix test whether the gender-type of the domain influences participants' decisions to speak up when controlling for the correctness of the individual answer and for overall ability. The ability proxies are individuals' scores in the first and second periods, which are positively and significantly correlated ($p<0.001$). Table A.2 displays the results from the pre-treatment period. It shows that, even when controlling for ability and performance on the task, women are significantly less likely to speak up on the male-typed math tasks (coef.=-0.179, $p<0.001$) – while this is not true when they work on verbal tasks. The pooled regression includes a dummy variable corresponding to the math tasks (0 for verbal tasks) and an interaction term for *Female* and *Math*. The results show that the confidence to speak up is generally lower on the math tasks, including for men (coef.=-0.424, $p<0.001$), and the interaction with *Female* is negative and significant (coef.=-0.206, $p=0.004$). This corresponds to the gender differences in relative performance beliefs, which were elicited in the beginning of the study with an incentivized question about which performance quartile the participants thought they would belong to, for verbal and math tasks separately (see Figure A.4 and Table A.4 in the Online Appendix).

Table A.3 uses the same regression models as Table A.2 but considers the post-treatment period. Being female is still negatively related to speaking up on math tasks (coef.=-0.0939, $p=0.040$), but with a smaller coefficient and lower significance level compared to the pre-treatment period (seen in Table A.2). The pooled regression no longer shows a significant interaction term between *Math* and *Female* (coef.=-0.0697, $p=0.293$). Thus, domain stereotype is no longer a significant predictor of confidence to speak up in the post-treatment period (while it does significantly predict speak up patterns before the intervention, always controlling for ability). In an exploratory test to compare the confirmatory finding on self-stereotyping in the pre-treatment period to behavior in the post-treatment period, we find that the reduction in self-stereotyping across periods is indeed statistically significant (Table A.5, Online Appendix). This is an indication that the recognition may have reduced self-stereotyping and increased high-ability women's confidence to speak up in the male-typed domain, which we will look at next.

Symbolic recognition makes recipients more confident to speak up

Figure 1 shows graphical results for the first main hypothesis, using the raw data (i.e., not adjusting for control variables). The y-axis displays the difference between confidence to speak up in the post- and pre-treatment periods, for both the non-recipients and recipients of the recognition. The figure illustrates that both groups have a higher confidence to speak up in the post-treatment period, but that the difference is larger for the treatment recipients. This suggests that the recognition increases its recipients' confidence to speak up.

[Insert Figure 1 here]

Table 3 tests this main hypothesis including control variables. As explained in the section on the econometric strategy, the main regression (column 1) consistently includes a dummy for whether the participant's answer to the question is correct or not, as well as the individual score on part 2 that, in comparison to the partner's score, determined whether the individual received the recognition or not. To that we add two regressions for robustness. The first one (column 2) includes control variables that correspond to the regressions in Baldiga Coffman (2014): score in the first part, size of the session, share of women in the session, race dummies, a dummy for whether the participant had attended high school in the US, and a dummy for being a student at the host university (Harvard University in our case). The regression in column 3 includes controls for risk preferences elicited from the lottery and beliefs about own ability on math tasks. The same approach is used for our main tests specified in models (2-4).

Table 3 displays several interesting findings. It shows that treatment recipients do not differ from non-recipients in their confidence to speak up before the recognition (part 2), controlling for ability. Also, whether or not the question answered comes from the pre- or post-treatment period does not make a difference, suggesting that the number of times a participant has answered questions does not influence confidence in own answers. From just comparing the means in Table 2 above, it is clear that confidence to speak up is higher in part 3 (*Post-treatment*), but this seems to be driven by the scores (which are

generally higher in part 3), by whether the answer to a given question is correct or not, and by the response to the recognition treatment among the recipients.

[Insert Table 3 here]

The main coefficient of interest, *Recognition*Post-treatment*, shows that the recognition overall has a significant positive impact on its recipients' subsequent confidence to speak up ($p<0.001$), controlling for individual performance and for whether each question is correctly answered or not. This implies that the effect stems from recipients' enhanced confidence in their own ability rather than actual ability, knowledge of mathematics, or luck in getting the answers right. Both groups do not differ in general in terms of their confidence to speak up before the intervention. However, after the recognition is given, there is a significant increase in recipients' confidence to speak up compared to the non-recipients, controlling for ability. Both the table and figure thus show that the recognition has a significant and positive effect on recipients' confidence to speak up in the focal domain. Furthermore, as expected, it does not produce spillover effects on confidence to speak up in the non-focal domain of verbal tasks (see Table A.6, Online Appendix).

Table 4 below tests if women react differently to the recognition (all three treatments combined) than men. We use model (2) specified in the econometric strategy section and find that the coefficient for female treatment recipients in the post treatment period, *Female*Recognition*Post-treatment,* is positive but statistically only marginally significant (coef.=0.0616, p=0.076). We conclude that the recognition overall has no or only a weak differential effect on women when controlling for ability and correctness of the answers. This finding is consistent with Jalava, Schrøter Joensen, and Pellas (2015), who found that girls were more motivated by the prospect of receiving a symbolic certificate than were boys in a school context.

[Insert Table 4 here]

Does the recognition's publicness matter?

Table 5 provides results for the third specification, testing whether the recognition differentially affects recipients depending on the degree of public visibility. As discussed above, it includes only the treatment recipients since the analysis and main research question focus on high-ability participants. The analysis reveals no significant treatment differences for women and men combined. Hence, the data do not provide evidence that the publicness of the recognition matters in this context (p=0.261 for the *Virtual Award*Post-treatment* coefficient; p=0.369 for *Award Ceremony*Post-treatment*). Again, we see that the introduction of the recognition provokes an increase in recipients' confidence to speak up.

[Insert Table 5 here]

Since the provision of recognition occurs at the same time as the change in periods (*Post-treatment* variable from specification 1), the size of the coefficient on *Post-treatment* in Table 5 corresponds closely in size to the sum of the *Recognition*Post-treatment* and *Post-treatment* coefficients in Table 3 (0.128 + 0.0439 = 0.1719, compared to 0.195). As expected, since Table 5 only considers treatment recipients, the coefficient on *Post-treatment* becomes significant because it picks up the effect of the recognition, which happened between parts 2 and 3.

Are female recipients differentially affected by the recognition's publicness?

Table 6 presents results from the fourth and main specification of interest, testing whether men and women are differentially affected by the treatments and analyzing the implications for the gender gap in confidence to speak up. The results show that female recipients react more strongly to the Award Ceremony treatment than do male recipients. This result is significant (p=0.027) and holds, with similar coefficient sizes and significance levels, for the alternative regression models using different controls, including ability and correctness. The effect size of the gender-specific reaction to the recognition in the Award Ceremony treatment is of similar magnitude as the *Female* dummy, indicating that it counters part of the gender gap in confidence to speak up. Figure 2 supports this conclusion by showing that among

treatment recipients, the gender difference in confidence to speak up almost disappears in the Award Ceremony treatment, while it remains present in the two other treatments. This finding is in line with Gerhards and Siemer (2016), who studied the *ex ante* incentive effects on effort provision of public and private feedback, finding that women provided more effort in the public treatment than did men.

[Insert Table 6 here]

[Insert Figure 2 here]

Figure A.5 (Online Appendix) shows the change in average confidence to speak up for the different treatment and gender groups between the pre- and post-treatment periods. The figure shows that the difference is the greatest for female recipients in the Award Ceremony treatment. This again suggests that women respond the most to the public ceremony compared to the two other recognition types and compared to men. It also shows visually that women react more strongly to recognition overall, as suggested by the results in Table 4.

Are female non-recipients differentially affected by the different recognition types?

Finally, we consider subjects who did not receive recognition and analyze whether the treatment changed their confidence to speak up. The analysis compares non-recipients in the Private Feedback treatment, who are unaware of the treatment, to non-recipients in the Virtual Award and Award Ceremony treatments. Treatment recipients are excluded from this analysis. Table 7 below shows that there are no significant differences between treatments, neither before (as expected) nor after the intervention. Which period subjects are in does not seem to matter, either. In Table 8 we see, again, that women are less confident to speak up in general, but they are not more or less affected by not receiving the recognition compared to men, leaving the gender gap largely unaffected.

[Insert Table 7 here]

[Insert Table 8 here]

Exploratory analyses

Performance effects

Table A.7 in the Online Appendix tests the effects of recognition, the different forms of recognition, and interactions with gender on individual performance on math tasks. We had chosen math tasks where performance would not respond much to effort but where instead the recognition on past performance would be predictive of individuals' future performance so as to more cleanly study the treatment effects on confidence to speak up (holding constant performance). Thus, we did not expect to find effects on performance, but we include the results for completeness. The presentation follows the same structure as the main analyses, but with the score per question (the dummy *Correct*) as the dependent variable instead of participants' confidence to speak up. As was visible in Table 2 above, the analyses show that participants overall perform better in the last part of the experiment (*Post-treatment*), possibly due to random differences in the difficulty of questions or learning. Treatment recipients perform better on average than non-recipients. The results from the full-sample analysis seem to suggest at first glance that the treatment negatively affects performance (*Recognition*Post-treatment*). However, this result is likely due to mean reversion. Given that treatment recipients are the high performers in the pre-treatment period, the potential for improvement is larger among the non-recipients.

Table A.7 moreover shows that the three different recognition types do not seem to have any significant differential effects on performance. The fifth column, including only the non-recipients, suggests that there is a significant negative effect among female non-recipients in the Award Ceremony treatment. This is an interesting exploratory finding. Future research could fruitfully study such performance effects of public awards on non-recipients from a gender perspective in order to allow us to draw firm conclusions.

Shifting from the individual to the group level, we do not find a significant performance effect as a result of the recognition, controlling for the correctness of the individually submitted answers. Table A.8

displays the results of the OLS regressions corresponding to model (5) in the econometric strategy section. We test the effect of the recognition on a dummy variable for whether the group's *submitted* answer (which depends on the group members' speak-up values) is correct. The analysis is on the group-question level, clusters on the group level and controls for the correctness of individual group members' answers and the group's performance in the previous part. The analysis does not allow us to rule out the null of no effect on group performance (the *Post-treatment* coefficient is not statistically significant when we include the various controls). As an addition to this analysis, Figure A.6 in the Online Appendix displays the mean number of correctly answered questions, both by individuals and by groups, respectively, before and after the treatment. Both individual answers and group answers improve between the pre- and post-treatment periods, but the difference between the individuals and the groups is stable at 10 percentage points in both periods. Figure A.7 (Online Appendix) also shows that the group scores are correlated across periods, which is expected.

Mechanisms behind the effects of the public award ceremony

Given its practical relevance and theoretical interest, we further explore what may account for the closing of the gender gap in the Award Ceremony treatment where face-to-face exposure is involved. Two non-exclusive mechanisms seem particularly relevant. First, we would expect women and men to respond differently to an intervention that increases their legitimacy as a group leader in a gendered domain (math). Thus, *being seen* by others and lauded publicly as the person who is the most competent in an atypical domain may increase women's sense of legitimacy, while men should not face a role conflict (e.g., Eagly 1987, Luhaorg and Zivian 1995). Second, *seeing* the non-recipients standing in the audience may have a differential effect on recipients' propensity to trust the award's signal about their relative competence. Previous research indicates that women's trust is more sensitive than men's to the details of the context (Azmat and Petrongolo 2014). For instance, seeing the partner's face has been found to make women more likely to trust compared to men and compared to merely being given written information (Eckel and Wilson 2002). Hence, women in our experiment may be more likely to believe the award's

signal about their own relative performance when seeing the audience of non-recipients standing in the room. We explore this possible explanation with the data we collected at the end of the experiment, capturing subjects' beliefs about their own relative performance compared to their team mate (Figure A.8 in the Online Appendix). While only indicative, the direction of the results suggests that women may be more likely to believe that they outperformed their partner in the Award Ceremony treatment compared to the Private Feedback and Virtual Award conditions. For men, the Award Ceremony treatment does not add to the signal's credibility.

Discussion

This study explores the effectiveness of recognition-based managerial interventions to reduce the gender gap in subjects' confidence to advance their ideas to their group in gender-typed knowledge work. In a setup where pure performance information provision was found to be ineffective (Baldiga Coffman 2014), it asks whether recognition with different degrees of publicness increases – and may have differential effects on – the confidence of high-ability individuals to speak up and contribute their ideas to their groups. We first extend previous work on gender-based self-stereotyping to the context of math, a field that is a significant contributor to the gender gap in STEM (see Kahn and Ginther 2017 for a recent review). We find that even high-ability women in this male-typed domain undercontribute their ideas to their group, while there is no undercontribution on a more gender neutral task, as self-stereotyping theory would predict. This is in line with recent empirical evidence on female elite mathematicians by Pope (2017), whose conclusion emphasizes the importance of turning our focus to policy implications: "Efforts intended to improve the accuracy of beliefs about mathematical ability are likely to be an important strategy in combating gender imbalance in STEM fields" (Aug 8). Providing recognition through awards is one such possible effort, and awards are indeed of widespread use in organizations in the private, public and non-profit sectors (Frey and Gallus 2017b). Given the limited empirical research on the topic, we do not yet have evidence on the mechanisms through which they impact behavior, and the possible

differential implications of the various forms of awards for different groups of the labor force. This is what our paper addresses.

We provide experimental evidence that symbolic recognition can increase the confidence of high-performers, encouraging them to speak up and advance their ideas in a male-typed domain. While the degree of publicness of the recognition did not seem to matter on average, the gender difference in confidence to contribute ideas disappeared in the Award Ceremony treatment, which involved face-to-face exposure. We discuss perceived legitimacy and trust in the award's signal as plausible mechanisms behind this effect.

The study raises a number of questions for follow-up research. For instance, further examining why women and men seem to react differently to face-to-face recognition as opposed to private or computer-mediated recognition seems both promising and relevant. It would be highly interesting to get a better sense of how emotions that are triggered by recognition influence belief formation and trust in the award's signal. Similarly, studying the effect of recognition on confidence and individuals' sense of identity, and looking at different age levels, would help decision-makers understand how to best design and target the intervention. If recognition can influence people's sense of identity, the effect on behavior would be expected to be relatively persistent (more persistent than, say, that of money). Hence, another intriguing question is how long the effects last, depending on the timing and form of recognition. For instance, will seeing a *Genius* plaque on one's wall lead to habituation, or can it subtly shape one's sense of identity through repeated exposure (and others' reactions to seeing the plaque)? Answering these questions will naturally also involve studying how different recognition and feedback schemes complement each other, and what is the optimal amount and frequency with which awards can be bestowed in organizations before they lose their value.

We also stand to gain important insights from replicating the experiment in the field, for instance in the context of collaborative innovation projects. Many if not most modern knowledge work platforms allow subjects to operate via pseudonyms and hence keep their gender undisclosed. While this is close to the experimental design, another extension would be to introduce gender visibility. It would be interesting to see whether this influences the recipients' and audience's reaction to awards, depending on gender. There is a potential that awards can contribute to shaping societal stereotypes. Understanding how this can be achieved and what conditions are important (e.g., with respect to the nature of the task and performance ambiguity, the involvement of a jury of experts, superiors, or peers) would allow us to design managerial policies with positive external effects, mitigating self-stereotyping by addressing the stereotypes as such rather than the individual's assessment of how (dis-)similar he or she is to a given group.

References

Ashraf, Nava, Oriana Bandiera, and Scott Lee, "Awards unbundled: Evidence from a natural field experiment," Journal of Economic Behavior and Organization, 100 (2014), 44-63.

Azmat, Ghazala and Barbara Petrongolo, "Gender and the labor market: What have we learned from field and lab experiments?" Labour Economics, 30 (2014), 32-40.

Babcock, Linda, and Sara Laschever. *Women Don't Ask: Negotiation And The Gender Divide* (Princeton and Oxford: Princeton University Press, 2009).

Baldiga Coffman, Katherine, "Evidence on self-stereotyping and the contribution of ideas," Quarterly Journal of Economics, 129 (2014), 1625-1660.

Bear, Julia B. and Linda Babcock, "Negotiating femininity: Gender-relevant primes improve women's economic performance in gender role incongruent negotiations," Psychology of Women Quarterly, 41 (2017), 163-174.

Bénabou, Roland J. and Jean Tirole, "Mindful economics: The production, consumption, and value of beliefs," Journal of Economic Perspectives, 30 (2016), 141-164.

Bhappu, Anita D., Terri L. Griffith, and Gregory B. Northcraft, "Media effects and communication bias in diverse groups," Organizational Behavior and Human Decision Processes, 70 (1997), 199-205.

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

Bohnet, Iris, Alexandra Vivien Van Geen, and Max H. Bazerman, "When performance trumps gender bias: Joint versus separate evaluation," Management Science, 62 (2016), 1225-1234.

Bowles, Hannah Riley, and Linda Babcock, "How can women escape the compensation negotiation dilemma? Relational accounts are one answer," Psychology of Women Quarterly, 37 (2013), 80-96.

Bradler, Christiane, Robert Dur, Susanne Neckermann, and Arjan Non, "Employee recognition and performance: A field experiment," Management Science, 62 (2016), 3085-3099.

Cardador, M. Teresa, "Promoted up but also out? The unintended consequences of increasing women's representation in managerial roles in engineering," Organization Science, 28 (2017), 597-617.

Cech, Erin, Brian Rubineau, Susan Silbey, and Caroll Seron, "Professional role confidence and gendered persistence in engineering," American Sociological Review, 76 (2013), 641-666.

Chan, Ho Fai, Bruno S. Frey, Jana Gallus, and Benno Torgler, "Academic honors and performance," Labour Economics, 31 (2014), 188-204.

Cranston, Susie, and Scott Keller, "Increasing the 'meaning quotient' of work," *McKinsey Quarterly*, January (2013).

Croson, Rachel, and Uri Gneezy, "Gender differences in preferences," Journal of Economic Literature, 47 (2009), 448-474.

Eagly, Alice, *Sex Differences in Social Behavior: A Social-Role Interpretation* (Hillsdale, NJ: Erlbaum, 1987).

Eckel, Catherine C. and Rick K. Wilson, "Conditional trust: sex, race and facial expressions in a trust game," Unpublished manuscript (2002).

Eil, David, and Justin M. Rao, "The good news-bad news effect: Asymmetric processing of objective information about yourself," American Economic Journal: Microeconomics, 3 (2011), 114-138.

Fischbacher, Urs, "z-Tree: Zurich Toolbox for Ready-made Economic Experiments," Experimental Economics, 10 (2007), 171-178.

Frey, Bruno S., "Awards as compensation," European Management Review, 4 (2007), 6-14.

Frey, Bruno S., and Jana Gallus, "Towards an economics of awards," Journal of Economic Surveys, 31 (2017a), 190-200.

——, *Honours Versus Money: The Economics Of Awards* (Oxford: Oxford University Press, 2017b).

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

Gallus, Jana, and Bruno S. Frey, "Awards: A strategic management perspective," Strategic Management Journal, 37 (2016), 1699-1714.

Gerhards, Leonie, and Neele Siemer, "The Impact of Private and Public Feedback on Worker Performance - Evidence from the Lab," Economic Inquiry, 54 (2016), 1188-1201.

Gibbons, Robert, "Piece-rate incentive schemes," Journal of Labor Economics, (1987), 413-429.

Gneezy, Uri, Muriel Niederle, and Aldo Rustichini, "Performance in competitive environments: Gender differences," Quarterly Journal of Economics, 118 (2003), 1049-1074.

Gneezy, Uri, and Aldo Rustichini, "Pay enough or don't pay at all," Quarterly Journal of Economics, 115 (2000), 791-810.

Gubler, Timothy, Ian Larkin, and Lamar Pierce, "Motivational spillovers from awards: Crowding out in a multi-tasking environment," Organization Science, 27 (2016), 286-303.

Heilman, Madeline, "Description and prescription: How gender stereotypes prevent women's ascent up the organizational ladder," Journal of Social Issues, 57 (2010), 657-674.

Holmström, Bengt, and Paul Milgrom, "Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design," Journal of Law, Economics, and Organization, 7 (1991), 24-52.

Holt, Charles A., and Susan K. Laury, "Risk aversion and incentive effects," American Economic Review, 92 (2002), 1644-1655.

Hoogveld, Nicky, and Nikolay Zubanov, "The power of (no) recognition: Experimental evidence from the university classroom," Journal of Behavioral and Experimental Economics, 67 (2017), 75-84.

Huberman, Bernardo A., Christoph H. Loch, and Ayse Önçüler, "Status as a valued resource," Social Psychology Quarterly, 67 (2004), 103-114.

Instone, Debra, Brenda Major, and Barbara B. Bunker, "Gender, self confidence, and social influence strategies: An organizational simulation," Journal of Personality and Social Psychology, 44 (1983), 322-333.

Jalava, Nina, Juanna Schrøter Joensen, and Elin Pellas, "Grades and rank: Impacts of non-financial incentives on test performance," Journal of Economic Behavior & Organization, 115 (2015), 161-196.

Kahn, Shulamit, and Donna Ginther, "Women and STEM," NBER Working Paper No. 23525, 2017.

Kosfeld, Michael, and Susanne Neckermann, "Getting more work for nothing? Symbolic awards and worker performance," American Economic Journal: Microeconomics, 3 (2011), 86-99.

Lacetera, Nicola, and Mario Macis, "Social image concerns and prosocial behavior: Field evidence from a nonlinear incentive scheme," Journal of Economic Behavior and Organization, 76 (2010), 225-237.

Larkin, Ian, "Paying $30,000 for a Gold Star: An empirical investigation into the value of peer recognition to software salespeople," Unpublished working paper, 2011.

Lazear, Edward P., "Performance pay and productivity," American Economic Review, 90 (2000), 1346-1361.

Lazear, Edward P., and Sherwin Rosen, "Rank-order tournaments as optimum labor contracts," Journal of Political Economy, 89 (1981), 841-864.

Lazear Edward P., and Kathryn L. Shaw, "Personnel Economics: The Economist's View of Human Resources," Journal of Economic Perspectives, 21 (2007), 91-114.

Lepper, Mark R., David Greene, and Richard E Nisbett, "Undermining children's intrinsic interest with extrinsic rewards: A test of the over-justification hypothesis," Journal of Personality and Social Psychology, 28 (1973), 129-137.

Markham, Steven E., K. Dow Scott, and Gail H. McKee, "Recognizing good attendance: A longitudinal quasi-experimental field study," Personnel Psychology, 55 (2002), 639-660.

Möbius, Markus M., Muriel Niederle, Paul Niehaus, and Tanya S. Rosenblat, "Managing self-confidence," NBER Working Paper No. 17014, 2011.

Neckermann, Susanne, and Xiaolan Yang, "Understanding the (unexpected) consequences of unexpected recognition," Journal of Economic Behavior and Organization, 135 (2017), 131-142.

Niederle, Muriel, and Lise Vesterlund, "Do women shy away from competition? Do men compete too much?," Quarterly Journal of Economics, 122 (2007), 1067-1101.

Nosek, Brian A., Mahzarin R. Banaji, and Anthony G. Greenwald, "Math= male, me= female, therefore math≠ me," Journal of Personality and Social Psychology, 83 (2002), 44-59.

Pope, Devin, "Women who are elite mathematicians are less likely than men to believe they're elite mathematicians," Washington Post, August 8 (2017).

Prendergast, Canice, "The provision of incentives in firms," Journal of Economic Literature, 37 (1999), 7-63.

Robinson, Carly, Jana Gallus, Monica Lee, and Todd Rogers, "The demotivating effect (and unintended message) of retrospective awards," HKS Working Paper No. RWP18-020 (2018).

Stuhlmacher, Alive F., Maryalice Citera, and Toni Willis, "Gender differences in virtual negotiation: Theory and research," Sex Roles, 57 (2007), 329-339.

Thomas-Hunt, Melissa C., and Katherine W Phillips, "When what you know is not enough: Expertise and gender dynamics in task groups," Personality and Social Psychology Bulletin, 30 (2004), 1585-1598.

WorldatWork, *Trends in Employee Recognition* (Scottsdale, 2017).

Wuchty, Stefan, Benjamin F. Jones, and Brian Uzzi, "The increasing dominance of teams in production of knowledge," Science, 316 (2007), 1036-1039.

TABLES

Table 1: Results on self-stereotyping

	Part 2 Math	Part 2 Verbal	Part 2 Pooled	Part 3 Math	Part 3 Verbal	Part 3 Pooled
Men	3.290	2.819	3.179	3.464	2.953	3.343
	(0.638)	(0.805)	(0.545)	(0.618)	(0.763)	(0.523)
Women	3.008	2.828	2.963	3.233	2.930	3.160
	(0.638)	(0.732)	(0.554)	(0.701)	(0.733)	(0.611)
P-value	0.00002	0.916	0.0002	0.001	0.769	0.002
Observations	378	378	378	378	378	378

Notes: Higher numbers indicate greater confidence to contribute one's idea ("speak up"). Columns show individual average *Speak up* values on math tasks, verbal tasks, and the pooled data, split into the two parts of the experiment where the *Speak up* variable was measured, i.e., parts 2 and 3. Rows indicate individual average scores among men and women, respectively, as well as standard deviations (in parentheses). *p*-values computed using *t-tests* for two independent samples.

Table 2: Average number of correct answers per part and gender

	Part 2 Math	Part 2 Verbal	Part 2 Pooled	Part 3 Math	Part 3 Verbal	Part 3 Pooled
Men	12.355	2.494	14.849	14.867	2.873	17.741
	(3.871)	(1.552)	(4.349)	(3.728)	(1.652)	(4.387)
Women	11.745	2.585	14.330	13.920	2.967	16.887
	(3.359)	(1.379)	(3.872)	(3.565)	(1.537)	(4.152)
P-value	0.102	0.548	0.221	0.012	0.571	0.054
Observations	378	378	378	378	378	378

Notes: Average number of correct answers out of a total of 19 math and 6 verbal questions per part. Standard deviations in parentheses. *p*-values computed using *t-tests* for two independent samples.

Table 3: Main effect of recognition on confidence to speak up

	(Model 1a)	(Model 1b)	(Model 1c)
Recognition	0.0469	0.0274	0.0427
	(0.0534)	(0.0502)	(0.0517)
Post-treatment	0.0439	0.0442	0.0432
	(0.0295)	(0.0294)	(0.0294)
Recognition*Post-treatment	0.128***	0.127***	0.129***
	(0.0341)	(0.0340)	(0.0340)
Correct	1.003***	0.984***	0.991***
	(0.0293)	(0.0292)	(0.0293)
Part 2 Score	0.0725***	0.0464***	0.0592***
	(0.00796)	(0.00990)	(0.00825)
Beliefs			0.0325***
			(0.00650)
Risk			-0.0396
			(0.0930)
Constant	1.547***	1.545***	1.297***
	(0.100)	(0.163)	(0.114)
Controls	No	Yes	No
Clusters	378	378	378
Observations	13766	13766	13766

Standard errors in parentheses

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

Notes: Diff-in-diff regression over parts 2 and 3 of the experiment predicting average confidence to speak up per question. Includes both treatment recipients and non-recipients.

Table 4: Testing whether the recognition differentially affects women's confidence to speak up

	(Model 2a)	(Model 2b)	(Model 2c)
Female	-0.246**	-0.231**	-0.178*
	(0.0796)	(0.0793)	(0.0751)
Recognition	-0.0155	-0.0434	0.000897
	(0.0717)	(0.0685)	(0.0682)
Female*Recognition	0.101	0.121	0.0669
	(0.0922)	(0.0890)	(0.0884)
Post-treatment	0.0440	0.0441	0.0434
	(0.0295)	(0.0294)	(0.0294)
Recognition*Post-treatment	0.0957*	0.0928*	0.0952*
	(0.0396)	(0.0395)	(0.0395)
Female*Recognition*Post-treatment	0.0616	0.0643	0.0632
	(0.0347)	(0.0347)	(0.0346)
Correct	0.998***	0.981***	0.989***
	(0.0294)	(0.0293)	(0.0294)
Part 2 Score	0.0714***	0.0469***	0.0601***
	(0.00795)	(0.00968)	(0.00828)
Beliefs			0.0283***
			(0.00645)
Risk			-0.0478
			(0.0911)
Constant	1.707***	1.561***	1.450***
	(0.111)	(0.158)	(0.125)
Controls	No	Yes	No
Clusters	378	378	378
Observations	13766	13766	13766

Standard errors in parentheses

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

Table 5: Testing the effect of recognition with different degrees of publicness on recipients' confidence to speak up

	(Model 3a) Recipients	(Model 3b) Recipients	(Model 3c) Recipients
Virtual Award	0.0347	0.0511	0.0354
	(0.0586)	(0.0529)	(0.0570)
Award Ceremony	0.0351	0.0525	0.0339
	(0.0562)	(0.0540)	(0.0559)
Post-treatment	0.195***	0.194***	0.194***
	(0.0305)	(0.0305)	(0.0305)
Virtual*Post-treatment	-0.0464	-0.0472	-0.0459
	(0.0412)	(0.0411)	(0.0412)
Ceremony*Post-treatment	-0.0395	-0.0390	-0.0394
	(0.0438)	(0.0440)	(0.0438)
Correct	1.065***	1.050***	1.061***
	(0.0440)	(0.0434)	(0.0438)
Part 2 Score	0.0633***	0.0433***	0.0557***
	(0.00790)	(0.00977)	(0.00868)
Beliefs			0.0156**
			(0.00553)
Risk			0.00662
			(0.0949)
Constant	1.654***	1.489***	1.534***
	(0.128)	(0.183)	(0.137)
Controls	No	Yes	No
Clusters	203	203	203
Observations	7463	7463	7463

Standard errors in parentheses

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

Table 6: Testing for gender differences among high-ability participants in the response to receiving different types of recognition

	(Model 4a) Recipients	(Model 4b) Recipients	(Model 4c) Recipients
Female	-0.155**	-0.144**	-0.139**
	(0.0478)	(0.0445)	(0.0467)
Virtual Award	0.0493	0.0667	0.0483
	(0.0568)	(0.0521)	(0.0557)
Award Ceremony	0.0332	0.0503	0.0324
	(0.0550)	(0.0528)	(0.0549)
Post-treatment	0.203***	0.199***	0.196***
	(0.0431)	(0.0443)	(0.0436)
Virtual*Post-treatment	-0.0968	-0.0861	-0.0879
	(0.0595)	(0.0602)	(0.0592)
Ceremony*Post-treatment	-0.118	-0.120	-0.108
	(0.0615)	(0.0621)	(0.0625)
Female*Private*Post-treatment	-0.0182	-0.0114	-0.00558
	(0.0593)	(0.0580)	(0.0592)
Female*Virtual*Post-treatment	0.0707	0.0575	0.0671
	(0.0579)	(0.0556)	(0.0569)
Female*Ceremony*Post-treatment	0.141*	0.152*	0.133*
	(0.0632)	(0.0624)	(0.0617)
Correct	1.064***	1.051***	1.061***
	(0.0443)	(0.0436)	(0.0440)
Part 2 Score	0.0597***	0.0413***	0.0540***
	(0.00779)	(0.00931)	(0.00845)
Beliefs			0.0126*
			(0.00558)
Risk			0.00575
			(0.0942)
Constant	1.784***	1.573***	1.672***
	(0.138)	(0.185)	(0.152)
Controls	No	Yes	No
Clusters	203	203	203
Observations	7463	7463	7463

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7: Testing for treatment differences among non-recipients

	(Model 3a) Non-Recipients	(Model 3b) Non-Recipients	(Model 3c) Non-Recipients
Virtual Award	-0.0689	-0.0620	-0.0529
	(0.104)	(0.0939)	(0.0944)
Award Ceremony	-0.186	-0.155	-0.131
	(0.102)	(0.103)	(0.0952)
Post-treatment	0.0631	0.0620	0.0630
	(0.0472)	(0.0465)	(0.0468)
Virtual*Post-treatment	0.00790	0.00763	0.00697
	(0.0673)	(0.0671)	(0.0671)
Ceremony*Post-treatment	-0.0450	-0.0402	-0.0438
	(0.0727)	(0.0724)	(0.0722)
Correct	0.950***	0.924***	0.928***
	(0.0390)	(0.0397)	(0.0396)
Part 2 Score	0.0807***	0.0484**	0.0656***
	(0.0136)	(0.0161)	(0.0135)
Beliefs			0.0466***
			(0.0112)
Risk			-0.114
			(0.150)
Constant	1.582***	1.683***	1.190***
	(0.178)	(0.278)	(0.200)
Controls	No	Yes	No
Clusters	175	175	175
Observations	6303	6303	6303

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 8: Testing for gender differences among non-recipients

	(Model 4a) Non-Recipients	(Model 4b) Non-Recipients	(Model 4c) Non-Recipients
Female	-0.276***	-0.245**	-0.177*
	(0.0793)	(0.0877)	(0.0763)
Virtual Award	-0.0440	-0.0416	-0.0377
	(0.102)	(0.0933)	(0.0941)
Award Ceremony	-0.149	-0.125	-0.112
	(0.0997)	(0.103)	(0.0936)
Post-treatment	0.0866	0.0838	0.103
	(0.0808)	(0.0792)	(0.0742)
Virtual*Post-treatment	-0.0932	-0.0651	-0.122
	(0.135)	(0.126)	(0.124)
Ceremony*Post-treatment	-0.181	-0.205	-0.203
	(0.148)	(0.147)	(0.139)
Female*Private*Post-treatment	-0.0451	-0.0427	-0.0777
	(0.139)	(0.135)	(0.128)
Female*Virtual*Post-treatment	0.129	0.0840	0.146
	(0.134)	(0.123)	(0.127)
Female*Ceremony*Post-treatment	0.174	0.220	0.183
	(0.144)	(0.138)	(0.142)
Correct	0.945***	0.922***	0.928***
	(0.0386)	(0.0393)	(0.0391)
Part 2 Score	0.0817***	0.0509**	0.0677***
	(0.0136)	(0.0159)	(0.0136)
Beliefs			0.0424***
			(0.0113)
Risk			-0.131
			(0.147)
Constant	1.718***	1.566***	1.323***
	(0.178)	(0.280)	(0.206)
Controls	No	Yes	No
Clusters	175	175	175
Observations	6303	6303	6303

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

FIGURES

Figure 1: Main result for hypothesis 1

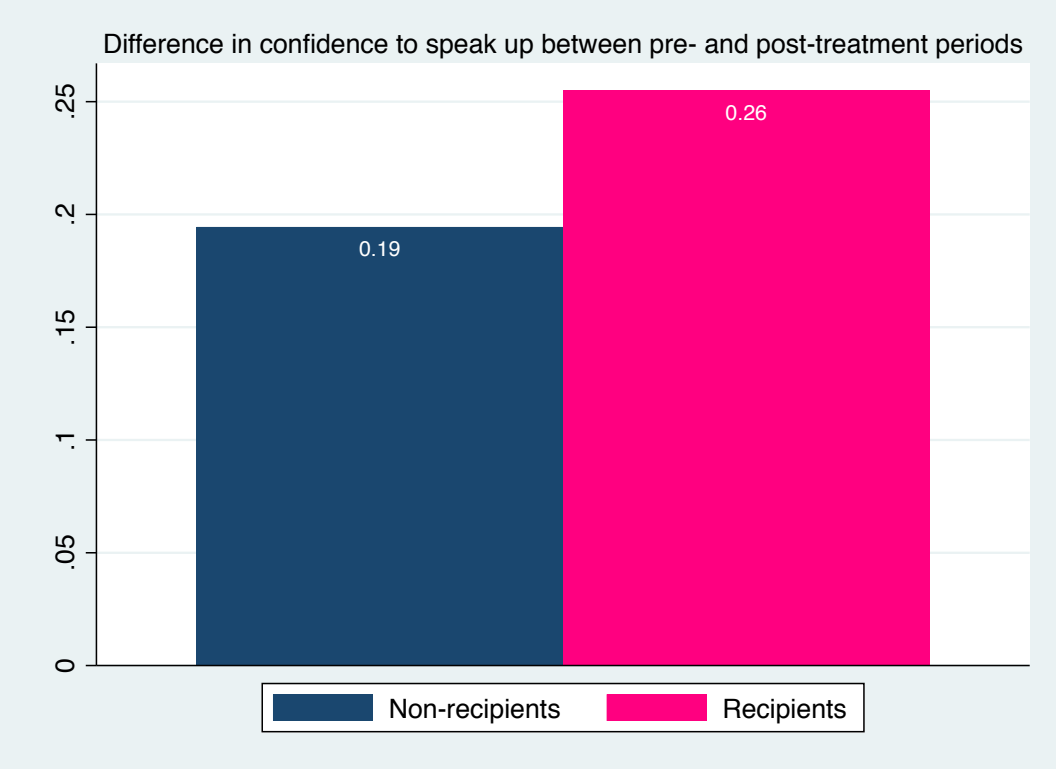
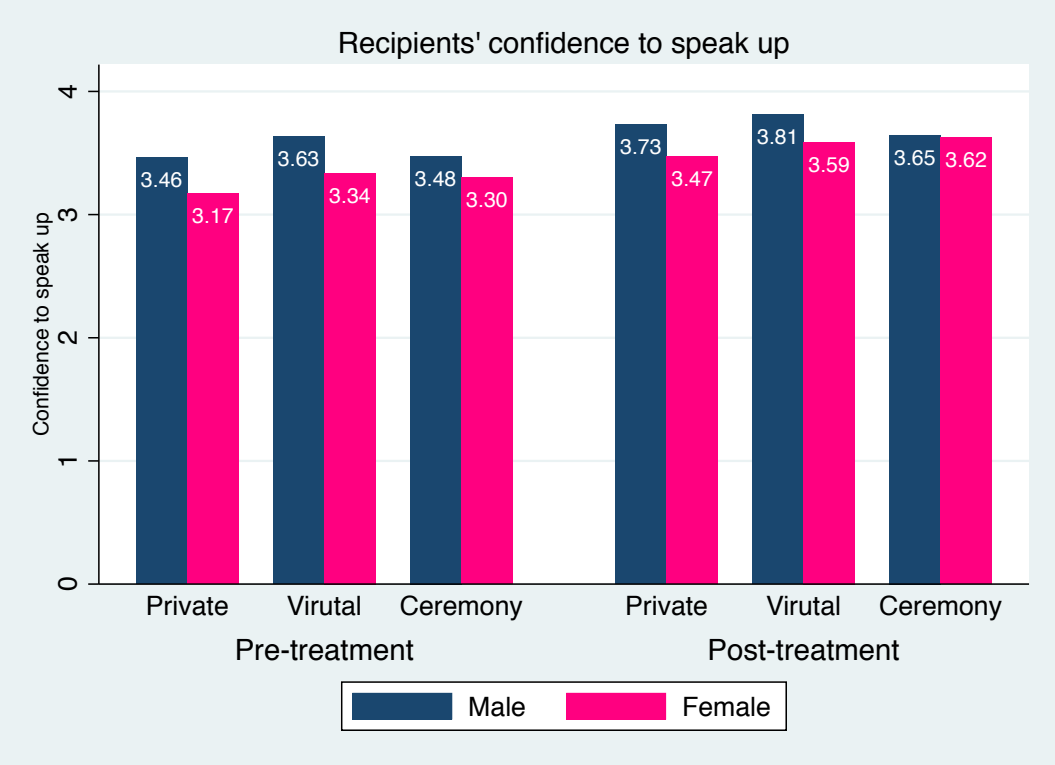


Figure 2: Gender differences in the response to receiving different types of recognition



Online Appendix

Table A.1: Balance across treatments

Treatment	Private Feedback	Virtual Award	Award Ceremony	Pooled
Age	22.41	22.33	22.42	22.39
Gender (=1 if female)	0.516	0.603	0.565	0.561
Degree attained				
High-school	0.414	0.373	0.444	0.410
Bachelor	0.336	0.452	0.355	0.381
Masters	0.188	0.143	0.153	0.161
Other	0.063	0.032	0.048	0.048
Educational field				
Economics	0.172	0.079	0.113	0.122
Political Science	0.078	0.079	0.056	0.071
Mathematics	0.016	0.024	0.073	0.037
Psychology	0.086	0.087	0.073	0.082
Humanities	0.039	0.079	0.073	0.063
Biology	0.102	0.159	0.105	0.122
Arts	0.031	0.056	0.016	0.034
History	0.023	0.016	0.056	0.032
Other Social Sciences	0.102	0.119	0.129	0.116
Other Nature Sciences	0.133	0.127	0.105	0.122

Other	0.219	0.175	0.202	0.198
Current Harvard stud.	0.578	0.579	0.589	0.582
Ethnicity				
African American	0.156	0.095	0.121	0.124
Asian / Pacific Island.	0.266	0.310	0.282	0.286
Hispanic / Latino	0.086	0.087	0.081	0.085
Native American / American Indian	0.000	0.008	0.000	0.003
White	0.375	0.421	0.460	0.418
Other	0.117	0.079	0.056	0.085
Performance				
Part 1, math	12.91	12.94	12.62	12.83
Part 1, verbal	3.09	3.26	3.24	3.2
Part 2, math	11.80	12.40	11.83	12.01
Part 2, verbal	2.43	2.59	2.62	2.54
Confidence to speak up, Part 2	3.05	3.11	3.01	3.06
Beliefs				
# correct math	13.73	13.85	13.25	13.61
4th quartile math	0.297	0.341	0.258	0.299
3rd quartile math	0.469	0.429	0.492	0.463
2nd quartile math	0.211	0.183	0.185	0.193
1st quartile math	0.023	0.048	0.065	0.045

# correct verbal	4.40	4.33	4.24	4.32
4th quartile verbal	0.344	0.365	0.331	0.347
3rd quartile verbal	0.469	0.444	0.468	0.460
2nd quartile verbal	0.172	0.135	0.153	0.153
1st quartile verbal	0.016	0.056	0.048	0.040
Risk taking				
Risk preference lottery	0.460	0.479	0.485	0.474
Risk preference survey	3.36	3.22	3.21	3.26
Observations	128	126	124	378

Notes: The table reports means for the three experimental conditions (Private Feedback, Virtual Award, Award Ceremony) and the pooled sample. *Performance* variables indicate, for part 1 and 2, how many of the 19 math and 6 verbal questions subjects on average answered correctly. *Confidence to speak up* is the main outcome variable of interest and takes values 1 to 4. The variables in the category on *Beliefs* were elicited in part 1 of the experiment. They indicate how many of the 19 math and 6 verbal questions subjects on average expected to answer correctly. They also capture, for each treatment, the fraction of subjects who expected to end up in the different performance quartiles (1–4). *Risk preference lottery* (à la Holt and Laury) is coded to take values 0 to 1 (a higher number indicates more risk-taking). *Risk preference survey* takes values 1 to 5 and captures the subject's self-reported willingness to take risk.

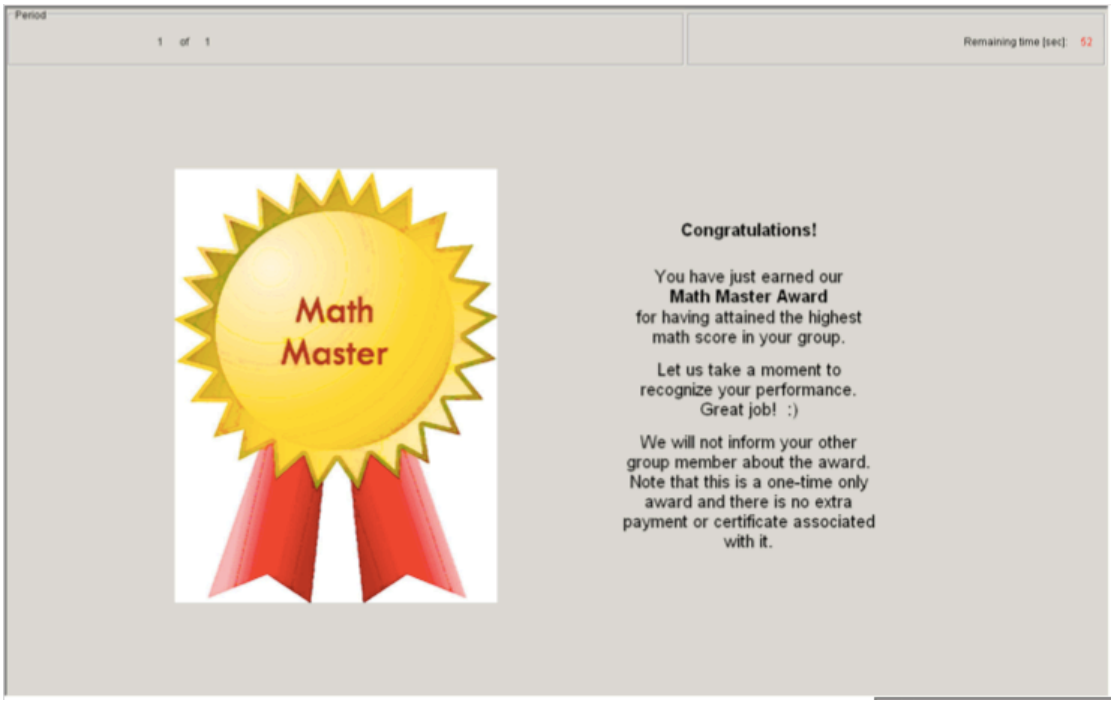


Figure A.1. Screen view for treatment recipients in the Private Feedback arm. Recipients in the other treatment arms saw the same picture and wording, with the exception of the first sentence of the final paragraph, which would be replaced with: "We will inform your other group member about the award" for recipients in the Virtual Award treatment. Recipients in the Award Ceremony treatment would be told that "We will now have a small award ceremony for all the award recipients. Please get ready to stand up and come to the front" – followed by the standard sentences saying that this was a one-time only award.

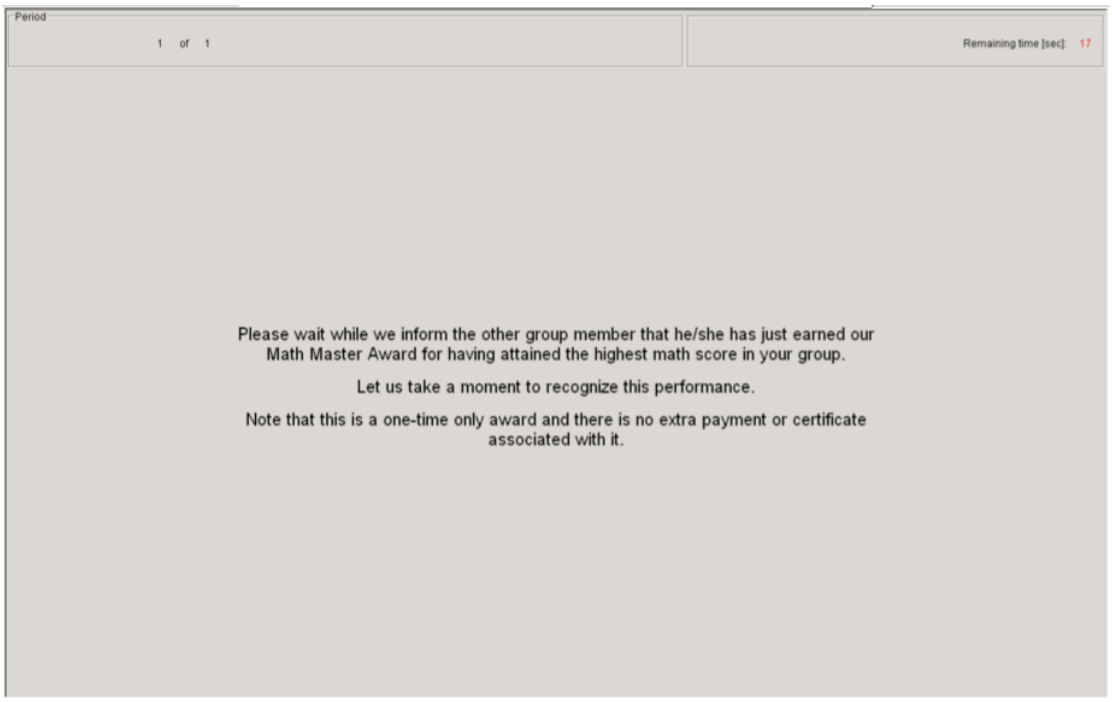


Figure A.2. Screen for non-recipients in the Virtual Award treatment arm.

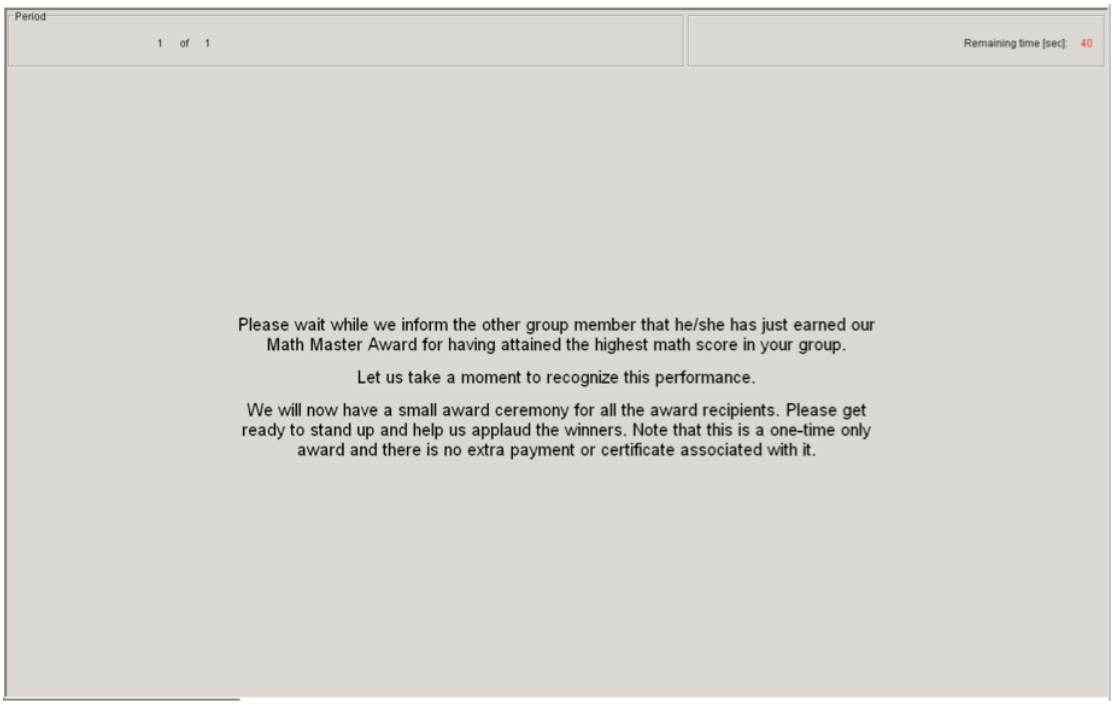


Figure A.3. Screen for non-recipients in the Award Ceremony treatment arm.

Table A.2: The role of domain stereotype in predicting confidence to speak up pre-treatment

	Math	Verbal	Pooled
Female	-0.179***	-0.0583	0.00801
	(0.0443)	(0.0673)	(0.0711)
Math			-0.424***
			(0.102)
Female*Math			-0.206**
			(0.0709)
Correct	1.008***	0.598***	0.925***
	(0.0327)	(0.0409)	(0.0273)
Part 2 Score	0.0526***	0.106***	0.0669***
	(0.00950)	(0.0247)	(0.00836)
Constant	1.513***	1.775***	1.895***
	(0.164)	(0.211)	(0.162)
Controls	Yes	Yes	Yes
Clusters	378	378	378
Observations	6755	2129	8884

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.3: The role of domain stereotype in predicting confidence to speak up post-treatment

	Math	Verbal	Pooled
Female	-0.0939*	-0.0840	-0.0365
	(0.0455)	(0.0672)	(0.0683)
Math			-0.648***
			(0.117)
Female*Math			-0.0697
			(0.0662)
Correct	0.860***	0.778***	0.839***
	(0.0386)	(0.0483)	(0.0316)
Part 2 Score	0.0658***	0.0506*	0.0746***
	(0.00969)	(0.0241)	(0.00886)
Constant	1.270***	2.000***	1.891***
	(0.203)	(0.239)	(0.183)
Controls	Yes	Yes	Yes
Clusters	378	378	378
Observations	7011	2203	9214

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

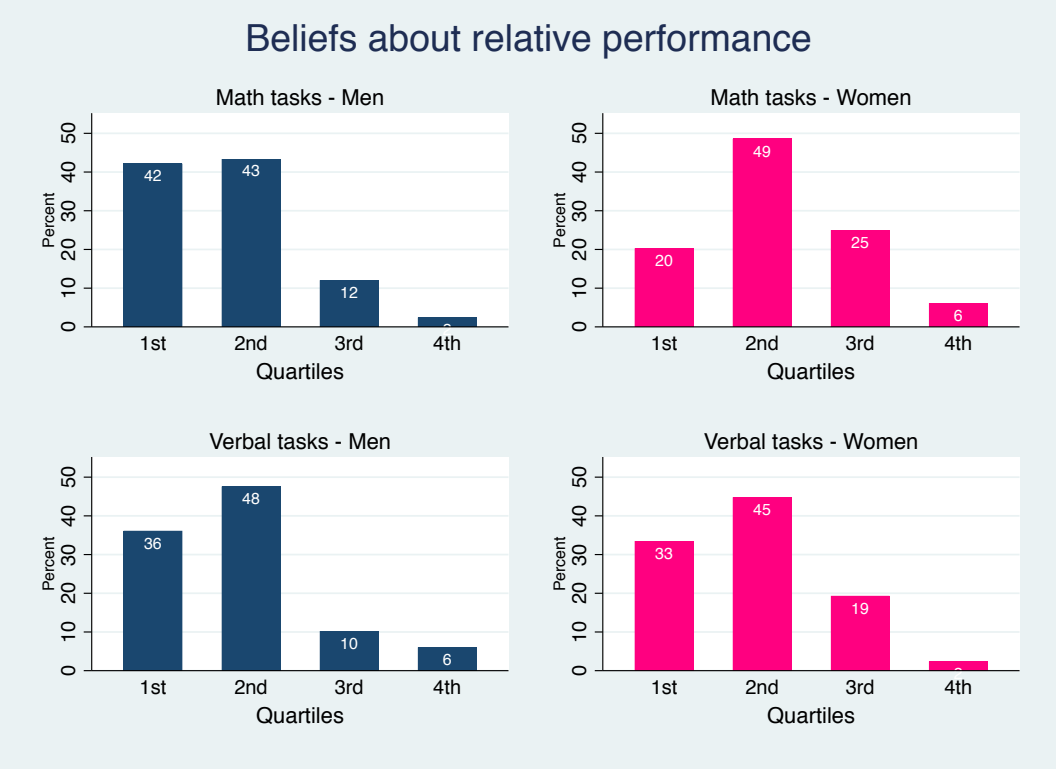


Figure A.4. Relative performance beliefs, by gender and stereotype of domain (math and verbal separately). "1st" indicates participant's belief to be among the best 25%, "4th" indicates belief to be among the bottom 25%. Beliefs elicited in an incentive compatible manner.

Table A.4 Testing the difference in reported beliefs about own performance quartile between men and women, for math and verbal tasks, respectively

	Men	Women	Diff.	Std. Error	P-value	Obs.
Performance Quartile Beliefs, Math	1.75	2.17	-0.42	0.08	0.000	378
Performance Quartile Beliefs, Verbal	1.86	1.91	-0.04	0.08	0.597	378

Note: p-value for gender difference in performance beliefs for math tasks is 0.0000005.

Table A.5: Testing whether self-stereotyping among women is reduced in the post-treatment period compared to the pre-treatment period

	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)
Math*Post-treatment			-0.162**
			(0.0512)
Female*Post-treatment	0.0833*	-0.0451	-0.0482
	(0.0373)	(0.0576)	(0.0593)
Female*Math*Post-treatment			0.134*
			(0.0667)
Correct	0.943***	0.688***	0.885***
	(0.0293)	(0.0342)	(0.0247)
Part 2 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
Clusters	378	378	378
Observations	13766	4332	18098

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.6: Testing for spillover effects on confidence to speak up in non-focal domain (verbal), across treatments and gender

	(Model 1)	(Model 2)	(Model 3)	(M4 Recip)	(M4 Non-Recip)	(M4 Allobs)
	Verbal	Verbal	Verbal	Verbal	Verbal	Verbal
Female		-0.0118		-0.00365	0.00898	0.0128
		(0.0971)		(0.0976)	(0.110)	(0.102)
Recognition	-0.0121	-0.0197				0.00189
	(0.0853)	(0.120)				(0.118)
Female*Recognition		0.0166				-0.0261
		(0.137)				(0.130)
Virtual Award			0.0674	0.0686	-0.113	-0.0111
			(0.114)	(0.114)	(0.131)	(0.0866)
Award Ceremony			0.119	0.119	-0.197	-0.0205
			(0.115)	(0.115)	(0.132)	(0.0870)
Post-treatment	0.0611	0.0610	0.0791	0.0362	0.0310	0.0314
	(0.0438)	(0.0438)	(0.0635)	(0.0922)	(0.129)	(0.0858)
Recognition*Post-treatment	0.0211	0.0650				0.0102
	(0.0564)	(0.0684)				(0.0559)
Female*Recognition*Post-treatment		-0.0821				
		(0.0713)				
Virtual*Post-treatment			0.00152	0.163	-0.103	0.0331
			(0.0839)	(0.132)	(0.185)	(0.113)
Ceremony*Post-treatment			-0.0130	0.105	0.226	0.158
			(0.0929)	(0.144)	(0.190)	(0.116)
Female*Private*Post-treatment				0.0823	0.157	0.120
				(0.141)	(0.173)	(0.111)
Female*Virtual*Post-treatment				-0.200	0.0853	-0.0556
				(0.140)	(0.171)	(0.110)
Female*Ceremony*Post-treatment				-0.153	-0.250	-0.201
				(0.145)	(0.179)	(0.113)
Correct	0.901***	0.902***	0.984***	0.991***	0.786***	0.904***
	(0.0373)	(0.0372)	(0.0495)	(0.0496)	(0.0551)	(0.0372)
Part 2 Score	0.00428	0.00381	-0.00363	-0.00513	0.0126	0.00436
	(0.0119)	(0.0121)	(0.0152)	(0.0161)	(0.0182)	(0.0122)
Constant	2.371***	2.382***	2.370***	2.389***	2.437***	2.372***
	(0.128)	(0.147)	(0.226)	(0.257)	(0.204)	(0.152)
Clusters	378	378	203	203	175	378
Observations	4332	4332	2345	2345	1987	4332

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

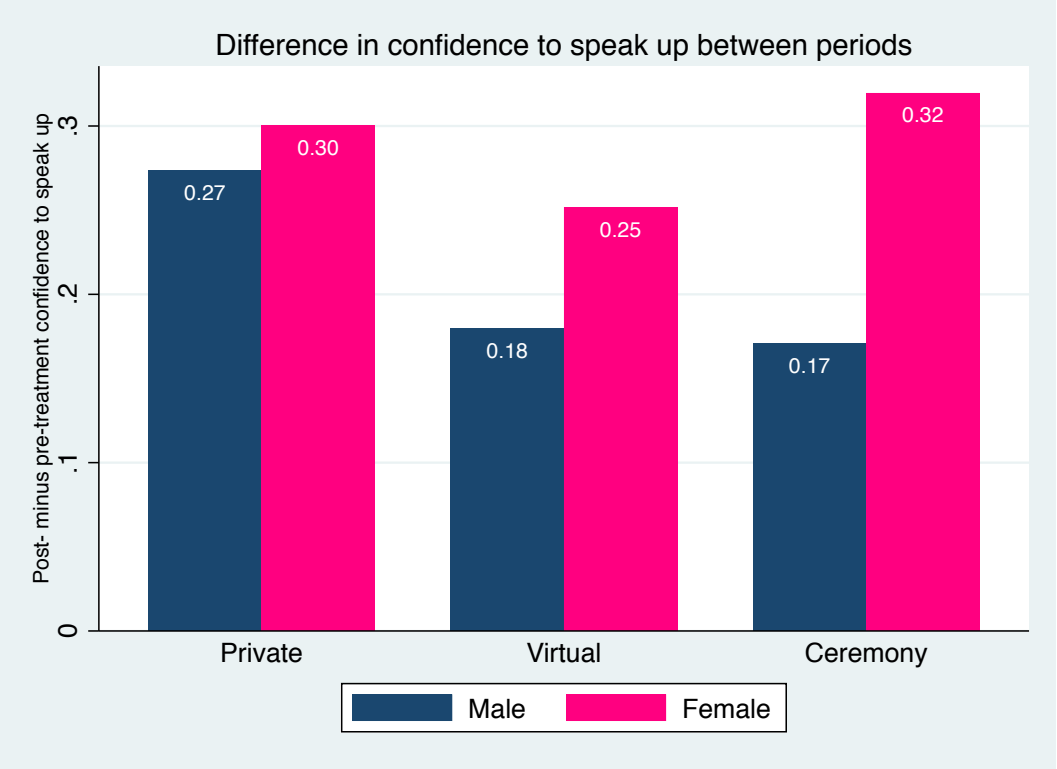


Figure A.5. Responsiveness to different recognition types by gender (recipients only). Higher values on y-axis indicate a greater difference in confidence to speak up, post- minus pre-treatment periods. The figure shows that women are more responsive to the Award Ceremony treatment compared to both men and to the other treatments (Private Feedback, Virtual Award). Hence, as seen in Figure 2, the gender gap almost disappears in the Award Ceremony treatment.

Table A.7: Testing for individual performance effects

	(Model 1) All obs	(Model 2) All obs	(Model 3) Recipients	(Model 4) Recipients	(Model 4) Non-Recipients	(Model 4) All obs
Female		-0.0310*		-0.00941	0.0000421	-0.0222
		(0.0145)		(0.00807)	(0.0108)	(0.0126)
Recognition	0.0350***	0.0188				0.0156
	(0.00922)	(0.0130)				(0.0133)
Female*Recognition		0.0264				0.0345*
		(0.0162)				(0.0173)
Virtual Award			0.0162	0.0170	0.00965	0.0119
			(0.0100)	(0.00990)	(0.0128)	(0.00804)
Award Ceremony			0.0122	0.0121	0.000718	0.00705
			(0.0102)	(0.0102)	(0.0117)	(0.00768)
Post-treatment	0.125***	0.125***	0.0947***	0.0913***	0.139***	0.135***
	(0.0114)	(0.0114)	(0.0120)	(0.0174)	(0.0293)	(0.0194)
Recognition*Post-treatment	-0.0394**	-0.0483**				-0.0408**
	(0.0139)	(0.0169)				(0.0139)
Female*Recognition*Post-treatment		0.0168				
		(0.0160)				
Virtual*Post-treatment			-0.0197	-0.0222	0.0346	0.00375
			(0.0193)	(0.0292)	(0.0456)	(0.0265)
Ceremony*Post-treatment			-0.00724	-0.0229	0.0409	0.00285
			(0.0185)	(0.0282)	(0.0422)	(0.0247)
Female*Private*Post-treatment				0.00648	-0.0136	-0.00106
				(0.0248)	(0.0407)	(0.0230)
Female*Virtual*Post-treatment				0.0101	-0.0772	-0.0286
				(0.0294)	(0.0426)	(0.0255)
Female*Ceremony*Post-treatment				0.0381	-0.0958*	-0.0242
				(0.0272)	(0.0378)	(0.0233)
Part 2 Score	0.0424***	0.0424***	0.0375***	0.0374***	0.0471***	0.0423***
	(0.00146)	(0.00148)	(0.00200)	(0.00199)	(0.00215)	(0.00149)
Constant	0.139***	0.157***	0.233***	0.239***	0.0890***	0.146***
	(0.0159)	(0.0191)	(0.0307)	(0.0310)	(0.0238)	(0.0186)
Clusters	378	378	203	203	175	378
Observations	13766	13766	7463	7463	6303	13766

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: The dependent variable is *Correct* (=1 if the question is answered correctly, 0 otherwise) in all columns of this table.

Table A.8: Testing for group performance effects

	(Model 5) Groups
Correct	0.425***
	(0.0120)
Partner correct	0.350***
	(0.0102)
Post-treatment	0.00736
	(0.00707)
Group Score Pre-treatment	0.0125***
	(0.00231)
Constant	0.0929**
	(0.0318)
Clusters	189
Observations	7165

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: The dependent variable is *Groupcorrect* (=1 if the group submitted the correct answer, 0 otherwise).

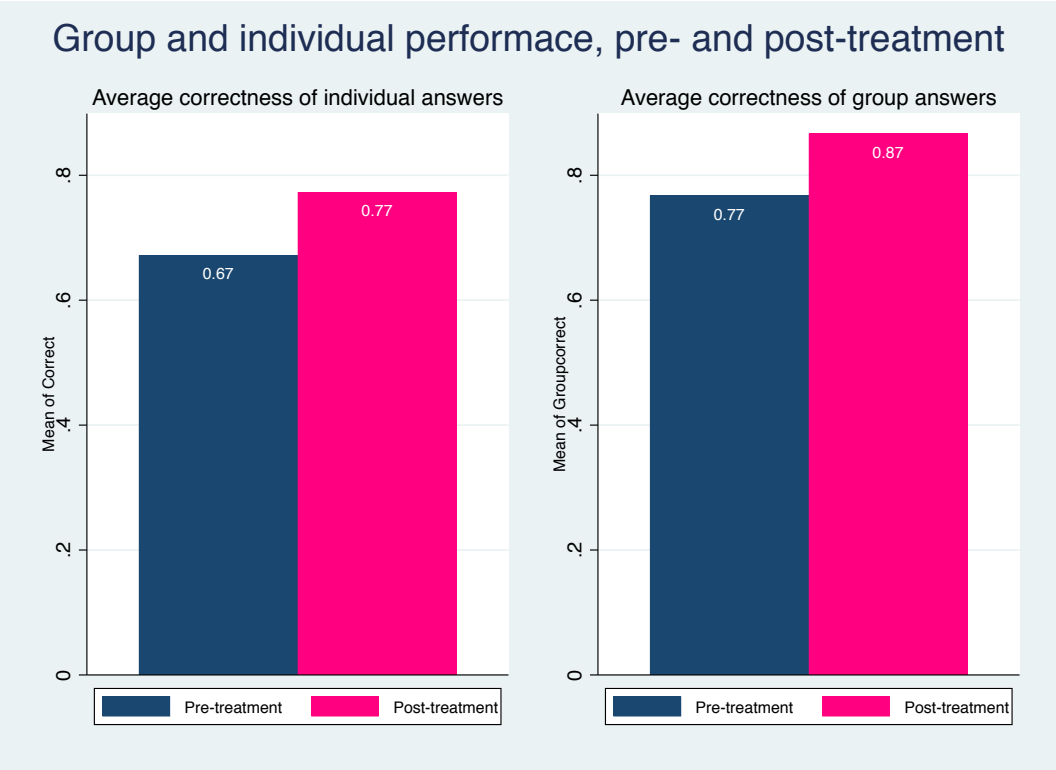


Figure A.6 Relationship between individual and group performance in the pre- and post-treatment periods

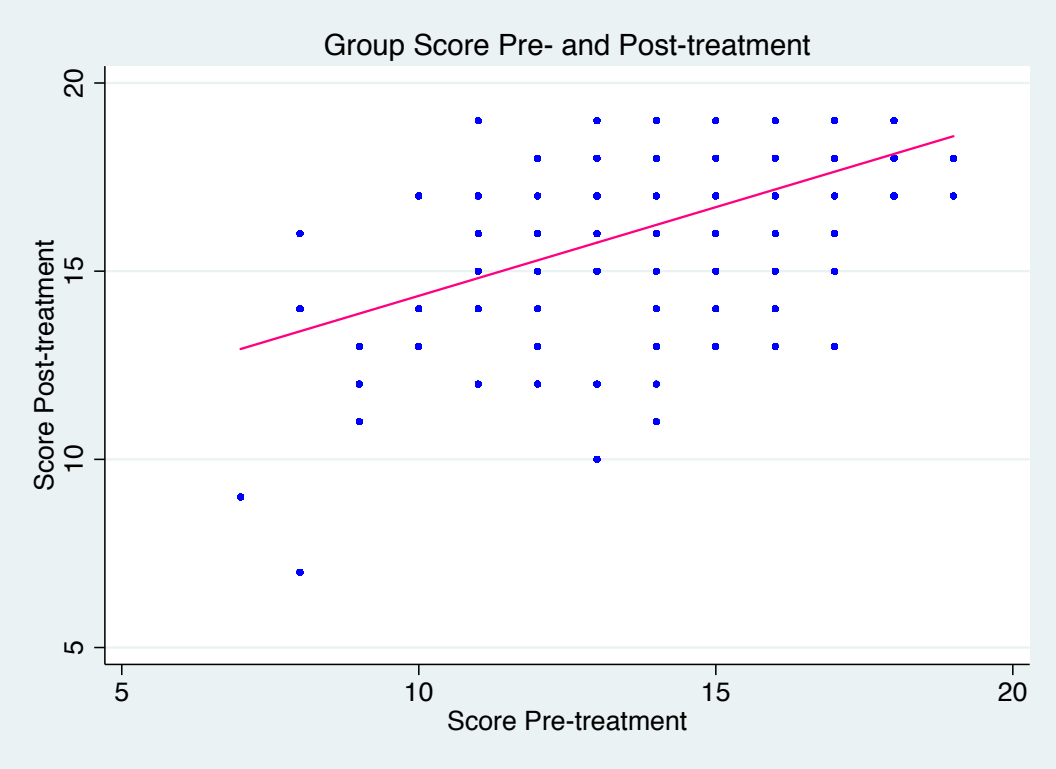


Figure A.7 Relationship between group performance in the pre- and post-treatment periods

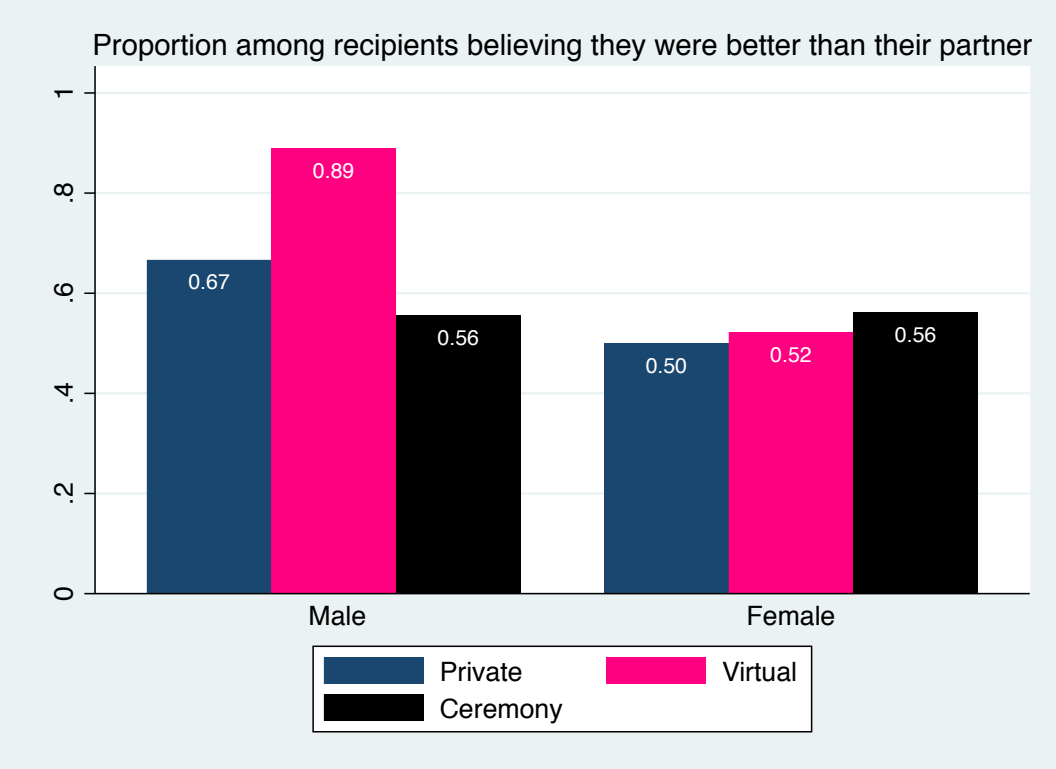


Figure A.8 Proportion of recipients who believe that they performed better than their partner

Notes: The question from the post-experimental survey was formulated as "In general, do you think that your partner answered more questions correctly than you did?" The answer choices were: "Yes, my partner answered more questions correctly." "No, I answered more questions correctly." "We answered an equal number of questions correctly." Not all subjects answered this question. For each of the six bars above, there are 9, 18, 18, 18, 23 and 16 observations, respectively. The directional differences seen among women indicate that they may have believed the award signal about their relative performance more when seeing the non-recipients standing in the room in the Award Ceremony treatment than when the recognition was privately communicated on the screen.