I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13871

The Effect of Manager Gender and Performance Feedback: Experimental Evidence from India

Martin Abel
Daniel Buchman

NOVEMBER 2020

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13871

The Effect of Manager Gender and Performance Feedback: Experimental Evidence from India

Martin Abel
Middlebury College, IZA and JPAL

Daniel Buchman
Middlebury College

NOVEMBER 2020

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

The Effect of Manager Gender and Performance Feedback: Experimental Evidence from India*

We hire 1,800 Indian gig economy workers for a real-effort transcription task and randomize the gender of the (fictitious) manager as well as the delivery of performance feedback. We find that negative feedback (i.e. criticism) leads to moderate deterioration in worker attitudes, but it increases effort provision in both mandatory and voluntary tasks. By contrast, praise affects neither attitudes nor effort provision. Importantly, feedback effects do not vary between workers assigned to female and male managers. Consistent with this finding, there is no evidence for attention discrimination towards female managers, implicit gender bias, or gendered expectations among workers. By contrast, Abel (2019) employs the same research design in the U.S. and finds substantial gender discrimination and no effect of feedback on effort. This highlights that the effects of feedback and manager gender vary across different contexts.

JEL Classification: J50, J70
Keywords: India, gender discrimination, gig economy, feedback

Corresponding author:
Martin Abel
Middlebury College
14 Old Chapel Rd
Middlebury, VT 05753
USA
E-mail: mabel@middlebury.edu

* This paper greatly benefited from discussions with and comments by Willa Brown, Tanya Byker, Jeff Carpenter, Siddharth George, Erik Gong, Caitlin Myers, Oyebola Okunogbe, Akhil Rao, Emcet Tas, and Maya Woser. This study was conducted in partnership with the World Bank's South Asia Gender Innovation Lab. The experiment was registered under registry number AEARCTR-0003810 and IRB approval was obtained from Middlebury College. Declarations of interest: none. All errors and omissions are our own.

1 Introduction

As in many other parts of the world, women in India are closing the education gap: 53% of undergraduate degrees and 42% of PhD degrees were awarded to women in 2019 (Catalyst, 2019). Yet, the female labor force participation rate stands at only 23.6%, which is low compared both to India's male labor force participation rate of 78.6% and the female labor force participation rates of countries at comparable income levels. This gender gap is even larger for leadership positions in the private sector, where women occupy 16% of mid-level management positions and 4% of senior-level positions (McKinsey, 2018). These massive gender differences despite a convergence of educational attainment suggest an inefficient allocation of talent. Aside from gender equity concerns, this also hampers economic growth (Hsieh et al., 2019). Negative economic consequences may be particularly severe in developing countries like India where firms often lack access to effective managers (Bloom et al., 2012, 2013; McKenzie and Woodruff, 2015).

Popular press narratives, personal anecdotes, and a large body of research documents accounts of gender discrimination across India's labor market (Batra and Reio Jr, 2016; Giri, 2014; Zimmermann, 2012; Budhwar et al., 2005). There are few studies, however, that systematically quantify the degree of gender discrimination, and those that do tend to focus on job candidates or entry-level workers (Batra and Reio Jr, 2016). We know less about the barriers women face as they begin to occupy leadership positions. As the share of women in such positions increases – both in India and around the world – understanding discrimination against female managers becomes all the more important. This paper tests whether gender discrimination by subordinates presents a barrier for female managers in India.

A critical component of discrimination by subordinates is the way in which workers react to feedback. An effective manager must be able to deliver feedback that elicits positive attitude and behavior change. Yet, the evidence on the effectiveness of feedback is mixed, suggesting that such effects are highly context-specific. This has two implications: first, it highlights the need to understand further the underlying mechanisms that may explain these conflicting results, and second, it raises questions about the generalizability of existing studies. Underscoring this point is the fact that most evidence on feedback effects comes from Western contexts.

This study aims to fill these gaps and provide some of the first evidence on the effect of feedback as well as the effect of manager gender in a developing country context. We recruit a sample of 1,800 Indian workers through Amazon's Mechanical Turk (MTurk) for a transcription task and randomize the gender of workers' (fictitious) manager as well as whether they receive feedback on their performance. This design replicates Abel (2019), which uses a sample of MTurk workers in the U.S. and will serve as a benchmark for our results.

Our study contributes four main results: First, in the absence of feedback, we find people

provide the same level of effort working under female and male managers. Second, feedback receipt leads to an increase in effort provision for both mandatory and voluntary tasks. This is driven by the positive effects of negative feedback (i.e. criticism) on effort provision, which we observe despite a slight deterioration in worker attitudes when criticized. By contrast, praise does not affect attitudes or effort provision, which could be explained by workers' expectations of praise. Indeed, we find that workers assess positive performance feedback to be highly accurate and appropriate.

Third, we find the effect of feedback does not vary by manager gender. Workers respond similarly to feedback from female and male managers for both praise and criticism. Fourth, as the above results may be surprising given the frequency of accounts of gender discrimination in India, we provide evidence on mechanisms that may explain the apparent absence of gender discrimination in our setting. We find no or limited evidence for the importance of attention discrimination (Bartos et al., 2016), *implicit* gender bias (Abel, 2019), and gendered expectations (Carli, 2001; Manne, 2017) in our setting. Workers spend the same amount of time reading and thinking about the feedback of female managers. In a test of the degree to which workers implicitly associate men with careers and women with family, we observe a modest level of implicit gender bias. Importantly, implicit bias does not negatively affect workers' performance or attitude when paired with a female manager. We also do not find that workers hold *explicit* biases about the qualification of female managers. Finally, workers are much less likely to associate certain management styles with a specific gender compared to their U.S. counterparts.

This paper relates to several stands of literature. We add to an established literature on the effects of feedback. Some papers find negative effects on productivity (Eriksson et al., 2009; Kuhnen and Tymula, 2012) while others find positive ones (Charness et al., 2014; Azmat and Iriberri, 2010; Bandiera et al., 2015; Blanes i Vidal and Nossol, 2011). We contribute novel evidence from the growing gig economy and shed light on some of the underlying mechanisms, further explaining why there may be conflicting evidence across different settings.

Many cultural differences affect workplace relationships and communication, including community orientation, comfort with ambiguity, hierarchy, etc. (Varma et al., 2005; Amba-Rao et al., 2000; Sully De Luque and Sommer, 2000). It is plausible, therefore, that these differences affect how people deliver and respond to feedback. One important example is the prevalence of gendered expectations. There is ample and very robust evidence showing that people in the U.S. and other Western countries strongly associate men with criticism and women with praise (Carli, 2001). Importantly, we see a different pattern of associations among gig economy workers in India. For example, strict expectations are equally associated with female and male managers. Receiving critical feedback from a woman therefore does not create a dissonance with expectations. The lack of defensiveness is also reflected by the fact that workers assess (negative) feedback from female and male managers as equally accurate and appropriate. This may be one of the explanations for why we observe less gender discrimination towards critical female managers than in the U.S. sample (Abel, 2019).

Our study also adds to a nascent literature investigating "discrimination from below" (Ayalew et al., 2018). Existing research employs lab and field experiments as well as evidence from survey data to explore subordinates' reactions to female managers.[1]. This literature has yet to arrive at a firm consensus. Some research finds that workers, particularly men, are less likely to trust the advice or follow the instructions of female managers, but other studies suggests these effects dissipate at the most senior levels (Budhwar et al., 2005; Grossman et al., 2016; Ayalew et al., 2018; Macchiavello et al., 2020).

One important aspect of our study design is that it looks at short-term supervisor-subordinate relationships without face-to-face interactions, which may limit the generalizability of results to more traditional office environments.[2] However, the gig economy and remote work agreements have become much more prominent in recent years and are expected to increase in frequency in the wake of the COVID-19 pandemic. In addition, research suggests that the type of text-based feedback we provide can be as salient as its verbal counterpart (Bracken et al., 2004). Firms increasingly gain access to large amounts of data and technology that allows them to provide real-time, automated feedback (Cecchi-Dimeglio, 2017). Our results suggest that in the context of remote gig economy workers in India, this type of feedback is effective in increasing worker effort, even for voluntary tasks.

The rest of the paper is structured as follows: Section 2 provides background information on gender discrimination in India. Section 3 describes the study design and Section 4 presents the main results. Section 5 discusses potential mechanism and Section 6 concludes with a discussion.

2 Background

2.1 Gender discrimination in India

Indian female labor force participation has fallen in recent years. Some attribute this to rising household incomes, but the point remains: less than a quarter of women in India work in the formal economy. Despite scoring lower on the Human Development Index and having lower per capita incomes, Bangladesh, Bhutan, and Nepal all have significantly higher female labor force participation rates at 36, 59, and 83 percent, respectively. In fact, India's rate is comparable with those of Pakistan and Saudi Arabia, countries where women struggle to attend school and are limited in their ability to work. This remarkably low rate is reflected in the fact that women's contribution to India's GDP is approximately 18%. McKinsey (2018)

[1] A related literature shows biases against female instructors in student evaluations (Buser et al., 2019; Boring, 2016; Mengel et al., 2019)

[2] Indeed, a recent study by Macchiavello et al. (2020) finds that in Bangladesh's garment sector, female managers are initially less effective as workers have negative beliefs about their ability.

suggests the biggest opportunity for advancing GDP growth in the next decade will be in bringing more women into the job market.

Unfortunately, numerous barriers still exist, which prevent women from entering the workforce, discrimination being chief among them. The literature on discrimination in India indicates that, from a young age, boys are given more attention and resources in the home (Barcellos et al., 2014; Dhar et al., 2018; Jayachandran and Pande, 2017). In the workplace, some suggest India's historically paternalistic culture has led to a consistent under-valuing of women's contributions (Batra and Reio Jr, 2016). Additionally, women face unique challenges in starting small business (Giri, 2014) and continue to receive less pay despite performing better than male counterparts in certain sectors (Ranganathan and Shivaramb, 2019). Some of the more recent evidence finds, however, that female professionals in India report discrimination does not limit their professional opportunities (Abbas, 2015).

2.2 Women in Management in India

The role of women in India's managerial class, particularly in the corporate sector, is growing, albeit from a very low rate. In 2005, the number of women in managerial positions ranged between 3 and 6 percent of administrative positions; today, it sits around 16 percent for mid-level management (Budhwar et al., 2005; McKinsey, 2018). Western HR practices and the celebration of certain "success stories" indicate a shift in cultural attitudes as well, though it is worth noting such stories may be celebrated for their exceptional nature. Most public institutions fully support gender diversity efforts, and some private-sector employers have begun gender diversity initiatives (Lockwood, 2009). A recent legal mandate helped bring board-level female representation up to 11%, though that number is still low relative to neighboring countries. Despite all this, women in top leadership roles are still in the minority, accounting for only 4% of senior-level management. (McKinsey, 2018).

Some older evidence suggests that there remains opposition towards women in leadership among men in the workforce. For example, Gupta et al. (1998) documents that men perceive women to be "less self-confident, more emotional, and less objective." Along similar lines, Budhwar et al. (2005) finds that many female managers perceive the ego of their male subordinates as an obstacle to their effectiveness as managers. More recently, however, Abbas (2015) uses survey data to show no significant negative bias against female managers. It remains unclear if these changes in reported attitudes reflect actual societal changes. Additionally, few studies discuss whether these shifts translate into better treatment of women in leadership roles.

2.3 India's Gig Economy

India has one of the youngest work forces in the world, and many of those tech-savvy, educated millennial workers will likely seek more flexible remote work arrangements as they enter the labor market. Ipsos (2011) found nearly a third of Indians work from home daily and over half are frequent remote workers, and a more recent 2016 survey revealed that one in two Indians preferred telecommuting (Basu, 2016). A recent "Future of Jobs" report found that Indians are a leading source of labor in the online gig economy with a 24% share of the global online gig workforce. The vast majority of gig workers, approximately 80%, joined within the last five years, and though the nature of gig work varies, approximately 31% are focused on repetitive tasks similar to those used in our study design (NobleHouse, 2019).

Though research is not yet definitive, the COVID-19 pandemic is likely to accelerate many of these trends in the gig economy and remote work. While some have proposed that these type of work agreements will benefit women, as they offer more flexibility (Goldin, 2014), these jobs also tend to have less regulatory oversight and may thus be particularly prone to discrimination.

3 Experimental Design

3.1 Recruitment and Sample Characteristics

We recruit study participants via Amazon's Mechanical Turk (MTurk).[3] Respondents are limited to MTurk workers from India over 18-years-old. We further exclude workers with HIT approval ratings below 95% and limited previous experience (less than 50 completed HITs) in order to improve data quality and prevent bots from completing the task.[4] The HIT description simply said "Participate in transcription task" and advertized a flat rate payment of Rs. 63 ($0.84) for a 15 minute task.[5]

[3]Mechanical Turk (MTurk) is an online platform that allows firms to recruit workers for simple tasks, known as Human Intelligence Tasks (HITs). MTurk has been widely used in market research and is becoming increasingly popular for academic research (see for example (DellaVigna and Pope, 2018)).

[4]We further divided data collection into small batches of 30 to 100 workers posted at different times on different days thus minimizing risks that our HIT is discussed and shared in online fora.

[5]The corresponding hourly rate of $3.35 is slightly above the average pay for MTurk workers in India. As discussed in more detail in Abel (2019), the flat rate increases reliance on workers' intrinsic motivation. Activating intrinsic motivation is one of the key challenges of successful managers.

Table 1: Sample Characteristics and Balance Test

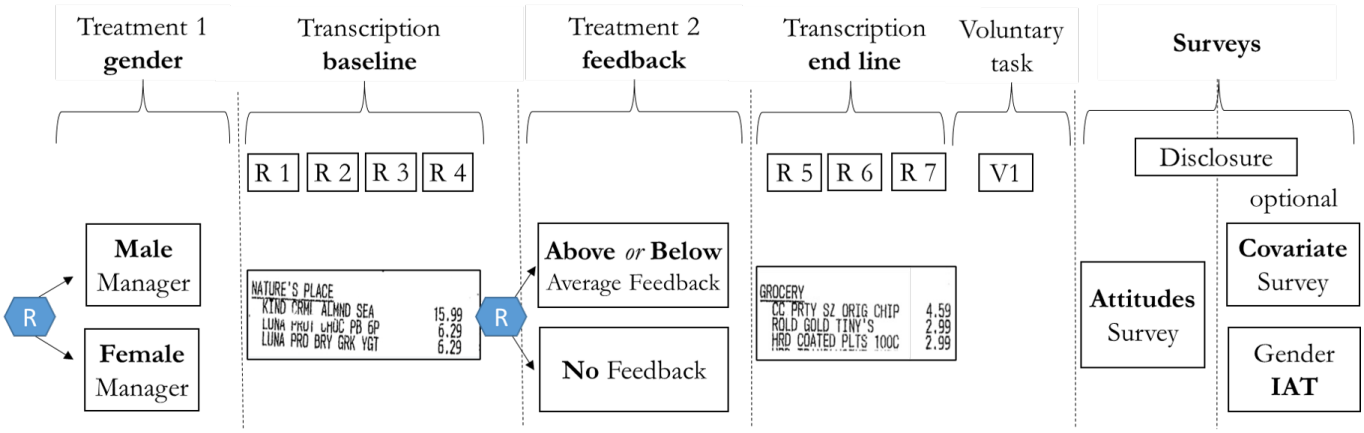
			Manager Gender			Feedback		
	N	Mean	Female	Male	p-value	Yes	No	p-value
Age	1751	32.15	32.15	32.15	.99	32.1	32.22	.77
Female	1765	.27	.26	.28	.38	.27	.26	.5
University	1755	0.45	0.45	0.46	.72	0.46	0.45	.59
Lower caste	1761	.46	.47	.45	.57	.47	.44	.2
Upper caste	1761	.21	.22	.2	.16	.2	.23	.11
Experience Fem Mgr	1274	.75	.75	.76	.81	.76	.74	.38
Experience Male Mgr	1273	.92	.92	.93	.23	.92	.93	.75
Fem Mgr Effective	961	3.17	3.14	3.2	.26	3.2	3.12	.13
Male Mgr Effective	1178	3.12	3.09	3.15	.14	3.12	3.12	.99
Gender Implicit Bias	448	.2	.2	.21	.68	.21	.2	.9
Joint Significance (p-v)					.60			0.45

Notes: The first two columns show the sample size and sample mean for each variable. The next three columns compare characteristics across the random *manager gender* assignment and provide a p-value for a test of equal means. The last three columns compare characteristics across the randomized *feedback* delivery and provide a p-value for a test of equal means.

A total of approximately 1,800 Indian workers completed the HIT between January and August 2020. The second column in Table 1 shows characteristics of the full sample. As in other parts of the world, Indian gig economy workers are not representative of the overall population. Our sample is younger and more educated, with 45% of participants having completed university. Only 27% of respondents are female, which is reflective of India's overall low female labor force participation. Additionally, 46% of respondents are members of lower castes (mostly OBC) and 21% are members of upper castes. While we have participants from 28 Indian states (particularly Tamil Nadu), respondents are clustered in southern India where most of India's IT industry is located.

Regarding previous work experience, 75% report having had a female supervisor in the past, which is below the share with former male supervisors (92%) but still high given the relatively small share of women in managerial positions. Levels of implicit bias are relatively low with an average score of 0.2 (compared to 0.33 in the U.S. sample) on a scale where 0 presents no implicit bias and 1 presents strong gender bias.

Figure 1: Design Overview



3.2 Randomization

The study design, summarized in Figure 1, closely follows the design of Abel (2019). There are two stages of randomization. First, half of participants are assigned to either a fictitious female or male manager. The introduction states that "*Our manager NAME might check in with you during the task.*" The gender of the manager is communicated via the title (Ms. vs. Mr.) followed by either a female or male first name.[6] First names may communicate other characteristics beyond gender (Fryer Jr and Levitt, 2004). We therefore elicited associations of different Indian names with levels of education, age, caste and location in an out of sample survey on MTurk. Overall, there is little variation among common Hindu first names and gender differences are small (see Figure A5). We matched female and male manager names so that these characteristics are balanced by manager gender. Female manager names include Anita, Seema, Sarika and Sharmila. Male manager names include Anil, Ashok, Dinesh, and Kamal. Table 1 suggests that randomization of manager gender was successful. None of the baseline characteristics differ between the groups at statistically significant levels and we can reject that these characteristics are jointly significant (p-value=0.60).

After recipients describe four receipts (discussed in more detail below), 60% are randomized into receiving performance feedback. The content of the feedback depended on their actual performance in completing tasks associated with the initial four receipts.[7] Specifically, the text for **positive** / **negative** feedback reads:

[6]Last names were randomized to be associated with oppressed and dominant castes. We henceforth refer to these as lower or upper castes. This randomization is used to estimate the effect of manager caste in a companion paper. We included common last names including Iyer, Kaushik, Menon Sharma, Shastri, Chamar, Kori, Pasi, and Manjhi. In all of our analyses, we control for the caste of the manager. As discussed in the companion paper, differences in manager caste perceptions are relatively small, arguably because of regional differences in familiarity with names. This limits the extent to which we can estimate interaction effects of the caste and gender of the manager.

[7]The scoring was automated so that we could provide instant feedback.

Hello,
This is MS./MR. NAME. As mentioned in the task introduction, I'm overseeing your performance in transcribing the receipts.

I just went over some of the receipts. Your performance has been ***above*** */* ***below*** *average. I was* ***pleased with*** */* ***disappointed by*** *your effort and attention to detail.*

Going forward, remember that your ***continued commitment will improve*** */* ***lack of commitment will harm*** *the quality of our services.*
MS./MR. NAME

As discussed in more detail in Abel (2019), the text includes the standard components of feedback. Table 1 shows that characteristics are balanced between the feedback and no feedback group. We can reject that the variables are jointly significant (p-value=0.45).

3.3 Transcription Task and Outcomes

We choose a real effort transcription task and ask participants to copy both the item description and price of a scanned copy of an actual receipt. This task includes both language and math components and may thus be seen as more gender neutral (Niederle and Vesterlund, 2007). In fact, female and male workers perform at similar levels in this task (Table 2). Furthermore, these types of transcription tasks are among the most common tasks on MTurk. The introduction of the HIT states that the transcription task can "*help businesses understand ... how to spend money better so they can increase their bottom line*".[8] Misleading participants about the purpose of the research is necessary for most discrimination studies to ensure that participants do not try to hide discriminatory attitudes and behaviors.

3.3.1 Effort outcomes

Transcription results offer us an objective measure of workers' effort and accuracy. Similar to List and Momeni (2017), we present workers with receipts of varying readibility and first ask them whether that receipt is legible (see Figure 2). On average, 75% of receipts were deemed legible. This presents a measure of effort at the extensive margin. In order to also measure the quality of the work, we construct a transcription score. Each correctly transcribed item

[8]The HIT introduction does mention that "*performance data will be recorded for research purposes*" This type of data collection arrangements become increasingly common, not just from customers but also from workers.

counts as 1 and each correct price as 0.5.[9] The average score is 65.6, which is almost identical to the performance of U.S. workers. These outcomes measure effort related to *mandatory* tasks, as specified in the HIT introduction.

We also measure effort in *voluntary* tasks. For many real-world tasks, individual effort or output is hard to observe. One of the key challenges of managers is to leverage intrinsic motivation of workers for tasks that may be hard to contractually specify (Podsakoff et al., 1990). We measure voluntary effort in two ways. For each legible receipt, we ask workers to add up all item prices. We emphasize that this is "*not a required task and will not affect your payment*". Overall, 62% of workers complete this task, which is about 30% higher than workers in our U.S. sample (Abel, 2019). One issue with this measure is that we only collect it for those who say the receipt is legible. To address potential selection issues, we introduce a second measure. After the completion of the transcription, we ask participants to complete the "completely voluntary" task of grouping receipts according to the total sum.[10] The combination of mandatory and voluntary effort allows us to test whether (negative) feedback crowds out intrinsic motivation and leads workers to reallocate effort to the incentivized outcomes similar to multitasking concerns (Holmstrom and Milgrom, 1991).

Figure 2: Receipt Example

3.3.2 Attitude

After the completion of the transcription but before the debrief, we collect data on workers' attitudes. Specifically, we first ask whether workers want to "sign up to do additional work for

[9]For receipts deemed illegible, we set values for effort measures (adding, transcription accuracy) to zero. Results are unchanged in magnitude and significance when we different scoring rules, e.g. assigning 60% of possible points for skipping a receipt. This rule reflects that very poor performance is often more detrimental than not completing the task at all.

[10]Specifically, the instructions read: *I need to group receipts into two categories, depending on whether the sum of all listed items is greater than 10.00. For each of the receipts below, please indicate whether the items listed add up to an amount greater than 10.00.*

us in the next few days?" This serves as a proxy for worker retention, which is an important challenge for many firms (Hoffman et al., 2017). Next, we ask how much workers agree that they are (a) satisfied with the "experience transcribing receipts" and (b) convinced that "the task was important". Last, we ask those who received feedback whether they agree that the feedback was accurate and appropriate (see Appendix A4 for the complete survey).

3.4 Estimation Strategy

As described, the experimental design features two stages of randomization: the gender of the manager (*FemMgr*) and whether workers receive feedback (*Feedback*). The effect of feedback and manager gender can thus be estimated through the following specification:

$$y_i = \beta_0 + \beta_1 FemMgr_i + \beta_2 Feedback_i + \beta_3 FemMgr\, x\, Feedb_i + \epsilon_i \quad (1)$$

Here, y_i measures outcomes y for worker i. As specified in the pre-analysis plan, we collect two types of outcomes y: measures of effort (legibility, adding, transcription score, voluntary task) and attitude (interest in future work, task satisfaction, task importance). We combine the three attitude measures into a standardized index to allay multiple hypothesis testing concerns (Kling et al., 2007). β_1 measures the effect of a female manager who does not provide feedback, β_2 is the effect of feedback from a male supervisor, and β_2+β_3 represents the effect of feedback from a female supervisor.

To estimate the effect of feedback *content*, we estimate:

$$y_i = \gamma_0 + \gamma_1 AboveAv_i + \gamma_2 Feedback_i + \gamma_3 AboveAv\, x\, Feedb_i + \epsilon_i \quad (2)$$

Here, *Aboveav* is a binary variable indicating if workers performed above average. Therefore, γ_2 measures the effect of negative feedback, and γ_2+γ_3 captures the effect of positive feedback.

4 Results

Following our pre-analysis plan, we first present the effect of manager gender before workers receive any feedback (4.1), then discuss the effects of feedback (4.2) and finally look at the interaction of manager gender and feedback (4.3).

4.1 Manager Gender and Effort Provision

Table 2 shows that before receiving feedback, manager gender has no effects on the level of effort workers exert for both mandatory and voluntary tasks. Coefficients are precisely estimated and very close to zero at about 0 to 0.02 standard deviations (Col. 1, 3, 5). The effect of manager gender also does not vary significantly by workers' gender, although it is notable that coefficients on the interaction of female managers and female workers is negative (Col. 4, 6).

Table 2: Effect of Manager Gender on Effort (Baseline)

	Legible		Adding		Transcription Score	
	(1)	(2)	(3)	(4)	(5)	(6)
Female Mgr	-0.001	-0.001	0.009	0.019	-0.006	0.167
	(0.015)	(0.017)	(0.019)	(0.022)	(0.279)	(0.330)
Female Worker	0.020	0.020	0.019	0.037	0.413	0.741*
	(0.016)	(0.023)	(0.021)	(0.031)	(0.307)	(0.436)
Fem Mgr x Fem Wkr		-0.000		-0.035		-0.648
		(0.033)		(0.043)		(0.617)
Observations	1765	1765	1765	1765	1765	1765
Sample Mean	0.75	0.75	0.62	0.62	13.45	13.45
Std Dev	0.31	0.31	0.40	0.40	6.02	6.02
P-value		0.96		0.65		0.36

Notes: The dependent variable in Column (1) and (2) is whether the worker says the receipt is legible. The dependent variable in Column (3) and (4) measure if the worker is willing to add up the amounts. The dependent variable in Column (5) and (6) captures the accuracy of transcribing receipts All estimations are OLS. Robust standard errors are in parentheses. *P-value* presents results testing if female managers have a positive effect for female workers. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.2 Effect of Feedback Content

4.2.1 Attitude

Table 3 shows the effects of feedback on the three attitude outcomes as well as the standardized attitude outcomes. The aggregate effect of feedback is negative but small in magnitude (about 0.05 s.d.) for all attitude outcomes (Col. 1, 3, 5, 7). These aggregate effects mask important differences by feedback content. The feedback coefficients in Col. 2, 4, 6 and 8

show that negative feedback leads to a drop in attitude outcomes by about 0.15 s.d.. This estimate is significant at the 5% level for the attitude index (Col. 8). By contrast, the effect of praise is very close to zero and not statistically significant, as indicated by p-values close to one reported in the bottom row.

Table 3: Effect of Feedback Content on Attitudes

	Work Future		Satisfaction		Task Import		Attitude Index	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Feedback (FB)	-0.010	-0.020	-0.038	-0.122	-0.040	-0.120	-0.043	-0.112**
	(0.009)	(0.014)	(0.046)	(0.075)	(0.053)	(0.080)	(0.034)	(0.055)
Above Av x FB		0.018		0.114		0.123		0.104
		(0.018)		(0.090)		(0.104)		(0.068)
Above Average		0.007		0.470***		0.275***		0.274***
		(0.013)		(0.073)		(0.086)		(0.053)
Observations	1747	1747	1747	1747	1747	1747	1747	1747
Sample Mean	0.96	0.96	5.27	5.27	5.16	5.16	-0.00	-0.00
Std Deviation	0.19	0.19	0.97	0.97	1.08	1.08	0.73	0.73
Pos FB=0 (p-v)		0.90		0.86		0.95		0.85

Notes: *Work future* is a binary variable whether workers are interested to work for the firm in the future. *Job satisf.* and *Task Import* measure how strongly workers agree that they were satisfied with the task and that the task was important, respectively (0=strongly disagree, 6= strongly agree). *Index* is a standardized index of the three outcome measures. *Abv Avg* is a dummy if the worker performed above average. All estimations are OLS. Robust standard errors are in parentheses. P-values report tests of equal coefficients across worker gender.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

These results are similar across worker gender, although it is notable that the negative effect of criticism on attitude tends to be larger for male workers (Table A1). While this difference is not statistically significant, it goes against a common narrative that women may react more negatively to criticism (Correll and Simard, 2016). As discussed in more detail in Section 6, it is noteworthy that the overall pattern of negative effects of criticism and no effect of praise are similar to results from the U.S. with the important difference that the negative effects of criticism in India are only about one third in magnitude compared to those in the U.S..

4.2.2 Effort

Table 4 shows the effects of feedback on effort. The aggregate effects are positive and moderate in size, ranging between 0.1 and 0.15 s.d. (Col. 1, 3, 5, 7). These positive results

are driven by the effects of criticism. Negative feedback increases the share of legible receipts by 5.1 p.p. (0.16 s.d.), increases the transcription score by 0.13 s.d. and even increases the share completing the voluntary task by 5 p.p. (0.13 s.d.). This last result is particularly surprising: despite the negative effect of criticism on workers' attitude, it increases effort for both mandatory and voluntary tasks.

Table 4: Effect of Feedback Content on Effort

	Legible		Adding		Transcrip Score		Voluntary Task	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Feedback (FB)	0.040***	0.051**	0.028	0.018	0.658***	0.651**	0.041**	0.050*
	(0.014)	(0.022)	(0.018)	(0.026)	(0.232)	(0.329)	(0.020)	(0.030)
Above Av x FB		-0.046*		-0.010		-0.443		-0.020
		(0.026)		(0.034)		(0.404)		(0.040)
Above Average		0.286***		0.323***		5.147***		0.048
		(0.020)		(0.026)		(0.314)		(0.032)
Observations	1747	1747	1747	1747	1747	1747	1610	1610
Sample Mean	0.68	0.68	0.56	0.56	9.82	9.82	0.82	0.82
Std Deviation	0.30	0.30	0.39	0.39	4.98	4.98	0.39	0.39
Pos FB=0 (p-v)		0.74		0.71		0.40		0.27

Notes: The dependent variable in Column (1) and (2) is whether the worker says the receipt is legible. The dependent variable in Column (3) and (4) measure if the worker is willing to add up the amounts. The dependent variable in Column (5) and (6) captures the accuracy of transcribing receipts Column (7) and (8) use the score as the dependent variable and divide the sample by worker gender. All estimations are OLS. Robust standard errors are in parentheses. P-values report tests of whether the effect of praise is different from zero. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The effect of praise on effort is very similar to the attitude results. For both mandatory and voluntary tasks, estimates are close to zero and not statistically significant. These effects are again similar by worker gender (Table A2). One explanation for the null effect of praise on both attitude and effort is that workers expect to be praised and thus do not change their attitude or effort. As discussed in more detail below, workers think that positive feedback is highly accurate and appropriate.

4.3 Effect of Manager Gender

This section looks at the effect of manager gender and feedback. The aggregate effect of feedback can be estimated through an interaction term (specification 1). To look at the interaction of manager gender and feedback content, we need to estimate models with a

triple interaction of manager gender, feedback, and performance. While these results are presented in the appendix, this section will show graphs with non-parametric results for greater transparency and ease of interpretation. This strategy effectively compares average outcomes across eight groups. One limitation is that the cell sizes are smaller (on average 220 workers), which reduces the precision of estimates.

4.3.1 Attitude

Figure 3 shows how the effect of feedback on the attitude index differs by manager gender for negative feedback (left panel) and positive feedback (right panel). The drop in attitudes in response to negative feedback is remarkably similar between male and female managers (p-value=0.86). Similar to results from the U.S. we find that for both female and male managers positive feedback has no effect. These results are similar for the three attitude outcomes that comprise the attitude index (Table A3). There is also no significant difference when we look at the interaction of manager and worker gender (results available on request).

4.3.2 Effort

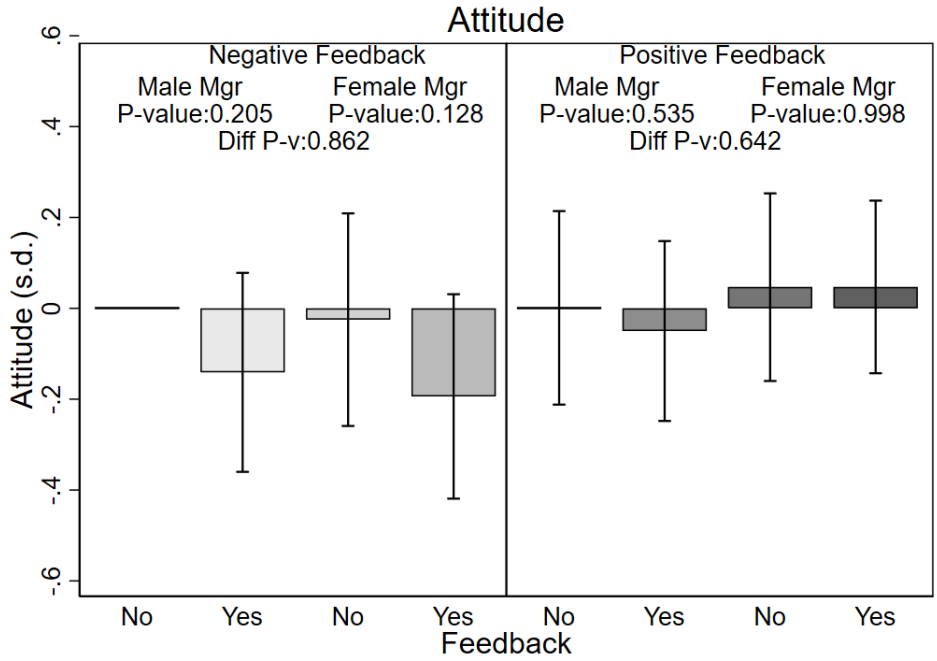
Similar to the results on worker attitudes, there is little evidence of gender discrimination in terms of effort provision. Figure 4 shows that responses to criticism are similar between female and male managers (left panel). While the positive effect of criticism on the transcription score is slightly larger for male managers, female criticism has slightly larger effects on effort in voluntary tasks, but neither of these differences are close to being statistically significant (Table A4).

The right panel shows that people tend to respond more positively to praise from female managers, although these differences are not statistically significant (p-value=0.15). These results are driven by responses of male workers, but worker gender differences are also not significant (results available on request).

5 Mechanism

The previous section showed that workers do not discriminate against female managers with regard to effort or attitudes. These findings may be surprising as they differ from results in the U.S. (Abel, 2019) and from other evidence out of India, which suggests gender discrimination poses a barrier to the career advancement of women (Batra and Reio Jr, 2016; Giri, 2014; Zimmermann, 2012). In this section, we take advantage of the rich data we collect to better understand the underlying mechanisms that may explain why our results

Figure 3: Feedback and Attitude, Manager Gender



Notes: This graph shows the effect of negative (left panel) and positive feedback (right panel) participants working under male and female managers on the standardized attitude outcome. This attitude outcome is normalized so that zero presents the attitude of a worker assigned to male managers who does not receive feedback. P-values are reported for a test of equal means across feedback assignment (P-value) and to test if the effect of feedback differs by manager gender (Diff P-v). 90% confidence intervals are displayed.

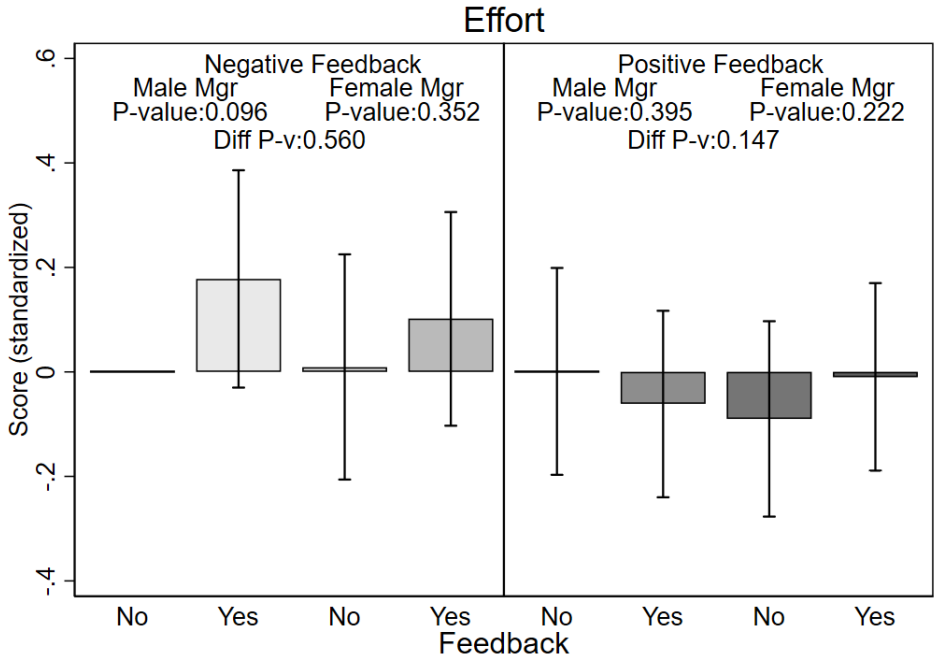
differ from other studies. As specified in the pre-analysis plan, our design tests for the role of attention discrimination (5.1), implicit bias (5.2), and gendered expectations (5.3).

5.1 Attention Discrimination

It has long been proposed that women are more likely to be ignored when trying to influence people (Carli, 2001). This would provide a significant challenge for managers to lead and change subordinates' behavior. Evidence on the role of attention discrimination is scarce due to data availability limitations.[11] In this study, we measure the exact time that workers spend reading and thinking about managers' feedback before continuing to the transcription task, which provides a unique opportunity to test the attention discrimination hypothesis.

[11]One notable exception is Bartos et al. (2016) who find attention discrimination towards immigrants in the rental market in Germany.

Figure 4: Effort and Feedback Content



Notes: This graph shows the effect of negative (left panel) and positive feedback (right panel) participants working under male and female managers on the standardized transcription score. This score is normalized so that zero presents the score of workers assigned to male managers who do not receive feedback. P-values are reported for a test of equal means across feedback assignment (P-value) and to test if the effect of feedback differs by manager gender (Diff P-v). 90% confidence intervals are displayed.

We do not find evidence for attention discrimination against women in our context. Table 5 shows that in aggregate, workers spend 0.3 seconds *longer* on feedback received from female managers (Col. 1). This difference is statistically insignificant and small compared to the control mean of 19 seconds. The amount of attention female managers receive also does not differ by workers gender (Col. 2) nor by the content of the feedback (Col. 3). The same results hold when we look at the role of feedback content separately by worker gender (Col. 4, 5).

5.2 Implicit Gender Bias

After the end of the transcription task and survey, we offer workers a small bonus for completing a "Gender-Career" implitic bias test (IAT).[12] The IAT measures the speed with

[12]Participants were randomized to complete either the gender or caste IAT. About 70% agreed to participate, resulting in a sample of about 500 who completed the gender IAT.

Table 5: Attention to Feedback

	(1)	Gender (2)	Content (3)	Fem Wkr (4)	Male Wkr (5)
Manager female	0.302	0.020	0.270	1.959	-0.496
	(0.730)	(0.854)	(1.111)	(2.003)	(1.335)
Worker female	-0.717	-1.223	-0.755		
	(0.820)	(1.133)	(0.813)		
Fem Mgr x Fem Wkr		1.045			
		(1.641)			
Above Average			3.180***	4.560**	2.550*
			(1.106)	(1.939)	(1.346)
Above Av x Fem Mgr			-0.146	-1.834	0.647
			(1.465)	(2.762)	(1.737)
Observations	1037	1037	1037	282	755
Sample Mean	19.10	19.10	19.10	19.10	19.10
Std Dev	12.03	12.03	12.03	12.03	12.03

Notes: The dependent variable for all columns is the time respondents spend reading the feedback from their manager (in seconds). All estimations are OLS. Robust standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

which someone associates male names with professional terms (e.g. "office," "manager," and "salary") and female names with family terms ("marriage", "home" and "children").[13] The difference in speed is transformed to a standardized score ranging between 1 and -1 where 0 indicates no bias. Slight, moderate, and strong biases correspond to scores of 0.15, 0.3, and 0.6, respectively.

The first surprising evidence is that levels of gender bias among Indian gig economy workers are substantially lower than among MTurk workers recruited from the U.S.. The average score in our Indian sample is 0.2 compared to 0.33 in the U.S. sample (Abel, 2019). A score of 0.2 indicates relatively low levels of implicit bias with regard to women and career.

Table 6 shows how workers' implicit gender bias is correlated with the effort and attitudes and how this relationship varies with the gender of their manager. Overall, more biased workers perform more poorly and have worse attitudes. However, this negative relationship is *reduced* if people are working under a female manager (Col 1, 5). Next we look at the role

[13]There is an active debate to what extent implicit biases are important drivers of behavior. One side claims that IATs are poor predictors of actual behavior (Oswald et al., 2013). The other side argues that implicit biases are far more common and predictive than explicit biases (Greenwald et al., 2015; Grossman et al., 2016). For example, Reuben et al. (2014) find that employers' IAT scores predict biased updating of expectations upon receiving performance information. Along the same lines, Glover et al. (2019) find that implicit biases among managers lower the job performance of minority workers.

of implicit bias in workers' reactions to feedback. The positive relationship between workers' gender bias and their performance when paired with a female managers is particularly strong when they receive feedback (Col. 2, 6). For effort, this relationship seems to be driven by the effects of negative feedback (Col. 3, 4). For attitudes, the content of the feedback does not matter (Col. 7, 8).

Overall, there is no evidence that implicit gender biases play an important role in the worker manager relationship in our setting. If at all, more biased workers tend to react more *positively* if paired with a female manager. One explanation is that the gender IAT measures people's association of women and career, and is thus less important once workers see women in a position of professional leadership. It is also possible that biased workers believe that female managers needed to overcome more barriers and are thus particularly qualified.

Table 6: Implicit Bias

	Transcrip. Score				Attitude Index			
	Aggreg. (1)	Feedb. (2)	Pos FB (3)	Neg FB (4)	Aggreg. (5)	Feedb. (6)	Pos FB (7)	Neg FB (8)
Gender IAT	-1.180**	-0.224	-0.361	-1.386	-0.156	0.015	-0.108	0.096
	(0.529)	(0.695)	(0.546)	(1.235)	(0.120)	(0.202)	(0.193)	(0.383)
Fem Mgr x IAT	0.698	-0.687	0.041	-0.730	0.246	0.075	0.268	-0.174
	(0.679)	(0.914)	(0.781)	(1.667)	(0.152)	(0.239)	(0.250)	(0.433)
IAT x FB		-1.683*	-1.098	-1.365		-0.289	-0.207	-0.277
		(0.997)	(1.079)	(1.644)		(0.245)	(0.253)	(0.457)
Fem Mgr x FB x IAT		2.442*	0.618	3.823*		0.290	0.001	0.612
		(1.307)	(1.304)	(2.311)		(0.306)	(0.334)	(0.543)
Observations	486	486	295	191	486	486	295	191
Sample Mean	13.03	13.03	13.03	13.03	0.01	0.01	0.01	0.01
Std Dev	2.327	2.327	2.327	2.327	0.707	0.707	0.707	0.707

Notes: The dependent variable in Column (1) through (4) is the transcription score described above. The dependent variable in Column (5) through (8) is the standardized attitude index. Larger values in the IAT variable means that respondents hold a stronger implicit gender bias. Models are fully interacted, but only coefficients of interest are reported for readability reasons. All estimations are OLS. Robust standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

We also collected evidence on the *explicit* beliefs about the qualification of female managers. In an important recent study in neighboring Bangladesh, Macchiavello et al. (2020) show that garment industry workers initially believe that female managers are of lower ability, resulting in lower performance. We collect data on workers' perception of their managers' level of education. 64.2% of workers thought their female manager attended university. This figure is very similar to the 62.4% of workers who thought their male manager attended university, suggesting that there are no gender differences in perceived manager ability.[14]

[14]This results contrasts with Macchiavello et al. (2020) who find that garment factory workers in

5.3 Gendered Expectations and the Perception of Feedback

People tend to react negatively when their expectations are violated.[15] This also applies to workers' reaction to feedback that contradict expectations about their own performance (Gjedrem, 2018; Kuhnen and Tymula, 2012). This is important in the context of gender discrimination, as numerous studies document that people hold gendered expectations. Women are seen to be more agreeable, altruistic, warm, compliant, modest, and sympathetic than men (Blau and Kahn, 2017; Bertrand, 2011; Eagly et al., 1995). Abel (2019) concludes that strong gendered expectations among gig economy workers in the U.S. can explain gender discrimination against critical female managers, confirming that gendered expectations have important implications for the effectiveness of female leaders (Eagly et al., 1995).

Table 7: Management Style Expectations

	N	Female	Male	Either	M/F Ratio
Strict Expectation	141	.33	.33	.34	1
Give Criticism	141	.22	.35	.43	1.59
Give Praise	141	.26	.29	.45	1.11
Appropriate Tone	141	.36	.23	.40	0.64
Communicate Expectations	141	.30	.24	.46	0.80
Give Feedback	141	.29	.26	.45	0.90

Notes: The table present results from an out of sample survey asking people whether they associate different management styles with female or male managers or either gender. M/F reported the response ratio of male versus female association.

Most of the evidence on gendered expectations comes from developed countries. We elicit associations of gender and management styles through an out-of-sample survey with 141 Indian MTurk participants. Table 7 shows the share of respondents that associate certain management styles either with women, men or either. Interestingly, we find that for many traits, people do not hold gendered expectations. The share associating female and male managers with strict expectations are identical. Likewise associations for giving praise and giving feedback are very similar. For other traits, we see somewhat stronger gendered expectations, although these are still much weaker than among gig workers in the U.S. (Abel, 2019).

Weaker gendered expectations could thus be one of the reasons for why we see less gender discrimination in the Indian context. This is also reflected by how workers interpret the feed-

Bangladesh falsely belief that female supervisors have less technical knowledge. One explanation is that gender stereotypes about ability are domain specific (Coffman et al., 2019).

[15] It is part of a broader literature showing that people react negatively if their expectations are violated (see Burgoon (1993) for a review).

back they receive. In the U.S., (male) workers were more likely to discard critical feedback from female managers as inaccurate. We do not observe the same pattern in India. The accuracy and appropriateness of feedback is very similar with a manager gender difference of less than 0.01 s.d. (Table A1, Col. 1, 5). Gender effects for accuracy and appropriateness are slightly larger for criticism than for praise, but the overall magnitude is still small at 0.05 s.d. (Table A1, Col. 2, 6). There are no significant differences or consistent patterns in the perception of feedback by worker gender (Col. 3, 4, 7, 8).

One may be concerned that participants recruited via MTurk may change their behavior as they may suspect they are participating in research.[16] While we cannot rule out these Hawthorne effects, it is reassuring that even after being prompted, only 20% of participants expected the task to be related to research. Importantly, when asked about the exact nature of the research questions, none of the participants thought it was related to gender or the identity of the manager.

6 Discussion

6.1 Comparison: India vs. USA

Our study provides some of the first experimental evidence on the effect of manager gender and feedback in a developing country context. It is, to our knowledge, also the first experimental replication on the effect of feedback across countries. Figure 5 compares our results to Abel (2019), who employs the same study design with workers in the U.S. gig economy.

The pooled results show that praise has no effect on attitude or effort in either the United States or India (bottom panel, Figure 5). By contrast, the negative effect of criticism on attitude is three times as large in the U.S. sample (0.45 s.d.) than the Indian counterpart (0.15 s.d.). Moreover, criticism decreases effort among workers in the U.S. sample (0.1 s.d.) but increases it among Indian counterparts (0.17 s.d.).[17]

In other words, not only do the Indian workers in our sample react less poorly to critical feedback, our findings suggest they also increase their effort provision in response to being criticized. This is consistent with research on the effects of India's high status identity, high power distance culture on workplace dynamics. This increased sense of hierarchy can result in greater inherent expectations of top-down feedback delivery (Sully De Luque and Sommer, 2000) and respect for paternalistic authority (Amba-Rao et al., 2000). Additionally, Kwantes

[16]MTurk is increasingly used in the U.S. as a platform to recruit participants for research. We are aware of fewer MTurk-based research studies in developing countries.

[17]Both of these cross-country differences are sizable in magnitude at around 0.3 s.d. and statistically significant at the 5% level.

(2009) suggests that loyalty plays a much bigger role in Indian workers' commitment to their organizations, which implies effort provision and attitudes towards supervisors are driven by a sense of obligation rather than a subjective evaluation of their job satisfaction.

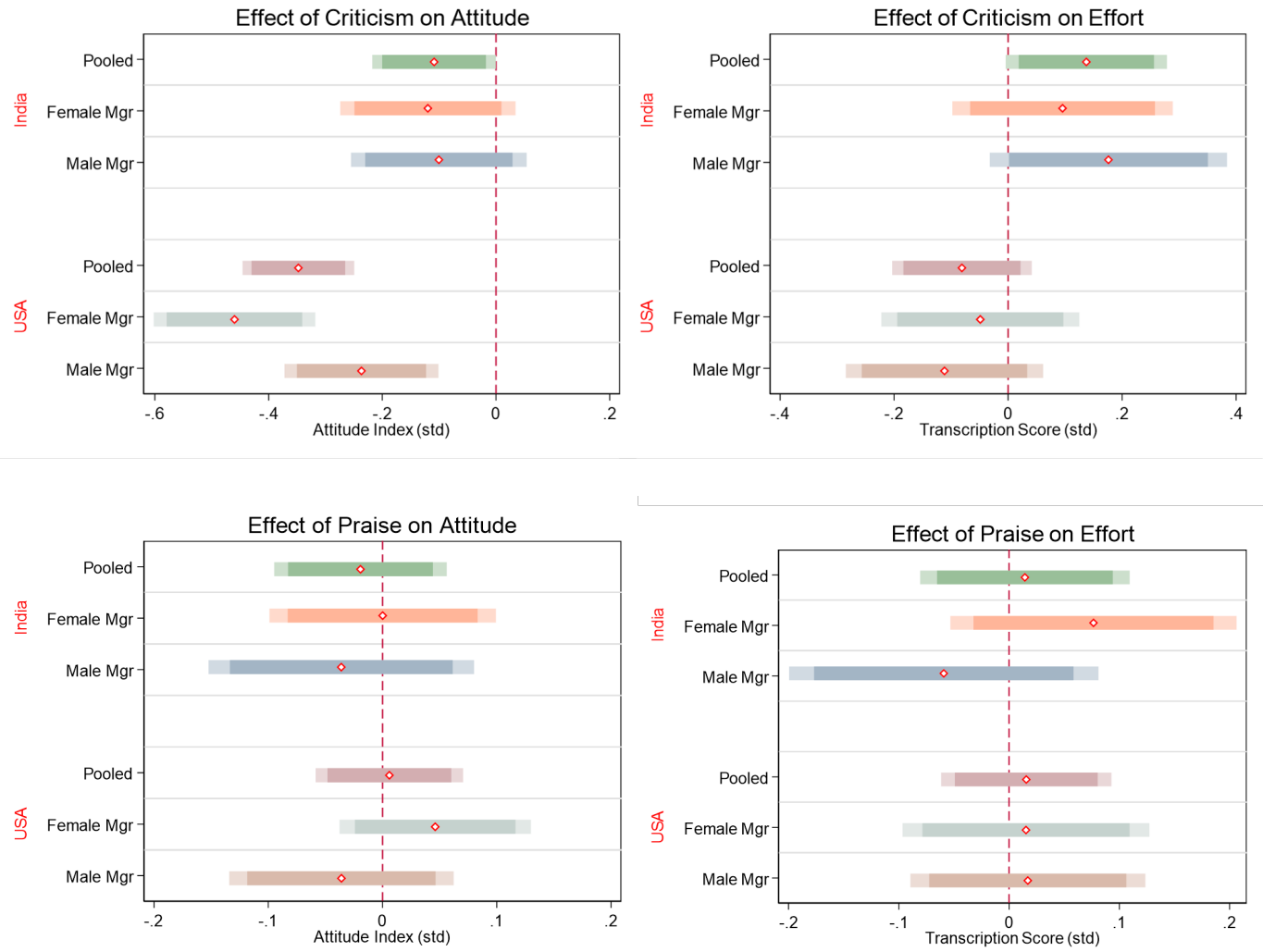
In our study, the increased acceptance of hierarchy is reflected in Indian workers' perception of feedback. Figure 6 shows the perceived accuracy of feedback as a function of workers' performance in India (left) and the United States (right). Specifically, we measure how much people agree with the statement that the feedback they received was accurate and record responses as 0 for "strongly disagree" through 6 for "strongly agree". Praise is perceived as highly accurate in both countries, as depicted by the high levels of perceived feedback accuracy to the right of the average threshold. In both contexts, criticism is perceived as less accurate than praise. Levels of agreement with criticism, however, vary substantially. On average U.S. workers "somewhat disagree" with the statement, while Indian workers on average "somewhat agree" or "mostly agree" that the criticism they received was accurate.

The second notable cross-country difference is in the role of manager gender. Criticism has a small negative effect (0.12 s.d.) on attitudes in India, regardless of manager gender. By contrast, in the U.S. the effect of negative feedback is twice as large if it comes from a female manager (Figure 5, top left panel). Consistent with this pattern, we observe that Indian workers do not assess (negative) feedback to be less accurate if it comes from a female manager, which contrasts with results from the U.S. sample, where workers think criticism is less accurate when it comes from a female manager.

Across both countries we do not find gender discrimination in response to praise. U.S. workers react slightly more positively to praise from women with respect to attitudes, but there is no differential response in effort provision. By contrast, Indian workers tend to increase effort more if praised by a female manager compared to a male manager, but these differences are not statistically significant (Figure 5, bottom panels).

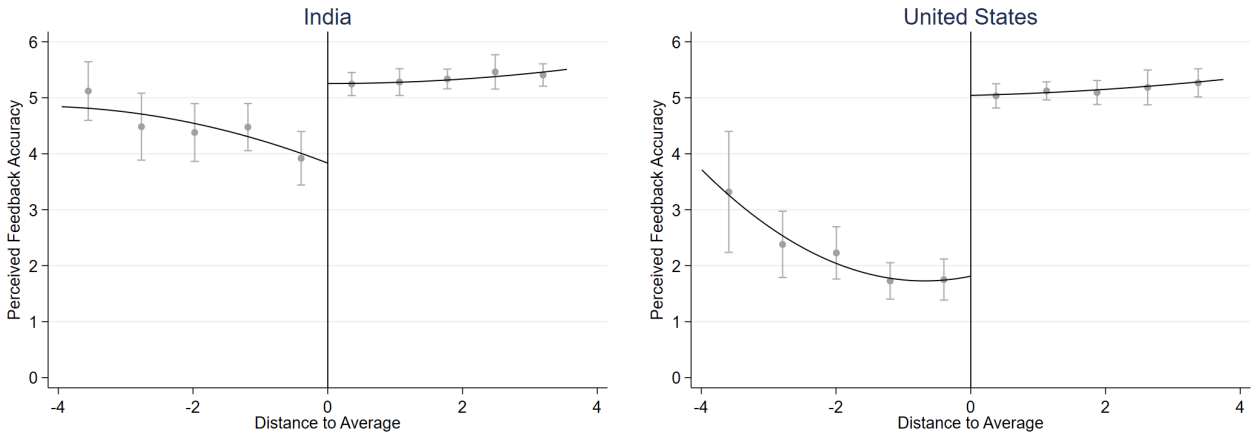
Our discussion of the underlying mechanisms highlights that gender discrimination and other aspects of workplace culture are highly context-specific. One's culture and belief system directly inform the gendered expectations one holds; one's upbringing and environment can change the expectations one has of a particular gender, and one's learned habits may be directly responsible for one's likelihood to pay attention to or ignore a person of a particular gender. The benefit of replicating study designs with comparable samples across countries is that it can help to identify context-specific factors that determine how workers respond to various treatments. This is particularly useful in studying culturally variable phenomena like gender-based discrimination from below.

Figure 5: Effect of Criticism: India vs. USA



Notes: The figure reports the perceived accuracy of feedback in India (left panel) and the U.S. (right panel). The x axis measures the transcription score. Feedback to the left (right) of the average is negative (positive). The y axis measures whether workers say that the feedback was accurate, ranging from 0=strongly disagree to 6=strongly agree."

Figure 6: Perception of Feedback: India vs. United States



Notes: The figure reports effects of criticism (top panels) and praise (bottom panels) on attitude (left panels) and effort (right panels) for both criticism (top panels) and praise (bottom panel). Each point estimate reports differences in responses relative to the group of workers that did not receive feedback. Results are presented for both the pooled sample and separately by manager gender for both India and the U.S.. The shades of the bars present 90% and 95% confidence intervals.

6.2 Treatment Heterogeneity and External Validity

The specific work arrangement we are investigating, remote work in the gig economy, is growing in popularity. Particularly in the wake of the COVID-19 pandemic, Indian workers, who had long expressed a growing interest in telecommuting, are likely to pursue more remote work arrangements as well as flexible gig-based work (Basu, 2016; Ipsos, 2011). Few studies have explored the unique nature of the supervisor-subordinate relationship in remote work arrangements in India or the other parts of the Global South more broadly.

While studying this growing sector is important in its own right, it is also important to acknowledge that our sample is not representative of the Indian workforce. Gig economy workers are disproportionately young, male and university-educated. To test the extent to which this selection affects the generalizability of our findings, Figure 7 shows how our results vary across relevant subgroups. We find the effect of criticism and praise on attitudes is consistent across all groups, and none of the differences in attitude effects exceed 0.1 s.d. (left panels, Figure 7). There is slightly more variation in the effect of negative feedback on effort. Most notably, older respondents react more negatively to criticism from female managers. A similar pattern holds for members of upper castes, although given the relatively small sample size, this effect is estimated imprecisely.

Another concern is that our sample is predominantly drawn from India's south. Nearly 65% of workers in our sample reside in either Tamil Nadu or Kerala, reflecting that most of India's IT industry is located in the south. Traditionally, these states tend to be more gender-

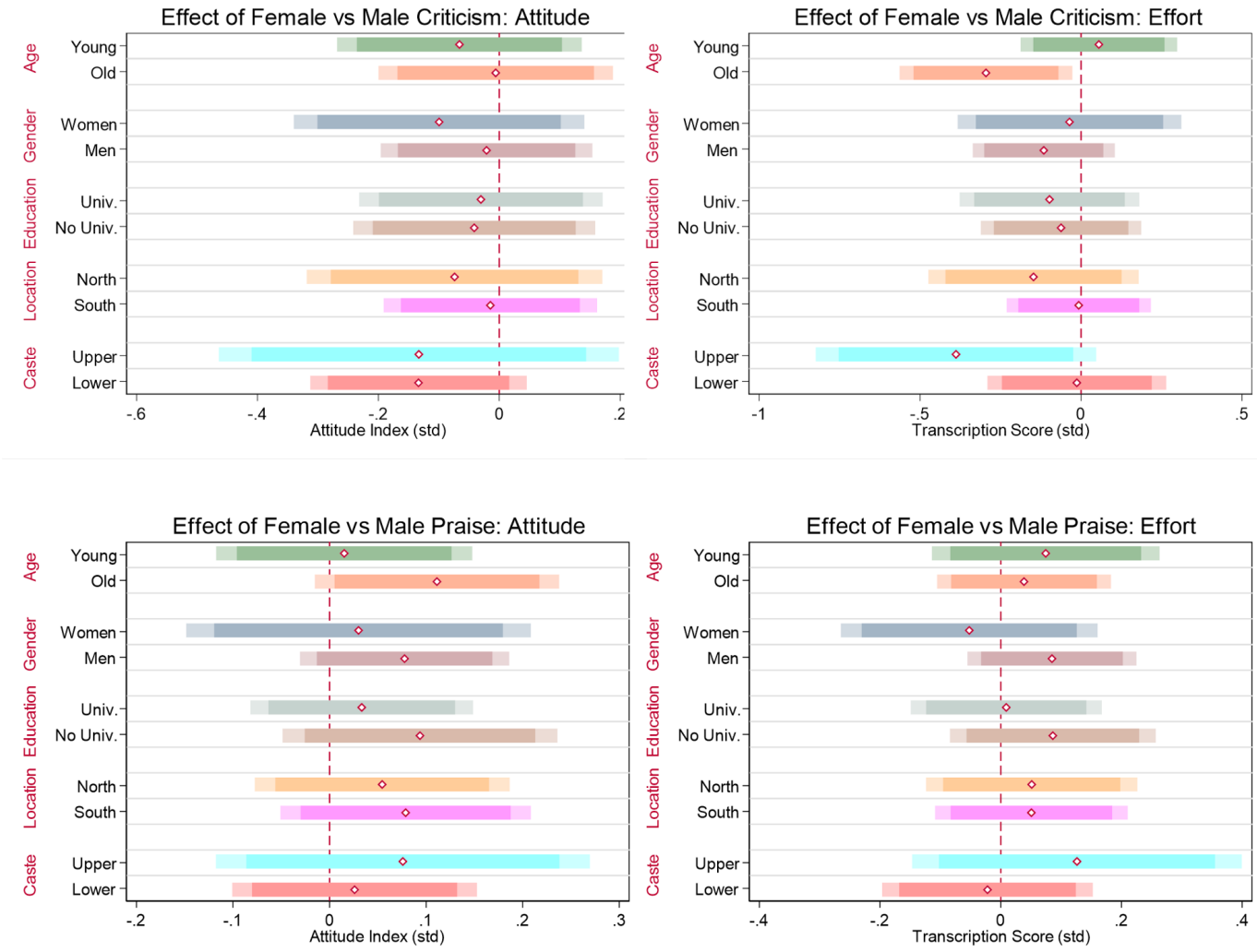
progressive, have fewer gender disparities in education, and boast higher levels of female labor force participation and female autonomy (Rammohan and Vu, 2018). Comparing results between workers in the south and other parts of India indeed shows that there tends to be less gender discrimination in the south. These differences, however, are relatively small and not statistically significant.[18]

Overall, these results suggest there is relatively little variation across subgroups, which partially allays concerns over the generalizability of our results. Nevertheless, we want to reiterate that we do not make any wider claims about whether results generalize to other parts of India's economy. Indeed, many studies show that gendered expectations and gender discrimination is still prevalent in Indian society (Ranganathan and Shivaramb, 2019; Batra and Reio Jr, 2016; Giri, 2014; Barcellos et al., 2014; Zimmermann, 2012).

There is, however, evidence that gender discrimination may be on the decline in India. Abbas (2015) finds that gender expectations in India have been converging over time, and Beaman et al. (2012) show that the increase of women serving as village leaders led to a reduction in gender-biased attitudes and increase in support for female political leaders. By contrast, there is little evidence on the interplay of hierarchical norms and traditional gender expectations in India's private sector. Our study presents some initial evidence to suggest that deference to hierarchy may play a role in mitigating the effect of traditional gender norms on workplace interactions.

[18]Our results remain consistent when we include participants from Karnataka, Telangana, and Andhra Pradesh in the sample of southern workers.

Figure 7: India: Treatment Heterogeneity



Notes: The figure reports heterogeneous effects of having a female manager on attitude (left panels) and effort (right panels) for both criticism (top panels) and praise (bottom panel). Each point estimate reports differences in responses to female compared to male managers within that specific subgroup. The shades of the bars present 90% and 95% confidence intervals.

References

Abbas, S. (2015). Indian organizations and attitude towards women managers: Gender differential, sex bias, reaction and perceptive reality. *Yobe Journal of Economics 2*, 117–128.

Abel, M. (2019). Do workers discriminate against female bosses? *IZA Discussion Paper*.

Amba-Rao, S. C., J. A. Petrick, J. N. Gupta, and T. J. V. D. Embse (2000). Comparative performance appraisal practices and management values among foreign and domestic firms in india. *International Journal of Human Resource Management 11*(1), 60–89.

Ayalew, S., S. Manian, and K. Sheth (2018). Discrimination from below: Experimental evidence on female leadership in ethiopia. WPS 079, Center for Effective Global Action.

Azmat, G. and N. Iriberri (2010). The importance of relative performance feedback information: Evidence from a natural experiment using high school students. *Journal of Public Economics 94*(7-8), 435–452.

Bandiera, O., V. Larcinese, and I. Rasul (2015). Blissful ignorance? a natural experiment on the effect of feedback on students' performance. *Labour Economics 34*, 13–25.

Barcellos, S. H., L. S. Carvalho, and A. Lleras-Muney (2014). Child gender and parental investments in india: Are boys and girls treated differently? *American Economic Journal: Applied Economics 6*(1), 157–89.

Bartos, V., M. Bauer, J. Chytilová, and F. Matějka (2016). Attention discrimination: Theory and field experiments with monitoring information acquisition. *American Economic Review 106*(6), 1437–1475.

Basu, S. (2016). One in two indian employees prefer telecommuting, reveals randstad survey. *The Economic Times*.

Batra, R. and T. G. Reio Jr (2016). Gender inequality issues in india. *Advances in Developing Human Resources 18*(1), 88–101.

Beaman, L., E. Duflo, R. Pande, and P. Topalova (2012). Female leadership raises aspirations and educational attainment for girls: A policy experiment in india. *science 335*(6068), 582–586.

Bertrand, M. (2011). New perspectives on gender. In *Handbook of Labor Economics*, Volume 4, pp. 1543–1590. Elsevier.

Blanes i Vidal, J. and M. Nossol (2011). Tournaments without prizes: Evidence from personnel records. *Management Science 57*(10), 1721–1736.

Blau, F. D. and L. M. Kahn (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature 55*(3), 789–865.

Bloom, N., B. Eifert, A. Mahajan, D. McKenzie, and J. Roberts (2013). Does management matter? evidence from india. *The Quarterly Journal of Economics 128*(1), 1–51.

Bloom, N., R. Sadun, and J. Van Reenen (2012). The organization of firms across countries. *The quarterly journal of economics 127*(4), 1663–1705.

Boring, A. (2016). Gender biases in student evaluations of teaching. *Journal of Public Economics 145*, 27–41.

Bracken, C. C., L. W. Jeffres, and K. A. Neuendorf (2004). Criticism or praise? the impact of verbal versus text-only computer feedback on social presence, intrinsic motivation, and recall. *Cyberpsychology & behavior 7*(3), 349–357.

Budhwar, P. S., D. S. Saini, and J. Bhatnagar (2005). Women in management in the new economic environment: The case of india. *Asia Pacific Business Review 11*(2), 179–193.

Burgoon, J. K. (1993). Interpersonal expectations, expectancy violations, and emotional communication. *Journal of Language and Social Psychology 12*(1-2), 30–48.

Buser, W., J. Hayter, and E. C. Marshall (2019). Gender bias and temporal effects in standard evaluations of teaching. *AEA Papers and Proceedings 109*, 261–265.

Carli, L. L. (2001). Gender and social influence. *Journal of Social Issues 57*(4), 725–741.

Catalyst (2019). Quick take: Women in the workforce – india. Technical report.

Cecchi-Dimeglio, P. (2017). How gender bias corrupts performance reviews, and what to do about it. *Harvard Business Review*.

Charness, G., D. Masclet, and M. C. Villeval (2014). The dark side of competition for status. *Management Science 60*, 38–55.

Coffman, K. B., C. B. Flikkema, and O. Shurchkov (2019). *Gender Stereotypes in Deliberation and team decisions*. Harvard Business School.

Correll, S. and C. Simard (2016). Vague feedback is holding women back. *Harvard Business Review*.

DellaVigna, S. and D. Pope (2018). What motivates effort? evidence and expert forecasts. *Review of Economic Studies 85*(2), 1029–1069.

Dhar, D., T. Jain, and S. Jayachandran (2018). Reshaping adolescents' gender attitudes: Evidence from a school-based experiment in india. Technical report, National Bureau of Economic Research.

Eagly, A. H., S. J. Karau, and M. G. Makhijani (1995). Gender and the effectiveness of leaders: A meta-analysis. *Psychological Bulletin 117*(1), 125.

Eriksson, T., A. Poulsen, and M. C. Villeval (2009). Feedback and incentives: Experimental evidence. *Labour Economics 16*(6), 679–688.

Fryer Jr, R. G. and S. D. Levitt (2004). The causes and consequences of distinctively black names. *The Quarterly Journal of Economics 119*(3), 767–805.

Giri, A. V. (2014). Women entrepreneurs from india: Problems, motivations and success factors. *International Journal of Innovative Research and Practices 2*(1).

Gjedrem, W. G. (2018). Relative performance feedback: Effective or dismaying? *Journal of Behavioral and Experimental Economics 74*, 1–16.

Glover, D., A. Pallais, and W. Pariente (2019). Discrimination as a self-fulfilling prophecy: Evidence from french grocery stores. *The Quarterly Journal of Economics 132*(3), 1219–1260.

Goldin, C. (2014). A grand gender convergence: Its last chapter. *American Economic Review 104*(4), 1091–1119.

Greenwald, A. G., M. R. Banaji, and B. A. Nosek (2015). Statistically small effects of the implicit association test can have societally large effects.

Grossman, P. J., C. Eckel, M. Komai, and W. Zhan (2016). It pays to be a man: Rewards for leaders in a coordination game. Monash Economics Working Papers 38-16, Monash University, Department of Economics.

Gupta, A., M. Koshal, and R. Koshal (1998). Women managers in india: challenges and opportunities. equal opportunities international. *Equal Opportunities International 17*(8), 14–26.

Hoffman, M., L. B. Kahn, and D. Li (2017). Discretion in hiring. *The Quarterly Journal of Economics 133*(2), 765–800.

Holmstrom, B. and P. Milgrom (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *JL Econ. & Org. 7*, 24.

Hsieh, C.-T., E. Hurst, C. I. Jones, and P. J. Klenow (2019). The allocation of talent and us economic growth. *Econometrica 87*(5), 1439–1474.

Ipsos (2011). Telecommuting: Citizens in 24 countries assess working remotely for a total global perspective.

Jayachandran, S. and R. Pande (2017). Why are indian children so short? the role of birth order and son preference. *American Economic Review 107*(9), 2600–2629.

Kling, J. R., J. B. Liebman, and L. F. Katz (2007). Experimental analysis of neighborhood effects. *Econometrica 75*(1), 83–119.

Kuhnen, C. M. and A. Tymula (2012). Feedback, self-esteem, and performance in organizations. *Management Science 58*(1), 94–113.

Kwantes, C. T. (2009). Culture, job satisfaction and organizational commitment in india and the united states. *Journal of Indian Business Research*.

List, J. A. and F. Momeni (2017). When corporate social responsibility backfires: Theory and evidence from a natural field experiment. Working Paper 24169, National Bureau of Economic Research.

Lockwood, N. R. (2009). Perspectives on women in management in india.

Macchiavello, R., A. Menzel, A. Rabbani, and C. Woodruff (2020). Challenges of change: An experiment promoting women to managerial roles in the bangladeshi garment sector. Technical report, National Bureau of Economic Research.

Manne, K. (2017). *Down girl: The logic of misogyny*. Oxford University Press.

McKenzie, D. and C. Woodruff (2015). *Business practices in small firms in developing countries*. The World Bank.

McKinsey (2018). The power of parity: Advancing women's equality in asia pacific.

Mengel, F., J. Sauermann, and U. Zölitz (2019). Gender bias in teaching evaluations. *Journal of the European Economic Association 17*(2), 535–566.

Niederle, M. and L. Vesterlund (2007). Do women shy away from competition? do men compete too much? *The quarterly journal of economics 122*(3), 1067–1101.

NobleHouse (2019). Gig workforce: The future of work is anywhere.

Oswald, F. L., G. Mitchell, H. Blanton, J. Jaccard, and P. E. Tetlock (2013). Predicting ethnic and racial discrimination: A meta-analysis of iat criterion studies. *Journal of personality and social psychology 105*(2), 171.

Podsakoff, P. M., S. B. MacKenzie, R. H. Moorman, and R. Fetter (1990). Transformational leader behaviors and their effects on followers' trust in leader, satisfaction, and organizational citizenship behaviors. *The leadership quarterly 1*(2), 107–142.

Rammohan, A. and P. Vu (2018). Gender inequality in education and kinship norms in india. *Feminist Economics 24*(1), 142–167.

Ranganathan, A. and R. Shivaramb (2019). Getting their hands dirty: How female managers motivate female worker productivity through subordinate scut work. Technical report, Working Paper.

Reuben, E., P. Sapienza, and L. Zingales (2014). How stereotypes impair women's careers in science. *Proceedings of the National Academy of Sciences 111*(12), 4403–4408.

Sully De Luque, M. F. and S. M. Sommer (2000). The impact of culture on feedback-seeking behavior: An integrated model and propositions. *Academy of Management Review 25*(4), 829–849.

Varma, A., S. Pichler, and E. S. Srinivas (2005). The role of interpersonal affect in performance appraisal: evidence from two samples–the us and india. *The International Journal of Human Resource Management 16*(11), 2029–2044.

Zimmermann, L. (2012). Reconsidering gender bias in intrahousehold allocation in india. *Journal of Development Studies 48*(1), 151–163.

A Appendix

A.1 Tables

Table A1: Feedback effects on attitude by worker gender

	Work Future		Satisfaction		Task Import		Attitude Index	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Female Workers								
1=receive feedback	-0.003	0.010	-0.006	-0.128	-0.052	-0.106	-0.013	-0.051
	(0.018)	(0.029)	(0.082)	(0.129)	(0.097)	(0.145)	(0.061)	(0.096)
Above x FB		-0.028		0.182		0.066		0.038
		(0.037)		(0.161)		(0.191)		(0.123)
AboveAverage		0.025		0.340**		0.263		0.253***
		(0.029)		(0.136)		(0.161)		(0.097)
Observations	466	466	466	466	466	466	466	466
Panel B: Male Workers								
1=receive feedback	-0.012	-0.031**	-0.051	-0.120	-0.036	-0.128	-0.054	-0.135**
	(0.010)	(0.016)	(0.056)	(0.091)	(0.063)	(0.095)	(0.041)	(0.067)
Above x FB		0.036*		0.088		0.150		0.130
		(0.021)		(0.108)		(0.124)		(0.081)
AboveAverage		-0.000		0.517***		0.271***		0.276***
		(0.015)		(0.087)		(0.103)		(0.063)
Observations	1281	1281	1281	1281	1281	1281	1281	1281
Sample Mean	0.96	0.96	5.27	5.27	5.16	5.16	-0.00	-0.00
Std Dev	0.19	0.19	0.97	0.97	1.08	1.08	0.73	0.73
P-v: FB: M=F	0.68		0.64		0.89		0.58	
P-v: Neg FB: M=F		0.20		0.96		0.90		0.47
P-v: Pos FB: M=F		0.39		0.44		0.68		0.93

Notes: Work future is a binary variable whether worker are interested to work for the firm in the future. *Job satisf.* and *Task Import* measure how strongly workers agree that they were satisfied with the task and that the task was important, respectively (0=strongly disagree, 6= strongly agree). *Index* is a standardized index of the other three outcome measures. P-values report tests whether the effect of positive feedback is different from zero. All estimations are OLS. Robust standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A2: Feedback effects on effort by worker gender

	Legible		Adding		Transcrip Score		Voluntary Task	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Female Workers								
1=receive feedback	0.033	0.038	0.014	0.020	0.580	0.369	0.037	0.069
	(0.026)	(0.041)	(0.035)	(0.049)	(0.415)	(0.610)	(0.039)	(0.057)
Above x FB		-0.035		-0.042		-0.098		-0.067
		(0.049)		(0.066)		(0.762)		(0.078)
Observations	466	466	466	466	466	466	434	434
Panel B: Male Workers								
1=receive feedback	0.042**	0.056**	0.034	0.020	0.679**	0.759*	0.042*	0.043
	(0.017)	(0.026)	(0.022)	(0.030)	(0.280)	(0.392)	(0.023)	(0.035)
Above x FB		-0.049		-0.001		-0.568		-0.006
		(0.031)		(0.039)		(0.478)		(0.047)
Observations	1281	1281	1281	1281	1281	1281	1176	1176
Sample Mean	0.68	0.68	0.56	0.56	9.82	9.82	0.82	0.82
Std Dev	0.30	0.30	0.39	0.39	4.98	4.98	0.39	0.39
P-v: FB: M=F	0.78		0.63		0.84		0.91	
P-v: Neg FB: M=F		0.71		0.99		0.59		0.69
P-v: Pos FB: M=F		0.89		0.42		0.88		0.57

Notes: The dependent variable in Column (1) and (2) is whether the worker says the receipt is legible. The dependent variable in Column (3) and (4) measure if the worker is willing to add up the amounts. The dependent variable in Column (5) and (6) captures the accuracy of transcribing receipts Column (7) and (8) measure effects on the voluntary task. All estimations are OLS. Robust standard errors are in parentheses. The p-values reported in the last rows test if responses to female and male managers are different by worker gender. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A3: Effect of Feedback Content on Attitudes by Manager Gender

	Work Future		Satisfaction		Task Import		Att Index	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Manager Female	-0.000	-0.019	-0.035	-0.086	0.020	-0.027	-0.001	-0.051
	(0.013)	(0.019)	(0.074)	(0.119)	(0.085)	(0.126)	(0.053)	(0.083)
Feedback	-0.015	-0.033*	-0.072	-0.132	-0.026	-0.093	-0.057	-0.120
	(0.013)	(0.019)	(0.064)	(0.103)	(0.078)	(0.114)	(0.049)	(0.077)
Mgr Fem x FB	0.011	0.025	0.065	0.012	-0.026	-0.060	0.029	0.012
	(0.018)	(0.028)	(0.092)	(0.149)	(0.105)	(0.159)	(0.068)	(0.110)
Above x FB		0.035		0.071		0.104		0.097
		(0.027)		(0.124)		(0.154)		(0.097)
Fem x FB x Above		-0.031		0.089		0.049		0.021
		(0.037)		(0.179)		(0.209)		(0.135)
Observations	1747	1747	1747	1747	1747	1747	1610	1610
Sample Mean	0.96	0.96	5.29	5.29	5.17	5.17	0.01	0.01
Std Deviation	0.19	0.19	0.93	0.93	1.06	1.06	0.71	0.71

Notes: Work future is a binary variable whether worker are interested to work for the firm in the future. *Job satisf.* and *Task Import* measure how strongly workers agree that they were satisfied with the task and that the task was important, respectively (0=strongly disagree, 6= strongly agree). *Index* is a standardized index of the three outcome measures. All estimations are OLS. Robust standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A4: Effect of Feedback Content on Effort by Worker Gender

	Legible		Adding		Transcrip Score		Voluntary Task	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Manager Female	0.011	0.015	0.004	0.000	0.032	0.008	0.000	0.014
	(0.023)	(0.034)	(0.029)	(0.038)	(0.370)	(0.502)	(0.032)	(0.047)
Feedback	0.042**	0.066**	0.011	0.004	0.631*	0.845*	0.025	0.045
	(0.021)	(0.032)	(0.027)	(0.037)	(0.343)	(0.478)	(0.030)	(0.044)
Mgr Fem x FB	-0.003	-0.029	0.035	0.030	0.058	-0.401	0.033	0.011
	(0.029)	(0.044)	(0.037)	(0.051)	(0.465)	(0.659)	(0.040)	(0.060)
Above x FB		-0.075**		-0.020		-0.943		-0.045
		(0.038)		(0.050)		(0.600)		(0.059)
Fem x FB x Above		0.056		0.015		0.961		0.046
		(0.052)		(0.069)		(0.816)		(0.081)
Observations	1744	1744	1744	1744	1744	1744	1608	1608
Sample Mean	0.67	0.67	0.55	0.55	9.90	9.90	0.81	0.81
Std Dev	0.31	0.31	0.39	0.39	4.84	4.84	0.39	0.39

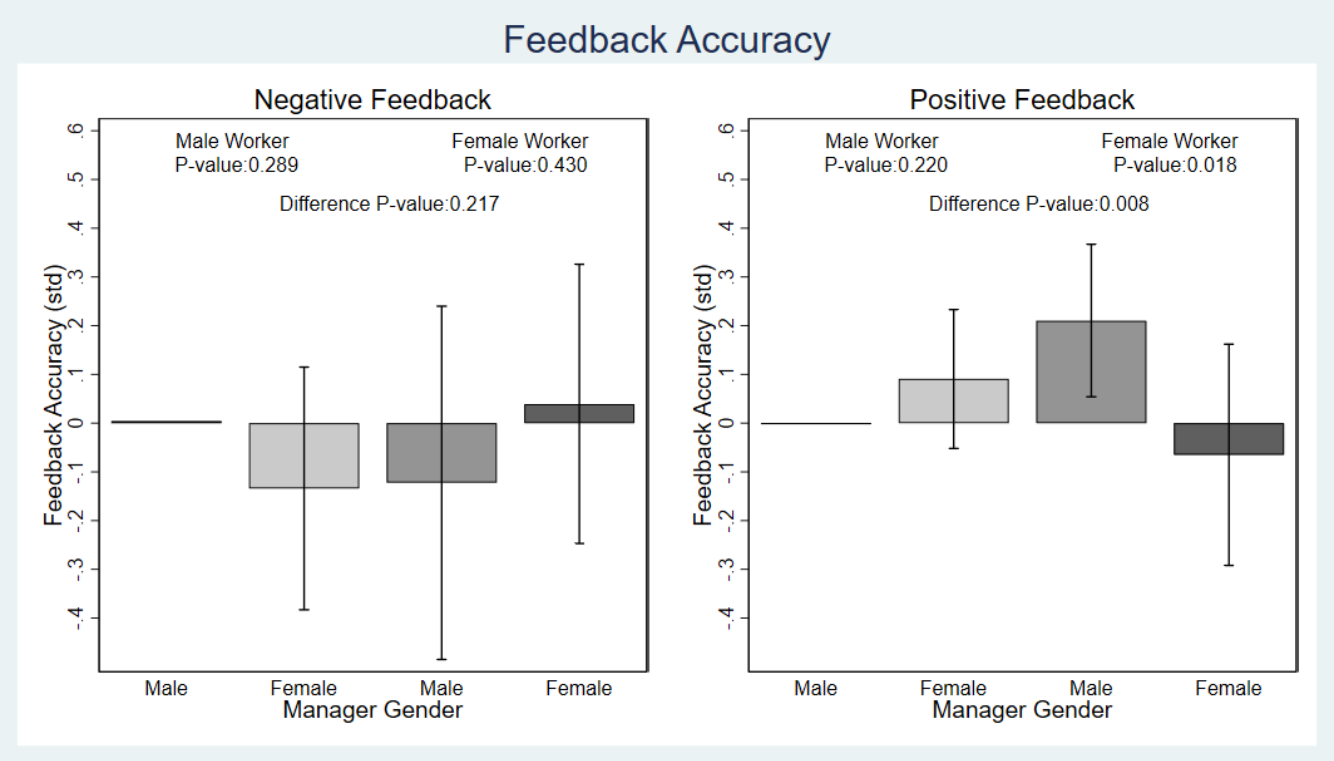
Notes: The dependent variable in Column (1) and (2) is whether the worker says the receipt is legible. The dependent variable in Column (3) and (4) measure if the worker is willing to add up the amounts. The dependent variable in Column (5) and (6) captures the accuracy of transcribing receipts Column (7) and (8) use the score as the dependent variable and divide the sample by worker gender. All estimations are OLS. Robust standard errors are in parentheses. The p-value reported in the last row tests if the sum of the two feedback coefficients are different from zero. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A5: Effect of feedback perception accuracy and appropriateness of Feedback

	Feedback Accurate				Feedback Appropriate			
	Full	Full	Women	Men	Full	Full	Women	Men
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female Manager	-0.011	-0.080	0.102	-0.171	0.003	-0.086	-0.148	-0.087
	(0.082)	(0.143)	(0.262)	(0.168)	(0.076)	(0.133)	(0.258)	(0.154)
Above x Fem Mgr		0.074	-0.475	0.304		0.111	0.013	0.173
		(0.166)	(0.304)	(0.194)		(0.152)	(0.295)	(0.176)
Above Average		0.762***	1.073***	0.617***		0.757***	0.904***	0.680***
		(0.120)	(0.224)	(0.140)		(0.108)	(0.206)	(0.126)
Observations	1037	1037	282	755	1037	1037	282	755
Sample Mean	4.95	4.95	4.91	4.95	5.01	5.01	4.90	5.04
Std Deviation	1.332	1.332	1.232	1.354	1.221	1.221	1.276	1.206
Pos FB=0 (p-v)		0.947	0.021	0.174		0.740	0.339	0.315

Notes: Dependent variables are how much worker agree that the feedback was accurate (col. 1-4) and appropriate (col. 5-8) on a scale from 0 (strongly disagree) to 6 (strongly agree). All estimations are OLS. Robust standard errors are in parentheses. The p-value reported in the last row tests if the effect of positive feedback is different from zero. The table is estimated for participants who received feedback. A programming error prevented collecting data on this variable for the second half of the sample. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure A1: Feedback Accuracy



Notes: Figure A1 shows how workers perceive the accuracy of positive (left panel) and negative (right panel) feedback. Bars show the difference of each combination relative to the assessment of a male worker paired with a male manager. Effects are reported for a standardized measure of feedback accuracy (mean=0, s.d.=1).

B Online Appendix

Figure A2: Introduction: Task Example

Now, please do the following:

1) **List the items** in the order that they appear on the receipt in the text entry fields below. Write the item name exactly as it appears on the receipt

2) For each item, **enter the exact price** that appears on the receipt in the price field. Enter only the number and do not omit any decimal places, even if they are 0. Do not enter a dollar sign, stars, or letters unrelated to the price.

3) If you have a calculator available, **add** the **prices of all the items** in the list and enter the total cost when prompted. This is **not a required task** and will not affect your payment.

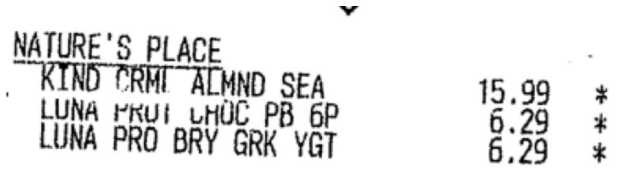
GROCERY

CC PRTY SZ ORIG CHIP	4.59	*
ROLD GOLD TINY'S	2.99	*
HRD COATED PLTS 100C	2.99	A

Example Receipt (Above)

Write "CC PRTY SZ ORIG CHIP" in the first name field and "4.59" in the first price field
Write "ROLD GOLD TINY'S" in the second name field and "2.99" in the second price field
Write "HRD COATED PLTS 100C" in the third name and "2.99" in the third field
Enter "10.57" into the total cost field

Figure A3: Task



Can you read the receipt?
*Selecting **"NO"** will **skip** to the **next receipt**. Please **only select "NO"** if the receipt is **genuinely illegible**!*

- ○ YES
- ○ NO → **Next receipt**

Enter the name and price of each item listed on the receipt. Even if the entry is not clear, please try your best to transcribe what you see. You may consult the instructions here.

What is the total cost of the all the items in the receipt?

Complete this if you have a calculator available. Otherwise, you may leave this field blank.

Item 1 (Name)

Item 1 (Price)

Item 2 (Name)

Figure A4: Survey

You have now **completed all the receipts** in need of transcription.

Please spend **60 seconds** responding to the following **multiple choice questions**. Answers to these sorts of questions help improve communication between managers and MTurkers.

Thank you,

${e://Field/MaleName}

Please enter the name of the manager assigned to you.

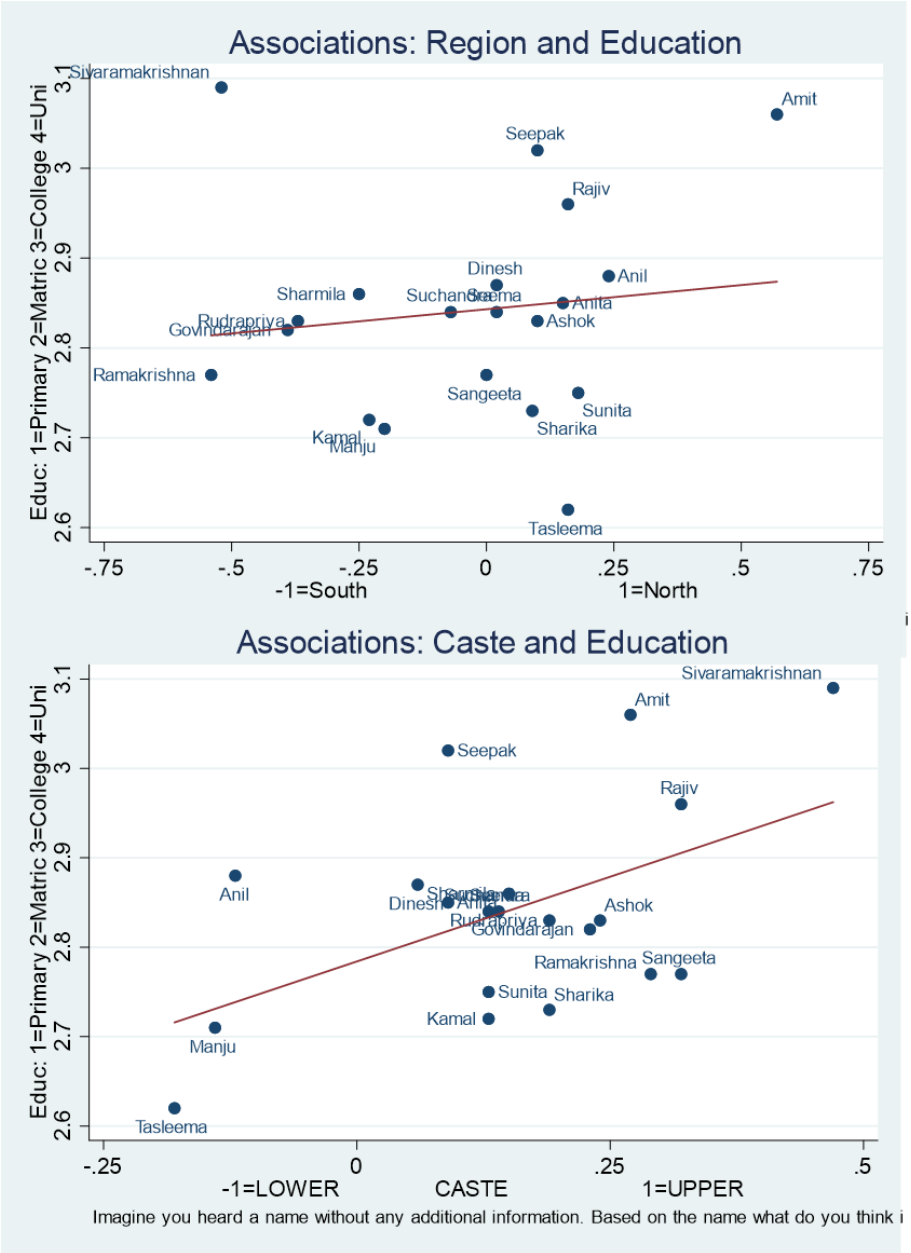
Would you be interested in doing more work for us in the future?

- ○ Yes
- ○ No

Please rate how strongly you agree with the statements below.

	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
I am satisfied with my experience transcribing receipts	○	○	○	○	○	○	○
The task (transcribing receipts) was stressful	○	○	○	○	○	○	○
I was convinced that the task (transcribing receipts) was important	○	○	○	○	○	○	○
The feedback I received was accurate	○	○	○	○	○	○	○
The tone of the feedback was appropriate	○	○	○	○	○	○	○

Figure A5: First Name Associations



Notes: The figure shows associations of first names with education levels, caste, and location. Specifically, we asked people "Without knowing anything else about this person, what do you think is the caste/ education / region of a person with the following name?"