Peer Bargaining and Productivity in Teams: Gender and the Inequitable Division of Pay

Lamar Pierce*
Olin Business School, Washington University in St. Louis, pierce@wustl.edu

Laura Wang*
University of Illinois, Urbana-Champaign, lauraww@illinois.edu

Dennis J. Zhang*
Olin Business School, Washington University in St. Louis, denniszhang@wustl.edu

August 14, 2018

A recent trend in organizational design is to reduce hierarchy and allow employee teams to self-manage tasks, responsibilities, and rewards. Yet we know little about the implications of this arrangement for worker productivity, pay equity, and organizational performance. We provide the first firm-based evidence that when worker teams are allowed to internally allocate compensation, the ensuing peer bargaining process can generate inequitable outcomes for women. We demonstrate this using risk-adjusted fixed-effect models to identify productivity and peer bargaining traits in 965 workers at 32 large Chinese beauty salons. We measure individual productivity through service and card sales and measure bargaining through the division of team-based commissions. We also build a parsimonious bargaining model to explain the mechanisms driving our empirical results. We find that although productivity and bargaining outcomes are positively correlated, female workers consistently receive bargaining outcomes below their productivity level, while men are consistently overcompensated. Importantly, we provide evidence that our results can only be explained by a combination of higher prosociality and lower bargaining power in women. Our findings provide unique organizational evidence on how the delegation of pay authority generates bargaining among peers that might impact firm operations and performance. Furthermore, we provide important evidence that the discriminatory social dynamics observed throughout society are evident in operational designs that delegate decision rights to teams. Managers seeking to implement self-management by peers must anticipate the myriad of productivity, retention, and ethical implications that can result when peer workers bargain over tasks and rewards.

Key words: Productivity, Bargaining, Service Operations, Gender, Negotiation, Fairness, Compensation

---

* The authors are listed alphabetically and contributed equally to the paper. We thank Seth Carnahan, David Harrison, Ashleigh Shelby Rosette, and participants at seminars at Duke, Carnegie Mellon, and Stanford for helpful comments and suggestions.

1. Introduction

Team-based production is essential to modern firms' operations (Huckman et al. 2009, Coff 1997, Wuchty et al. 2007), yet managers face the constant challenge of allocating tasks and rewards to team members when individual ability, effort, and performance are hard to measure (Prendergast 1999). To counter this information problem, managers often rely on team members to allocate tasks (Jung et al. 2017, Huckman and Staats 2011) and assign credit and rewards (Shaw et al. 2001, Bamberger and Levi 2009) among their peers. For example, firms like Verve and Zappos have implemented extreme self-management systems such as "holacracies" intended to replace hierarchy with peer-based management (Bernstein et al. 2016). Ideally, such self-managed teams would efficiently allocate tasks based on relative productivity advantages and equitably assign rewards based on contributions. Recent evidence in academia, however, suggests that this allocation process is inherently inequitable to female team members—female faculty end up with the less-desirable tasks (Babcock et al. 2017) and lower recognition (Sarsons 2017). But does this same inequity emerge in more common operational settings in firms?

We argue that inequity in self-managed teams is a likely outcome because the peer-based allocation of tasks and rewards is fundamentally a "peer bargaining" process, where tasks and earnings are negotiated among coworkers. Peer bargaining, like productivity (Huckman and Pisano 2006, Hamilton et al. 2003, Mas and Moretti 2009, Chan et al. 2014a), represents a persistent and estimable individual trait, and that this trait is based in distributional preferences (Fehr and Schmidt 1999) and bargaining power (Card et al. 2015) that can vary by gender (Bowles and McGinn 2008, Croson and Gneezy 2009). We empirically identify these separate productivity and bargaining traits in a service operations setting, where 965 workers in a chain of 32 large beauty salons in Beijing, China provide salon services and sell prepaid cards over 50 months. We show that when peer workers are allowed to divide their own team-based compensation, women are severely underpaid for their productivity due to both stronger preferences for fairness and lower bargaining power.

Our research setting provides several unique characteristics that allow us jointly identify worker productivity and bargaining traits. First, salon workers form small quasi-random teams by rotating through clients and teammates, providing rich variation in work shifts, peers, clients, and services. Second, workers receive fixed individual commissions for services (e.g., hair, cosmetics) and team-based commissions for prepaid cards that they attempt to sell following the services. Services and card sales allow us to estimate two productivity measures for each worker. Third, the team-based commissions from selling cards are divided entirely at the discretion of team members, providing a separate measure of worker bargaining outcomes. Finally, the workforce is evenly split between male (52 percent) and female workers, which gives us rich variation in gender mix within teams. We estimate each worker's bargaining trait and service and card sales productivity as risk-adjusted

fixed effects that control for transaction characteristics based on time, task, and coworkers (e.g., Huckman and Pisano 2006, Mas and Moretti 2009).

We find clear evidence of heterogeneity in workers' unique productivity and peer bargaining traits, but most strikingly show that gender strongly predicts under- or over-compensation relative to productivity. Although female workers have slightly higher productivity distributions than men, they have significantly worse bargaining outcomes. Men make up a disproportionate number of highly compensated yet unproductive workers, while women over-represent star employees with poor bargaining outcomes. We implement the Extreme Boosted Gradient Trees machine learning algorithm (Liaw and Wiener 2002) to validate our regression results, showing that gender is the strongest predictor of bargaining outcomes while productivity is primarily based on age and experience.

In order to examine whether the observed inequity in bargaining is driven by greater prosociality or reduced bargaining power, we propose a parsimonious bargaining model based on the ultimatum game, which is commonly used to model fairness aversion in operations and marketing literature (Haitao Cui et al. 2007, Chen et al. 2008). This model includes two mechanisms that shape bargaining outcomes and differ between men and women. First, a worker may accept less advantageous bargaining outcomes because of distributional preferences for prosociality (Bénabou and Tirole 2006, Frey and Meier 2004) or fairness or equity (Huseman et al. 1987, Rabin 1993, Fehr and Schmidt 1999)—preferences more commonly found in women than in men (Croson and Gneezy 2009). Second, a worker with lower bargaining power compared to her co-workers may consistently achieve less advantageous bargaining outcomes. While bargaining power can reflect the inherent bargaining ability (Cahuc et al. 2006, Malhotra and Bazerman 2008, Elfenbein 2015), it can also be based on the embedded social dynamics of ethnicity (Tinsley and Pillutla 1998), status (Blader and Chen 2011), and gender (Bowles and McGinn 2008, Babcock et al. 2017, Bowles et al. 2007, Stevenson and Wolfers 2006).

By comparing simulation results based on this model with statistical moments in our data, we demonstrate that the our data can only be explained by women having both less bargaining power and stronger pro-social preferences. More specifically, we observe that bargaining outcomes of all-male teams have much higher variance than those of all-female teams, which can only be rationalized by women having stronger pro-social preferences. We also observe that men achieve significantly higher bargaining outcomes compared to women in mixed-gender teams, a phenomenon consistent only with women having lower bargaining power.

Our paper contributes to several important research streams. First, we provide unique evidence that the division of value within the firm is heavily embedded in gender-based social dynamics (Bowles and McGinn 2008, Babcock and Laschever 2009, Castilla 2008). Although extensive

work shows inequity in hierarchical decisions, our work demonstrates that discretionary bargaining among peer workers also generates consistent inequity toward women. Importantly, we can also show that both equity preferences and social structure contribute to this pay gap. Our work is consistent with the gender equity literature (Bidwell et al. 2013, Fernandez-Mateo 2009, Cook et al. 2018, Castilla 2015, Sterling and Fernandez 2018) and with the argument that the dynamics of mixed-sex work groups are embedded in social and cultural norms that are crucial to team performance (Chatman 2010). It is also consistent with the argument that less formalized pay structures produce particularly inequitable pay dynamics in bargaining between men and women (Abraham 2017). To the best of our knowledge, ours is the first paper to show that the discretionary division of compensation among *peers* generates gender inequity in an organizational setting, particularly in mixed gender bargaining. Our mechanism evidence further supports the important distinction between "giving" and "giving in" in organizational settings (Cain et al. 2014).

Second, we contribute to a growing literature on personnel operations that studies how to improve worker efficiency and productivity in operations (Huckman and Staats 2011, Buell et al. 2016, Tan and Netessine 2017). Most of the personnel operations literature focuses on how to optimize the allocation of tasks and rewards to workers (Tan and Staats 2016, Ibanez et al. 2017) in order to improve operations, while we demonstrate that this process has particular potential to generate bias against women when it is left to the discretion of teams of peers. We also contribute to the literature on personnel operations that focuses on how to improve worker productivity (Staats et al. 2016, Song et al. 2017) by shedding light on an unfair bargaining process against women that may hinder female workers' productivity and increase their attrition rate.

Third, we add to the literature on star employees (Aguinis and O'Boyle 2014, Groysberg and Lee 2009) and their influence on peers. Most of this literature studies how high-productivity workers increase peer performance through social pressure (Mas and Moretti 2009, Tan and Netessine 2018), helping (Chan et al. 2014a), knowledge transfer (Azoulay et al. 2010, Chan et al. 2014b, Castilla 2005, KC et al. 2013) and recruiting better talent (Fernandez et al. 2000, Agrawal et al. 2017). Our paper builds on recent work that categorizes worker value on dimensions beyond task productivity. Oettl (2012), for example, defines and measure the value of employees not only on their own output, but also on their helpfulness toward peers. Burbano (2016) identifies worker preferences for social responsibility. Pierce and Snyder (2008) measure worker propensity for dishonesty. These studies collectively show that star employees cannot be defined purely on the direct and indirect impact of standard productivity measures.

Finally, since we focus on a typical service operations setting, we also expand the work on empirical service operations (Ang et al. 2015, Akşin et al. 2016, Yu et al. 2016, Ibanez et al. 2017). The existing service operations literature typically studies organizations where tasks and rewards

are dictated by managers, such as in call centers (Akşin et al. 2016, Yu et al. 2016) and hospitals (KC and Terwiesch 2009, Ibanez et al. 2017), and focuses on how to optimize the corresponding task routing or delay announcement policies. To the best of our knowledge, we are the first to study a service setting where tasks or rewards are split through peer bargaining, and provide empirical evidence that this design may not be fair and can significantly impact operational efficiency.

2. Empirical Setting

Our empirical setting is service operations in a chain of 32 large beauty salons in Beijing, China. The salons provide an array of services that include nails, hair, massage, and other beauty treatments for an average price of 143 Chinese Yuan (CNY), or equivalent to 21USD. Worker teams also sell pre-paid cards to customers who can use these cards for future services and enjoy a 10% discount. Customers are on average loading 2634CNY (about 381USD) to their pre-paid cards. Services are provided through teams of on average 1.49 workers. The average salon has 17.8 employees in a given week and makes 357.9 unique transactions per week. The salons are commonly owned by a family and operate seven days a week from 11 a.m. to 11 p.m. Our data span 50 months between April, 2009 and May, 2013 and include every service and card sales transaction as well as detailed demographic information of the workers assigned to those transactions. Our data do not have unique customer identifiers, such that we cannot observe repeated business.

2.1. Staffing and Worker Assignments

The 965 workers in our dataset originate from throughout China, with the highest density in the area surrounding Beijing. Approximately half (52%) are men, with 26% having a high school education. The average age is 29.6. The recruitment and hiring process are not observable in our data, although management explained that employees are hired through multiple processes. Workers can be divided into two broadly job classifications: stylist and beauticians. Stylists are either designated as junior- or senior-level, and are responsible for services related to hair cutting. Beauticians provide services such as facial spa and facial wax. Each transaction includes a combination of multiple tasks, which are classified in four broad categories: “senior task”, “intermediate task”, “junior task”, and ”beautician task”. Beauticians complete all beautician tasks. Senior-level stylists might conduct senior, intermediate, or junior tasks, while junior-level stylists are only assigned junior-level tasks. Senior tasks average 25.15CNY in commissions, intermediate tasks average 26.13CNY, and junior task average 4.21CNY. Beautician tasks average 50.7CNY. A key characteristic of our setting is that worker schedules, team assignment, and customer matching are set through a quasi-random process. Most workers are staffed ten hours per day and five or six days per week, with workers evenly rotating which days they take off and whether or not they open or close on a given day. For a salon of about 20 employees, on average only 2 to 3 take off a given day.

The salons provide customers with one or more services through rotating teams of workers. Since all employees are commission based, management focuses on ensuring that workers view customer and task assignment as fair. Consequently, all stylists and beauticians take turns providing services at their own level within their own group. For example, if a customer orders a "senior stylist wash, cut, and style," the junior-stylist group will send the next available worker to wash the hair (a junior task) while the senior-stylist group will send a senior stylist to cut and style the hair (a senior task). As a result, neither the senior nor the junior stylist can choose the partner he or she wants to work with. Because customer needs vary, requiring different combinations of workers, team composition and peer sets will vary substantially even within a given day. Thus, although our setting is not a true experiment, it roughly approximates one in worker assignment (see Section 5 for more robustness tests on quasi-random team formation).

2.2. Service and Sales Compensation

The database also includes the specific commission paid to each worker for both services and card sales. Workers in our setting receive no base wages, and are paid solely on commission. Each specific service has a fixed individual commission paid to the worker that is beyond their control (typically 21 percent). For example, a senior stylist normally receives 27 percent of the price for a hair cut, regardless of the identity of her teammates. We note that strong norms against tipping in China ensure that we observe total compensation for each worker. A worker on average earns 68.1 percent of her commission from service sales, the remaining 31.9 percent is from card sales.

In addition to services, workers attempt to sell customers prepaid cards that can be used for future services. This sales process starts when the customer first enters the salon and is greeted by an employee acting as a host. When the host hands a customer over to the next worker, he or she will subtly update them on the sales progress. For example, the initial employee may tell a senior stylist that "Mrs. Wang is really interested in getting a gold card. Can you tell her more about it?" All employees who interact with a customer throughout the visit can attempt to sell him/her a prepaid card, with those who contribute to a sale sharing a team-based commission of 9 percent. For example, if both employee A and B are involved in selling Mrs. Wang a card of 2000CNY, they will bargain over a team commission equal to $2000 \times 0.09 = 180$CNY. Card sales are designated in the data as either new cards or refills. Given that the hosts and service teams are quasi-randomly assigned to each customer, the card sales teams are also quasi-randomly assigned.

Our dataset contains $1,287,998$ total transactions, of which approximately 7.03 percent are card sales. Each service transaction identifies the the participating workers, the specific tasks they were assigned, the revenue generated by those tasks, and the individual commission earned by each worker for the task. The card sales transactions list the revenue from the card as well as the workers

involved in the sale and the commission received. We note again that commissions for service are a fixed percentage of revenue and not open for bargaining. In contrast, commissions for card sales are 9 percent for the entire team, and each person's commission is entirely discretionary within the team and bargained from the team's 9-percent commission. Although card sales are much less common than service transactions (7.03%), their share of worker income is higher (31.9%). In this way, we can observe the productivity on one task type where value division is fixed, and another where it is discretionary. This setting has several unique characteristics that make it ideally configured for identifying joint heterogeneity in worker productivity and peer bargaining.

We will restrict our analysis to two samples of workers who work long enough to provide reasonably precise worker fixed effect estimates.

The main sample is the 627 workers with at least 30 weeks of service and 30 card sales, with an alternative less restrictive set of 965 workers who have at least 10 weeks of service data and 10 card transactions. Table 1 presents summary statistics for $38,201$ card sale transactions involving two employees, and $1,197,337$ service transactions among those 965 workers (the remaining $52,460$ transactions can be divided into $39,658$ card transactions with one employee and $12,802$ card transactions with more than two employees). In this sampel of 965 workers, the average service costs 138.07CNY (20 USD), while the average prepaid card sale is $2,479$CNY (359 USD).

3. Identification Strategy

Our empirical strategy is to first estimate productivity and bargaining effects for all workers, and then compare the joint distribution of the individual measures across male and female workers. The strongest evidence of gender-based inequity would be equal or greater productivity by women with higher bargaining shares for men. Productivity will be measured across separate regressions for service revenue and card sales revenue. Bargaining share will be measured based solely on commissions from card transactions since there is no bargaining on service commissions.

We empirically identify productivity and peer bargaining traits as worker fixed effects in linear regressions that control for transaction/service type, schedule shifts, employee rank and classification, flexible time trends, salon identifiers, and coworker identities. This approach, often referred to as "risk adjustment", is widely used to estimate employee or facility performance in such industries as health care (Huckman and Pisano 2006), supermarkets (Mas and Moretti 2009), automotive service (Pierce and Snyder 2008), and sales (Chan et al. 2014a). The approach estimates an employee's average performance, conditional on the unique mix of positive and negative conditions they face. For the heart surgeons in Huckman and Pisano (2006), for example, a surgeon's performance measure is adjusted to account for risk factors, such as preexisting conditions, age, and gender. For the team-based settings in Mas and Moretti (2009) and Chan et al. (2014a), the model also controls for shift-specific customer traffic and the mix of coworkers.

Table 1 Summary Statistics of All Transactions and Workers

Variable	No. of Obs	Mean	Std	Min	Max
Panel A: Two-employee Card Transactions					
Revenue	38,201	2,479.0	3,283.4	8.0	49,000
Commission	38,201	180.9	251.7	0.0	441.0
Hour	38,201	16.9	3.7	0	23
Weekend	38,201	0.37	0.48	0	1
Month	38,201	6.44	3.54	1	12
Year	38,201	2011.7	0.92	2009	2013
New Card	38,201	0.51	0.50	0	1
Panel B: Service Transactions					
No. of Workers	1,197,337	1.48	0.79	1	22
Revenue	1,197,337	138.07	254.04	0.4	4,995
Commission	1,197,337	31.59	54.90	0.0	3,861
Hour	1,197,337	17.19	3.57	0	23
Weekend	1,197,337	0.35	0.48	0	1
Month	1,197,337	6.32	3.52	1	12
Year	1,197,337	2011.6	1.03	2009	2013
Beautician	1,197,337	0.15	0.35	0	1
Product Sale	1,197,337	0.01	0.12	0	1
Panel C: Worker Characteristics					
Age	965	28.59	4.94	19	47
Start Year	965	2010.6	1.07	2009	2013
Male	965	0.52	0.5	0	1
High School Education	965	0.26	0.44	0	1

Note: The table contains all observations related to the 965 workers who have at least worked 10 weeks in our sample period. The unit of observation is transaction. All currency is in Chinese yuan (CNY).

We estimate three separate regressions: two predicting weekly revenue generation and one predicting transaction commission split.[1] The two revenue regressions, which are used to estimate permanent productivity, separately use weekly service revenue and weekly card sales revenue as dependent variables. The commission regression, which estimates the peer bargaining trait, uses only those transactions with teams of two workers (75 percent of all card sales that involve more than one worker[2]) to regress each worker's commission share on worker fixed effects and control variables (see Appendix for distribution of card team size).[3] We detail each regression below.

3.1. Estimating Worker Service Productivity

We use the following equation to estimate each worker's permanent productivity in generating service revenue:

$$\begin{aligned} Log(\text{Service Revenue}_{it}) = & G_i^s + \text{TransactionType}_{it} + \text{Week}_t + \text{WorkerType}_{it} + \\ & \text{Shifts}_{it} + \text{Store}_i + \epsilon_{it}, \end{aligned} \tag{1}$$

[1] We aggregate our productivity metric at the weekly level since we measure productivity by how much revenue an individual can contribute to a salon in a given time. Our results are robust if we relax this to the day-level.

[2] Figure D.1 shows the histogram of number of employees for multi-employee card transactions.

[3] We focus on the card transactions with two employees because (a) we can control for the coworker fixed effect and (b) two-employee bargaining fits our theoretical model.

where the unit of analysis is the worker-week. The dependent variable is the average logged daily commission-adjusted revenue for worker i at their salon i in week t based on the actual number of days worked in week t. In our sample, all workers have only one salon at any given week; therefore, we use the individual subscript to represent the salon fixed effect. For each transaction, we observe the list of workers who have contributed and compute the commission-adjusted revenue for each worker by multiplying the total revenue of this transaction with the commission percentage split of that worker. Since service commissions are fixed and increase with the contribution from one worker, the worker will have higher commission-adjusted revenue from one transaction if she conducts a more important task and contributes more.

We also include controls that might influence a worker's weekly service productivity. Since we focus on all non-card transactions, some of which involve 10 workers, we control for the average number of coworkers in worker i's teams during week t. Note that we cannot directly control for individual coworker fixed effects in this specification but the effect of coworker on the service revenue should be indirectly controlled when we adjust the revenue assigned to each worker by her contribution to that transaction. We also control for the weekly percentage from different types of transactions. The six most commonly coded types of transactions involve more than 64 percent of all transactions: "facial beautician," "style," "stylist haircut," "massage," "simple haircut," and "hair care." The remaining 36% include over $1,000$ rare task descriptions that we collectively code as "other". We also control for the weekly percentage from different tasks within each transaction for a worker, including senior task, intermediate task, junior task and beautician task. We also include store dummies to address store-level sales differences. Lastly, we control for the 14 shift assignments (morning and evening) in a week with dummies and include week fixed effects. The variables of interest are the individual fixed effects in vector G^s, which represent the risk-adjusted permanent service productivity for all workers.

3.2. Estimating Worker Card Sales Productivity

We also estimate a worker's productivity in selling prepaid cards using the following regression model:

$$Log(\text{Card Revenue}_{it}) = G_i^c + \text{Card Initiation}_{it} + \text{WorkerType}_{it} + \text{Week}_t + \text{Shifts}_{it} + \text{Store}_i + \epsilon_{it}. \quad (2)$$

We estimate a similar model to explain weekly card productivity. The dependent variable is the log of total weekly revenue generated by worker i at salon j in week t through card sales. In this analysis, we limit our sample to only those transactions with teams of two workers (87 percent of total card transactions) to be consistent with our bargaining estimation sample below. Unlike the preivous analysis, we do not adjust the revenue by worker's commission split since, in

card transactions, worker's commission split does not only reflect her contribution to the card sale but also significantly depends on her bargaining ability and pro-social preferences. Card Initiation details the percentage of card sales that are refills. Also included are week and shift fixed effects. The variables of interest are the individual fixed effects in vector G^c, which represent the risk-adjusted permanent sales productivity for all workers.

3.3. Estimating Worker Bargaining Trait

We use the following equation to estimate each worker's bargaining trait as an individual fixed effect:

$$\begin{aligned}\text{Cut}_{ijt} =& C_i^c + \text{Coworker}_j + \text{TotalCommission}_{ijt} + \textit{TransactionType}_{ijt} + \\ & \text{CumulativeNumTransactions}_{it-1} + \text{Week}_t + \text{Store}_i + \text{Shift}_{it} + \epsilon_{ijkt}.\end{aligned} \tag{3}$$

As aforementioned, we limit our peer bargaining sample to only those transactions with teams of two workers and define the dependent variable Cut_{ijt} as the percentage of total commission received by worker i working with coworker j on transaction ijt. Unlike in the productivity regressions, our unit of analysis is an individual transaction; and we do not aggregate to weekly levels because we wish to control for the identity of coworkers in a particular transaction. We intentionally control for workers' cumulative number of card transactions before the transaction as a good approximation of their ability in generating card sales. Since the commission gained by a worker during a card sale may depend on her contribution to the card sale, bargaining ability and pro-social preference, by controlling for her cumulative number of sales, we intend to tease out the first factor and let her estimated worker bargaining trait only depend on her bargaining ability and pro-social preference. Again, we control for the store, week, shift, rank and job type, and type of transaction (i.e., card sales versus card refill) with fixed effects.

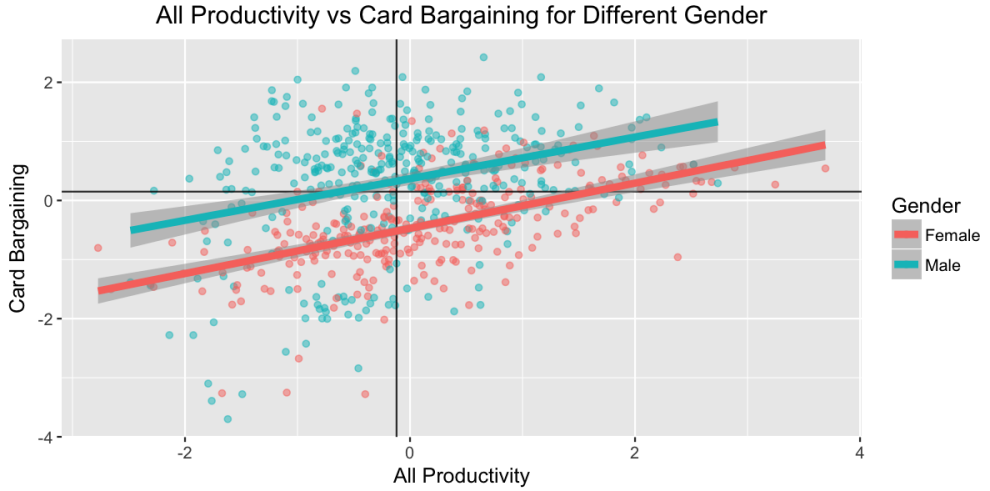
4. Results

4.1. Bargaining and Productivity Estimates

We estimate models 1-3 using ordinary least squares with standard errors clustered at the worker level,[4] identifying productivity traits G^s and G^c and bargaining trait C^c for our samples of 627 and 965 workers. Figure 1 presents the joint distribution of productivity and peer bargaining fixed effects for the 627 worker sample, broken out by gender. Vertical and horizontal lines represent the median worker on productivity and peer bargaining, respectively. Worker traits are substantively meaningful in predicting both productivity and bargaining. Worker productivity fixed effects in our service productivity model explain 20 percent of all variance, while worker fixed effects in our card sales productivity model explain 8 percent of variance. Worker fixed effects in our peer bargaining models explain 40 percent of the variance.

[4] Although clustering at the store level would be more conservative, such standard errors would be biased due to only 32 clusters (Cameron et al. 2008), so we cluster at the worker level. We note that the precision of our fixed effect

Figure 1 Joint Distribution of Total Productivity and Peer Bargaining Fixed Effects by Gender



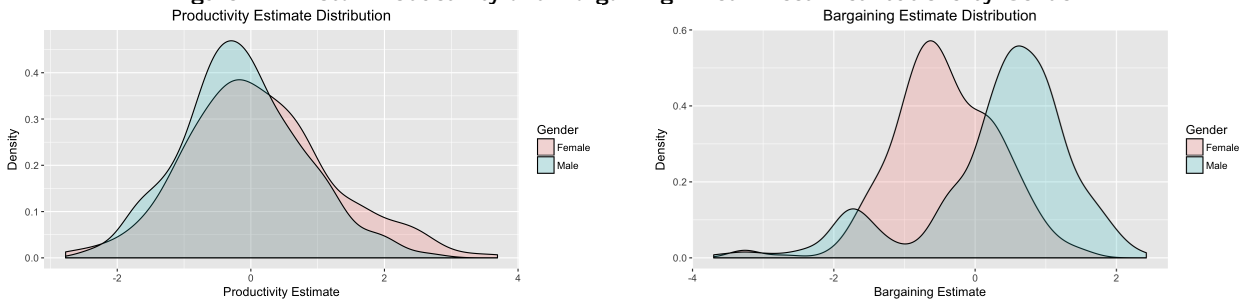
Note: This figure shows the unique bargaining and total productivity (card plus service) estimates for each of the 627 workers with at least 30 weeks of service and 30 card transactions. Men are represented in blue, while women are pink. Vertical and horizontal lines represent medians. Included are linear fits and confidence intervals.

A simple linear OLS model shows that average peer bargaining traits are positively related to task productivity (β=0.277, SE=.036), but this ignores the key differences in this relationship for men and women evident in Figure 1. Although the productivity distributions of women and men are similar, men dominate the top half of peer bargaining outcomes. Female workers receive considerably lower commission splits relative to their contributions compared with male coworkers. Women must reach the 90th percentile of productivity to achieve the bargaining share of the median male worker. The gender disparity is just as severe when examining only card sales productivity. Figure 2 presents kernel density distributions of total productivity and peer bargaining estimates by worker gender. Women show a weakly higher productivity distributions (Kolmogorov-Smirnov, p= 0.0317), while men have significantly better peer bargaining outcomes (p< 0.0001).

To further demonstrate this, we divide workers into four groups by median productivity and bargaining fixed effects for our two samples of 627 and 965 workers. Table 2 shows the remarkable differences by gender in the number of workers who fall in our four median-based quadrants. Based on the 627 worker sample, women are far more likely (32.2 percent) than men (9.9 percent) to have above-median productivity and below-median bargaining outcomes (p= .001), while men are much more likely (33.1 percent) than women (2.2 percent) to receive better bargaining outcomes than their productivity warrants (p< .001). Women are also far less likely (21.9 percent) to receive equitable bargaining outcomes commensurate with their high performance compared to men (36.7 percent, p< .001). We emphasize that this is not because women are less productive. Although

estimates is not of particular importance because we will be comparing the distribution of male and female fixed effects.

Figure 2 Total Productivity and Bargaining Fixed Effect Distributions by Gender



(a) Total Productivity Fixed Effect Distribution (b) Peer Bargaining Fixed Effect Distribution

Note: These figures present kernel density plots of worker fixed effect estimates. Panel (a) shows indistinguishable distributions for male and female productivity estimates. Panel (b) shows distinctly lower bargaining estimates for women. The Kolmogorov-Smirnov test shows that male and female workers have similar productivity fixed effects and the difference is marginally significant (p= 0.0317), while they have different bargaining fixed effects (p< 0.0001).

Table 2 Worker Categorization by Gender

		Sample A			Sample B		
G_i	C_i	Men	Women	P-value	Men	Women	P-value
High	Low	35	88	< 0.001	51	147	< 0.001
		(9.9%)	(32.2%)		(10.1%)	(32.2%)	
High	High	130	60	< 0.001	191	93	< 0.001
		(36.7%)	(21.9%)		(37.6%)	(20.4%)	
Low	Low	72	119	< 0.001	98	187	< 0.001
		(20.3%)	(43.7%)		(19.2%)	(40.8%)	
Low	High	117	6	< 0.001	168	30	< 0.001
		(33.1%)	(2.2%)		(33.1%)	(6.6%)	
Total Workers			627			965	

Note: This table shows the number of male and female workers who fall in each of four quadrants defined by median productivity and bargaining. Sample A requires a minimum of 30 weeks' service while Sample B requires 10 weeks. P-values based on chi-squared tests.

46.6 percent of men have high productivity, 54.1 percent of women meet this standard, a small and only weakly distinguishable difference (p= 0.071). The driving factor behind gender differences is bargaining outcomes; only 24.1 percent of women are in the top half of bargaining traits.

We more formally test this gender discrepancy by regressing each worker's peer bargaining fixed effect on their total productivity fixed effect while controlling for the worker's age, education, and job rank in the organization. It is possible, for example, that the female workers are all of lower rank, younger, or less experienced. We note that since we control for job task, rank, and store in the initial regressions that estimate the productivity and peer bargaining fixed effects, they are not omitted variables in these regressions. Our variable of interest is the worker's gender and the interaction between a dummy for male worker and the productivity fixed effect. Table 3 presents

Table 3 Productivity and Bargaining for Different Genders

	Dependent Variable: Bargaining Fixed Effects					
	Sample A			Sample B		
	I	II	III	IV	V	VI
Card Productivity (G_i^c)	0.231***			0.243***		
	(0.044)			(0.063)		
Service Productivity (G_i^s)		0.368***			0.276***	
		(0.036)			(0.053)	
All Productivity (G_i^a)			0.362***			0.305***
			(0.035)			(0.056)
Male	0.776***	0.682***	0.726***	0.603***	0.484***	0.531***
	(0.144)	(0.123)	(0.121)	(0.129)	(0.117)	(0.115)
Male $\times G_i^c$	−0.152**			−0.199***		
	(0.062)			(0.062)		
Male $\times G_i^s$		0.058			−0.158**	
		(0.078)			(0.079)	
Male $\times G_i^a$			−0.001			0.033
			(0.064)			(0.071)
Worker Characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Observations	627	627	627	965	965	965
R^2	0.363	0.477	0.454	0.344	0.333	0.350

Note: Standard errors, clustered at the employee level, are presented in parentheses, with significance levels: *p<0.1; **p<0.05; ***p<0.01. Sample A requires at least 30 weeks of observations while Sample B requires 10 weeks.

results for regressions that alternatively use service productivity and card sales productivity as dependent variables.

Columns I through III demonstrate how our three different productivity measures correlate with worker bargaining traits in both men and women for those 627 workers with at least 30 weeks of observations. Male workers consistently have higher bargaining outcomes given their productivity even when controlling for position, age, and education. Moreover, the interaction between being a male worker and card productivity is negative, which signals that card sales productivity is more strongly correlated with bargaining fixed effects for female workers than for male workers. Columns IV through VI show consistent results for the sample of workers with at least 10 weeks' observation.

4.2. Predicting Worker Type with Machine Learning

Our results show that male and female workers tend to have similar productivity fixed effects controlling for other factors, and female workers systematically have lower peer bargaining fixed effects compared to male workers. While we cannot control for all possible omitted variables that could jointly correlate with workers' productivity and bargaining outcomes and their gender, we can provide more evidence to this claim by using a predictive model. In particular, if gender indeed significantly affects workers' bargaining outcomes but not their productivity, gender should be an important predictor of workers' bargaining fixed effect but not their productivity fixed effect.

We implement a machine learning model to more formally identify how various demographics including gender predict a worker's combination of productivity and bargaining. Using this approach,

we seek to predict three outcomes with separate models: (a) productivity type, which equals one if the worker's productivity fixed effect is above median and zero otherwise; (b) peer bargaining type, which equals one if the worker's bargaining fixed effect is above median and zero otherwise; and (c) classification type, or in which of the four quadrants the worker falls. The independent variables represent time-invariant worker demographic information: age, gender, education level, applied job title, hometown, starting year, and store. In order to both model the complex function between workers' demographic information and their types and to prevent model over-fitting, we implement a widely adopted machine learning method, Extreme Gradient Boosted Trees, using the R software package (Chen and Guestrin 2016). Boosted tree models bootstrap a multitude of decision trees in order to classify observations (e.g, our worker types), focusing on maximizing out-of-sample predictive accuracy while correcting for overfitting. In each iteration, this approach builds a decision tree by drawing a bootstrapped sample and randomly selecting a subset of features. This approach also adjust the outcome variables in each iteration based on the prediction errors from the decision tree in the previous iteration, in order to improve the decision tree in each round. Based on these the bootstrapped samples with the selected features and the adjusted outcome variables, the algorithm builds a decision tree greedily (i.e., maximizes the information gains of each split from the top of the tree).

Extreme Gradient Boosted Trees have two major benefits over traditional classification methods, such as logistic and probit regression. The first substantial benefit is that it considers complex interactions between important features and automatically drops unimportant features. The second major benefit is the ability to trade off between in-sample accuracy and out-of-sample prediction power. In particular, the boosting tree algorithm has many "hyper-parameters" such as the depth of the trees and the number of features randomly selected in constructing each tree. Depending on the choice of these hyper-parameters, this algorithm can describe functions with different levels of complexity and in turn have different levels of overfitting.

In order to prevent overfitting, we use 3-repeats 10-fold cross-validation to estimate the out-of-sample prediction errors and search for the optimal hyper-parameters. Intuitively, 10-fold cross-validation divides the training sample into 10 pieces. For each piece, the algorithm will train the model based on the other nine pieces of the data and test the algorithm's out-of-sample prediction power on this remaining piece of the data. Therefore, for each set of hyper-parameters, the algorithm will have 10 performance measures (from 10 pieces) and the algorithm's overall performance is simply the average of these 10 performance measures. Three repeats ("3-repeats") means that we repeat the above 10-fold cross validation three times with different data divisions and compute the average out-of-sample performance of the model over these three repeats. The out-of-sample performance measures are then used to find the optimal set of hyper-parameters by searching

Figure 3 In-sample and Out-of-sample Prediction Errors of Bargaining and Productivity Type

(a) Productivity Type Prediction

(b) Bargaining Type Prediction

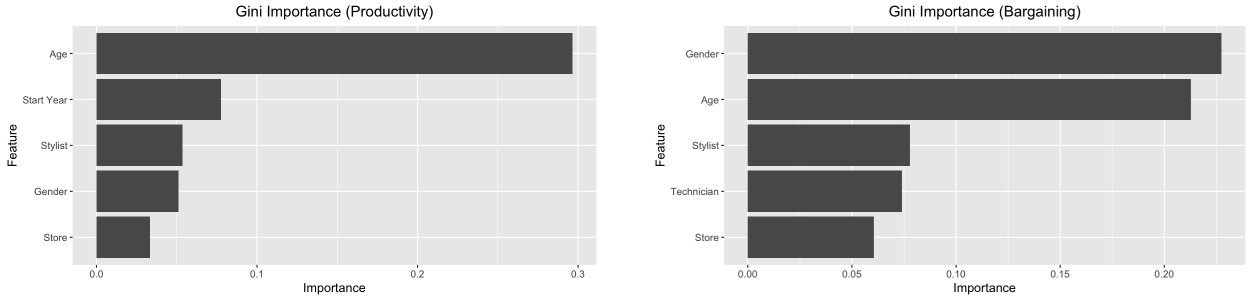
(c) Productivity and Bargaining Joint Type Prediction

Note: This figure presents the errors for Extreme Gradient Boosted Trees predicting above-median productivity and bargaining as well as worker classification type. Panels (a) and (b) show prediction errors for productivity and bargaining type versus 50 percent counterfactual, respectively. Panel (c) shows prediction errors for the four worker types versus 75 percent counterfactual.

through a grid of hyper-parameters. Once the algorithm finds the optimal set of hyper-parameters on the grid, it will construct the boosting tree model on all data (instead of nine pieces of the data) using this set of hyper-parameters.

We first present results for the two models predicting productivity and peer bargaining traits separately. Panels (a) and (b) of Figure 3 represent our in-sample and out-of-sample prediction accuracy for each separate trait. There are several important observations from the figure. First, the out-of-sample error rate of both productivity and the peer bargaining types are well below 0.5 (i.e., the accuracy of randomly guessing), which suggests that worker demographics could help us pre-classify workers into different quadrants. Second, we observe that, as the number of iterations increases, the in-sample error continues to fall while the out-of-sample errors stabilize. This suggests that if we did not use cross-validation to find the best hyper-parameters our model would suffer from severe overfitting problems. Third, we observe that the out-of-sample errors are considerably lower for peer bargaining types than for productivity types. As we will show later, this is because gender is a very strong predictor of bargaining type while only weakly predicting productivity. This is consistent with our previous analysis that female workers are consistently receiving lower

Figure 4 In-sample and Out-of-sample Prediction Errors of Bargaining and Productivity Type



(a) Gini Importance of Top Features (Productivity) (b) Gini Importance of Top Features (Bargaining)

Note: This figure presents the Gini importance for the top five predictors of productivity and bargaining types from our Extreme Gradient Boosted Trees models. Panel (a) shows that age most strongly predicts productivity. Panel (b) shows that both gender and age most strongly predicts bargaining.

bargaining cuts compared to male workers. Moreover, Panel (c) of Figure 3 shows the in-sample and out-of-sample errors for our random forest algorithms to classify the joint types. Similarly, we find that the algorithms' out-of-sample errors are well below the errors from randomly guessing (i.e., 75 percent), and that the algorithm is prone to overfitting.

Last, we analyze the prediction algorithms to demonstrate which demographics account for the difference in prediction power between the two traits. In particular, following traditional machine learning approaches (Strobl et al. 2007), we use the weighted Gini index of a feature to represent the feature's importance in our model. Intuitively, a feature's Gini index represents the average height of the feature on a decision tree. If a feature is at the top of a tree (i.e., the decision tree first splits on this feature), it is important in predicting outcomes. Therefore, the height of a feature on the final model is a good representation of the feature's importance in the algorithm. Panels (a) and (b) of Figure 4 demonstrate the top five features for both productivity and peer bargaining. One important observation is that although gender is only weakly important in predicting worker productivity, it is the most important feature for predicting peer bargaining. Again, this provides indirect evidence that female workers consistently get lower bargaining outcomes compared to male workers with comparable productivity. We note that age is highly important in predicting both productivity and bargaining. Although we also include starting year, this could represent career experience, or, alternatively, status.

5. Identification Tests

Our identification strategy relies on workers' quasi-random variation in customers, shifts, and peers to estimate permanent productivity and bargaining estimates for each worker as fixed effects, while controlling for other non-random confounding factors (such as service type). We next address several possible concerns that might reduce confidence in our results. One concern would be if

Table 4 Staffing Across Gender

	Dependent variable:	
	Weekend	Morning
	(1)	(2)
Male	−0.0002	−0.000
	(0.001)	(0.000)
Worker Characteristics	Yes	Yes
Observations	1,720,736	1,720,736
Note:	*p<0.1; **p<0.05; ***p<0.01	

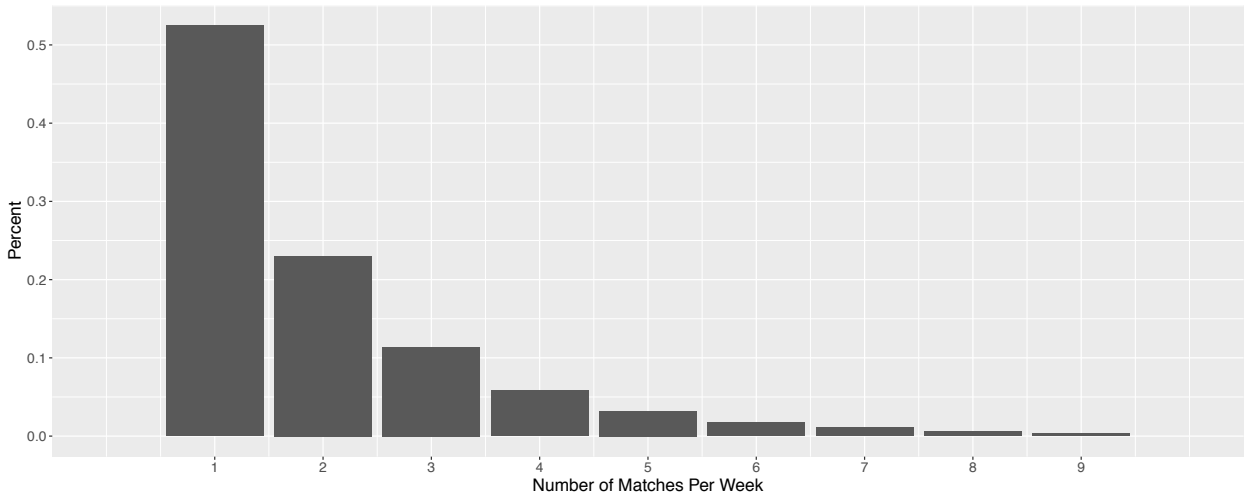
men and women were working very different combinations of shifts, which would suggest gender was driving endogenous staff scheduling decisions. A second concern would be if workers always formed the same teams, suggesting endogenous team formation that would reduce team variation and possibly bias our estimates. Finally, the biggest concern with our results would be if workers were consistently forming teams based on the unobserved ability. Although our interviews with management about the staffing and team formation processes reduce these concerns, we below provide further evidence that they are unlikely to explain our results.

Evidence of consistent staffing across gender: We first provide evidence that men and women are not staffed different in ways that might bias their productivity or bargaining estimates. First, we note that of the 965 workers (our primary sample) that work at least 10 weeks in our data, 773 have a median work week of either 5 or 6 days. Table 4 shows that male workers are no more likely to work on weekends or morning, controlling for their title, age, rank and store.

Evidence of rotational team formation: We next provide the distribution of worker team matches to show there is significant variation in who works with whom. Figure 5 provides a histogram showing the frequency that any two workers at a given store are on the same team in a given week. As is evident, the vast majority of worker pairs work together only one time in a week, which demonstrates the team formation variation crucial for our identification strategy. Furthermore, the figure shows that it is rare to have pairs of workers who consistently serve together—less than 2% of worker pairs appear more than 5 times in a week.

Evidence of consistent teammates' fixed effects across gender: One key concern with our data is that if workers endogenously select into teams based on similar productivity, the service productivity estimators (G_i^s) do not purely capture a worker's productivity, but instead also reflect her consistent pairing with similar workers. A similar problem would occur if managers were strategically pairing workers based on productivity. Such endogeneity problems would potentially endanger our main claim that female workers consistently receive lower bargaining outcomes compared to male workers with comparable productivity. In particular, our results will be confounded

Figure 5 Distribution of Weekly Peer Interactions in Team



Note: We group all transactions at the week and store level, and choose the week and store block if there are more than 5 employees and 100 transactions. We then compute the number of matches that happen between each pair of employees.

if female workers tend to work with more productive coworkers and there is a positive productivity spillover that is not completely captured by commission adjustment.

To formally explore this possible issue, we use service and card transactions with two-person teams (80.2 percent of all multi-worker service transactions) and compute each worker's average teammate productivity for each week. If female workers indeed tend to work with more productive workers and in turn have upwardly biased productivity fixed effects, we should find that female workers have higher coworker productivity. Table 5 shows the results of regressing average coworker productivity on worker gender. Columns (1) and (2) use our smaller sample requiring 30 weeks, while Columns (3) and (4) use the larger sample. The regressions show that whether or not one controls for observable characteristics, female gender appears to imprecisely correlate with coworkers' average productivity with inconsistent direction. This suggests that, even though there may be some endogenous staffing effects in the data, this endogeneity is very minor.

6. Evidence on Mechanisms

Although our results provide strong evidence that women receive low peer bargaining outcomes relative to their productivity, we also care whether these outcomes are freely chosen or instead result from power disparities. We focus on two mechanisms that might explain the gender-based differences in peer bargaining outcomes in our data—distributional preferences and bargaining power. These are two likely mechanisms because they both substantially influence bargaining outcomes and vary across gender. Substantial evidence across disciplines finds that women are more prosocial and fairness-oriented in dividing rewards (Andreoni and Vesterlund 2001, Eckel and Grossman

Table 5 Teammates' Fixed Effects Across Gender

	Dependent variable:			
	Coworker's G_i	Coworker's C_i	Coworker's G_i	Coworker's C_i
	(1)	(2)	(3)	(4)
Male Worker	−0.062	0.165	−0.021	−0.090
	(0.163)	(0.362)	(0.088)	(0.203)
Observations	627	627	965	965
Worker Characteristics	Yes	Yes	Yes	Yes

Note: Robust standard errors presented in parentheses, and are not clustered so as to provide a conservative identification test. *p<0.1; **p<0.05; ***p<0.01. Sample A requires at least 30 weeks of observations, while Sample B requires 10 weeks.

2001, Solnick 2001, Croson and Gneezy 2009) and have less bargaining power for a variety of reasons (Amanatullah and Tinsley 2013, Small et al. 2007, Walters et al. 1998, Lee et al. 2017, Heilman and Haynes 2005).

6.1. A Formal Model of Peer Bargaining

Following the past literature on inequality aversion in economics (Fehr and Schmidt 1999), marketing (Haitao Cui et al. 2007) and operations (Chen et al. 2008), we formally model how distributional preferences and bargaining power can generate peer bargaining outcomes using the one-shot Ultimatum Game. In particular, we assume that the bargaining game has a proposer and a responder, and without loss of generosity, that they split one unit of wealth. The bargaining game is as follows: The proposer proposes a share s to the responder. If the responder accepts the proposal, she receives s, while the proposer receives $1-s$. If the responder declines the proposal, she receives her outside option r_r and the proposer receives her own outside option r_s. In our context, this means that two workers have rational expectations over what splits they will receive if they team up to sell a card. If the expected value of teaming up is lower than the outside options for either worker, they will not form a team in the first place. Otherwise, they will form a team, generate the card sales, and divide the commission based on the expected agreeable split.

Following Fehr and Schmidt (1999), we assume that workers may have distributional preferences, such that they care about outcomes that are inequitable to both themselves and their partners. The worker i's utility for a split (x_i, x_j such that $x_i + x_j = 1$) is:

$$u_i(x_i, x_j) = x_i - \alpha_i \max\{x_j - x_i, 0\} - \beta_i \max\{x_i - x_j, 0\}, \tag{4}$$

where α_i and β_i represent the prosociality of the worker i. Similar to past work, we assume that $\beta_i < \alpha_i$, such that workers care more about disadvantageous inequity. Moreover, we assume that $\beta_i \leq 1$, or that no one is "absolutely" prosocial: the utility gained from getting 0.1 share of the pie is

lower than the disutility from having a 0.1 difference in the bargaining outcomes. Last, we assume that $\max\{r_r, r_s\} \leq 0.5$, representing that an agent, regardless of whether she is the proposer or the responder, would have a positive chance of gaining utility while bargaining.

With this utility function, our bargaining game can be described by the primitives on the proposer's and responder's preferences and outside options $\theta = (\alpha_s, \alpha_r, \beta_s, \beta_r, r_r, r_s)$. Given the primitives, we can characterize the bargaining outcomes:

PROPOSITION 1. *[Bargaining Outcome] Given the primitives of the bargaining game, $\theta = (\alpha_s, \alpha_r, \beta_s, \beta_r, r_r, r_s)$, the bargaining outcome is:*

$$(x_s, x_r) = \begin{cases} (0.5, 0.5), & \text{if } \beta_s \geq 0.5 \text{ and } \beta_r < 0.5 \\ (r_s, r_r) & \text{if } \beta_s \geq 0.5 \text{ and } \beta_r \geq 0.5 \\ (\frac{1+\alpha_r - r_r}{1+2\alpha_r}, \frac{r_r+\alpha_r}{1+2\alpha_r}) & \text{if } \beta_s < 0.5 \end{cases}$$

where (x_s, x_r) is the share that the proposer and responder receive and $x_s + x_r = 1$ if there is a trade.

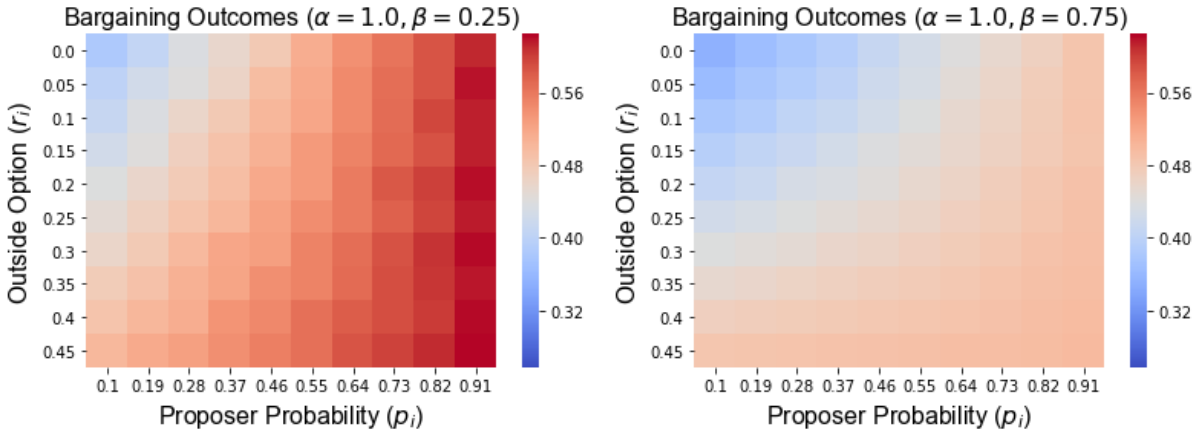
All proofs are in the Appendix. Proposition 1 shows that a worker's final split is co-determined by her distributional preference and outside option. If a person has a higher prosocial preference, she will have higher β in the model. Moreover, the person's outside option and her probability of being the proposer jointly determines her bargaining power. Therefore, a person will get a higher average bargaining outcome if she (a) is more likely to be a proposer (p_i is larger), (b) is less prosocial (β or α is larger), and (c) has a higher outside option (r is larger).

Suppose that a worker i with (α_i, β_i, r_i) is bargaining with a pool of workers with primitives drawn from three distributions ($\alpha \in \mathcal{A}$, $\beta \in \mathcal{B}$ and $r \in \mathcal{R}$). The probability that worker i is a proposer is denoted as $p_i \in (0,1)$. The average split of worker i when there is a trade is denoted as x_i. The following lemma (see Appendix for proof) formally summarizes the aforementioned comparative statics results related to average bargaining outcomes:

LEMMA 1. *[Comparative Statics] x_i is weakly increasing in p_i and r_i. x_i is weakly increasing in α_i and x_i is weakly decreasing in β_i.*

Given Proposition 1, we can simulate a person's share given his or her primitives. In particular, we assume that the worker bargains with a population of workers whose α is drawn from a uniform distribution on $[0,1]$, and β is drawn from a uniform distribution on $[0,\alpha]$, and r is drawn from a uniform distribution on $[0,0.5]$. We simulate the worker's average bargaining outcomes with respect to $r_i \in [0,0.5]$ and $x_i \in (0,1)$, $\alpha_i = 1.0$ and $\beta_i \in \{0.25, 0.75\}$. Panel (a) in Figure 6 represents the average payoff of the agent with respect to her outside option and proposing probability under low prosociality ($\beta = 0.25$). As the agent's bargaining power increases—her outside option (r_i) or proposing probability (p_i) increases—the agent has a higher average payoff regardless of α_i and β_i. Panel (b) in Figure 6 shows the average payoff under high prosociality ($\beta = 0.75$). Again, as the

Figure 6 Peer Bargaining Based on Bargaining Power and Fairness Preferences



(a) Average Payoffs with low prosociality (β) (b) Average Payoffs with high prosociality (β)

Note: This figure presents bargaining share based on different levels of distributional preferences, outside option, and proposer probability in our ultimatum game model. Each point represents $1,000$ simulations.

agent's bargaining power increases, the agent has a higher average payoff regardless of α_i and β_i. Moreover, comparing the two panels, we find that the agent will receive lower payoff (for each pair of x_i and r_i) if the agent has a higher β.

Our model implies that there are multiple reasons why a worker might have a lower peer bargaining trait. First, she may have a stronger fairness preference β. Second, she may have a lower outside option r. Third, she may have a lower probability of being the proposer, which in our case can represent bargaining ability.

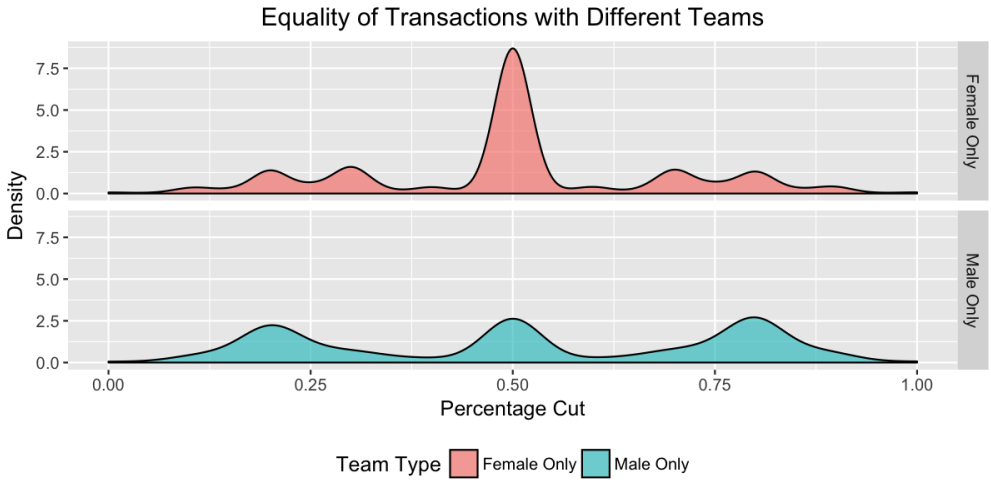
To help identify which mechanisms explain our results, we first present the raw commission splits based on the gender mix in teams of two: all-female, all-male, and mixed-gender. Figure 7 shows the kernel density plot of the commission split for different gender mixes. In particular, we observe two important moment conditions in Figure 7:

1. Panel (a) of Figure 7 shows that half of all cards sold by all-female teams result in a 50-50 commission split, while in all-male teams less than 30 percent of transactions result in a 50-50 commission split. This implies that, in same-gender groups, the average standard deviation of female-to-female splits is much smaller than that of male-to-male splits.

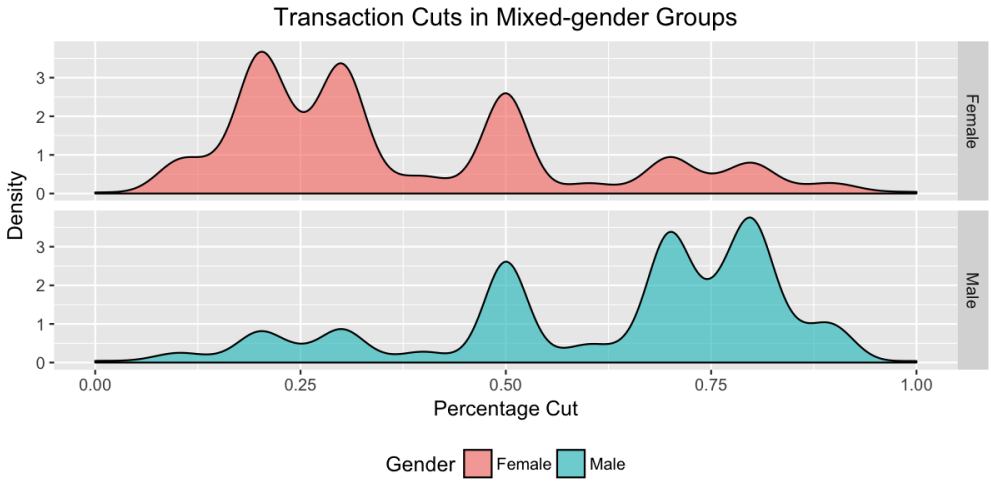
2. Panel (b) of Figure 7 shows that in mixed-gender groups, male workers' splits are much larger than those of female workers.

To more formally test for preference- and power-based mechanisms, we conduct a large-scale simulation based on our theoretical model to match the observed moment conditions above. In this simulation, we vary the gender-specific prosociality (i.e., $\alpha_{male}, \beta_{male}, \alpha_{female}, \beta_{female}$) and bargaining power of male and female workers (i.e., the proposing probability of male workers). The

Figure 7 All-Female Teams More Likely to Split Commissions Evenly



(a) Splits for All-Female and All-Male Teams



(b) Splits for Female and Male Workers in Mixed Team

Note: This figure presents kernel density plots of raw data on bargaining splits in two-person card sales teams. Panel (a) shows the distributions in same-gender teams. Panel (b) shows mixed-gender teams.

simulation, presented in detail in the Appendix A, numerically shows that female workers need to have *both* lower bargaining power and higher prosociality to explain the two aforementioned observations about the moment conditions in Figure 7. In other words, the result from Panel (a) of Figure 7 is consistent with the argument that female workers have stronger pro-social preferences and that these distribution preferences push them to seek more equal splits, while the result from Panel (b) is consistent with female workers having lower bargaining power and in turn receiving worse outcomes when bargaining with men. This also supports the argument by Abraham 2017 that women suffer greater inequity in less formalized pay systems when working with men.

7. Discussion and Conclusion

In this paper, we argue that two persistent traits are crucial in determining the value of a given employee in team-based production: (1) their productivity in crucial tasks and (2) their bargaining outcomes with peers. Our evidence from productivity and commission split bargaining among Chinese salon workers validates that there is indeed significant heterogeneity and explanatory power in these persistent traits—they explain 16.46 percent of the variance in productivity and 8.0 percent in commission splits. Just as importantly, our results imply that measuring a worker's net contribution purely on productivity ignores a crucial orthogonal dimension—the value extracted by the worker for that productivity.

These initial results, however, ignore a crucial predictor of their joint distribution—employee gender. In our sample, male workers consistently extract advantageous bargaining values from their female coworkers, despite having no observable productivity advantage. Consistent with overwhelming evidence on pay inequity, women in our sample earn less than their contributions would merit. And as Castilla and Benard (2010) highlight, this occurs even in the context of a pay-for-performance system. The uniqueness of our results, however, is that this inequity also results from bargaining outcomes with peers. Although we can only observe bargaining on commission splits, we believe the implications apply more broadly to inequitable task and effort divisions in teams as well (Babcock et al. 2017). A broad literature on married couples, for example, consistently finds that even as wives increase professional success and earnings relative to their husbands, they continue to contribute disproportionately to household labor and child-rearing (Pollak 1985).

In addition, our results have significant implications for operational effectiveness and other important performance metrics. Our results identify substantial variation in worker output relative to compensation, which implies that firms could reduce costs by replacing their overcompensated workers. Our results also provide more complex long-run implications based on social comparison and outside labor market options. Psychologists and economists have long recognized that employees engage in social comparisons with coworkers (Festinger 1954), with one of the key comparisons being the assessment of the fairness or equity of their own rewards and contributions versus that of their peers.

Workers with disproportionate bargaining outcomes relative to their productivity are likely to produce perceptions of inequity, since their bargaining traits are not rewarding them commensurate with their contributions through productivity. Underpaid workers may not perceive inequity, since they may have low peer bargaining traits because of preferences, but if their peer bargaining traits are due to ability or social determinants, perceptions of inequity are inevitable. Overpaid workers, however, will generate widespread perceptions of inequity because their rewards relative to their contributions will exceed each of the other three worker types. Star workers, for example,

will observe their overpaid counterparts as receiving equivalent peer bargaining outcomes despite substantially lower contributions through productivity.

Recent work has explained exactly how costly such widespread social comparison can be for a firm's performance when it reveals inequity (Larkin et al. 2012, Gartenberg and Wulf 2017, Roels and Su 2013, Breza et al. 2017). Wage and other reward comparisons can generate feelings of envy that can reduce effort and increase turnover (Nickerson and Zenger 2008, Card et al. 2012, Obloj and Zenger 2017). Perceptions of inequity can also increase unethical behavior and misconduct (Gino and Pierce 2009, 2010, Edelman and Larkin 2014, John et al. 2014, Larkin and Pierce 2015) and can even demotivate employees across unrelated tasks (Gubler et al. 2016).

Even in the absence of social comparison, the decoupling of productivity and peer bargaining can hurt the firm through attrition to outside labor markets. If some workers consistently receive less value than their peers because of either bargaining ability or social determinants, they are likely to look for outside employment where peer bargaining is less determinant of the rewards structure, such as firms where their productivity is directly rewarded and where their peer bargaining weakness is irrelevant because of task and reward assignments are either standardized or hierarchically assigned by managers. Indeed, the strategic human capital literature (Coff 1997, Chadwick 2017, Lepak and Snell 1999, Campbell et al. 2012, Lee et al. 2015) emphasizes that the division of value among stakeholders within the firm is one of the the most important challenges facing top management.

We note that our machine learning models directly address the common concern that correlations observed in regression models might be spurious. Standard errors in team- or network-based studies are notoriously difficult to estimate (Gel et al. 2017, Snijders and Borgatti 1999), and even with our cluster correction at the worker level we might be concerned that the joint distribution of productivity and bargaining fixed effects simply reflects noise. Our machine learning approach, however, provides validation that productivity and peer bargaining traits are meaningful and predictable based on observable worker characteristics. Most clearly, we can be confident that women's consistently lower bargaining outcomes are driving our classification.

Like most field studies, our results from China are embedded in a specific culture with unique norms and preferences around bargaining, fairness, and gender roles. Extensive work shows differences across cultures in coworker interaction (Tinsley and Brett 2001), negotiation strategies and outcomes (Gunia et al. 2016, Kopelman et al. 2016), and most specifically around fairness or equality (Gelfand et al. 2002). Furthermore, the gender pay inequality we observe is specific to China. Social scientists have found broad differences in pay equity across countries (Blau and Kahn 2003). We note, however, that this literature has typically found more equal negotiation outcomes in China than in Western countries (e.g., Tinsley and Pillutla 1998), which would suggest greater

bargaining trait variance in settings different from our own. As the world's largest economy with 20 percent of world population, even culturally specific results are important and generalizable.

Although one of the strengths of our setting is that peer workers directly bargain over monetary rewards, we acknowledge that this specific policy is not commonly used. Yet we believe the peer bargaining over *monetary* rewards is more broadly generalizable to non-monetary rewards commonly bargained over in teams of peers. These rewards, such as recognition, promotions, and awards, are also important components of equity calculations, and are also likely to suffer the same gender inequity observed here. Furthermore, given the trend toward self-management in many industries, our study is a warning that pay and other rewards may be poor candidates for allocation by peers.

We note several limitations of our study. First, we cannot observe why these workers choose to join or leave the salons, and the process through which teams are assigned is not truly random. But interviews with management indicate that it is sufficiently random to make our results very difficult to explain through endogenous sorting. Furthermore, our identification tests show little evidence that worker teams are strategically form based on productivity. Instead, workers appear to rotate through their many peers as teammates. So we are doubtful that strategic staffing decisions are seriously biasing our results, but we cannot fully dispel this concern. Second, we cannot observe any exchange between coworkers that occurs outside our data. It is possible, for example, that workers may give financial or other side payments in exchange for larger bargaining shares in commissions.

Finally, we note that if worker bargaining traits are known by coworkers, then those with inequitably high bargaining outcomes may impose an additional cost to the firm–that is, decreased joint gains (Tinsley et al. 2002). Workers anticipating a poor commission split may focus less effort on card sales with coworkers whom they know will likely extort unequal commission shares. Furthermore, the income inequality observed in our firm may represent broader costs to society detailed in extensive literature on wage and workload inequity.

References

Abraham, Mabel. 2017. Pay formalization revisited: Considering the effects of manager gender and discretion on closing the gender wage gap. *Academy of Management Journal* **60**(1) 29–54.

Agrawal, Ajay, John McHale, Alexander Oettl. 2017. How stars matter: Recruiting and peer effects in evolutionary biology. *Research Policy* **46**(4) 853–867.

Aguinis, Herman, Ernest O'Boyle. 2014. Star performers in twenty-first century organizations. *Personnel Psychology* **67**(2) 313–350.

Akşin, Zeynep, Baris Ata, Seyed Morteza Emadi, Che-Lin Su. 2016. Impact of delay announcements in call centers: An empirical approach. *Operations Research* **65**(1) 242–265.

Amanatullah, Emily T, Catherine H Tinsley. 2013. Punishing female negotiators for asserting too much... or not enough: Exploring why advocacy moderates backlash against assertive female negotiators. *Organizational Behavior and Human Decision Processes* **120**(1) 110–122.

Andreoni, James, Lise Vesterlund. 2001. Which is the fair sex? gender differences in altruism. *The Quarterly Journal of Economics* **116**(1) 293–312.

Ang, Erjie, Sara Kwasnick, Mohsen Bayati, Erica L Plambeck, Michael Aratow. 2015. Accurate emergency department wait time prediction. *Manufacturing & Service Operations Management* **18**(1) 141–156.

Azoulay, Pierre, Joshua S Graff Zivin, Jialan Wang. 2010. Superstar extinction. *The Quarterly Journal of Economics* **125**(2) 549–589.

Babcock, Linda, Sara Laschever. 2009. *Women don't ask: Negotiation and the gender divide*. Princeton University Press.

Babcock, Linda, Maria P Recalde, Lise Vesterlund, Laurie Weingart. 2017. Gender differences in accepting and receiving requests for tasks with low promotability. *American Economic Review* **107**(3) 714–747.

Bamberger, Peter A, Racheli Levi. 2009. Team-based reward allocation structures and the helping behaviors of outcome-interdependent team members. *Journal of Managerial Psychology* **24**(4) 300–327.

Bénabou, Roland, Jean Tirole. 2006. Incentives and prosocial behavior. *American economic review* **96**(5) 1652–1678.

Bernstein, Ethan, John Bunch, Niko Canner, Michael Lee. 2016. Beyond the holacracy hype. *Harvard business review* **94**(7) 8.

Bidwell, Matthew, Forrest Briscoe, Isabel Fernandez-Mateo, Adina Sterling. 2013. The employment relationship and inequality: How and why changes in employment practices are reshaping rewards in organizations. *Academy of Management Annals* **7**(1) 61–121.

Blader, Steven L, Ya-Ru Chen. 2011. What influences how higher-status people respond to lower-status others? effects of procedural fairness, outcome favorability, and concerns about status. *Organization Science* **22**(4) 1040–1060.

Blau, Francine D, Lawrence M Kahn. 2003. Understanding international differences in the gender pay gap. *Journal of Labor economics* **21**(1) 106–144.

Bowles, Hannah Riley, Linda Babcock, Lei Lai. 2007. Social incentives for gender differences in the propensity to initiate negotiations: Sometimes it does hurt to ask. *Organizational Behavior and Human Decision Processes* **103**(1) 84–103.

Bowles, Hannah Riley, Kathleen L McGinn. 2008. Untapped potential in the study of negotiation and gender inequality in organizations. *Academy of Management Annals* **2**(1) 99–132.

Breza, Emily, Supreet Kaur, Yogita Shamdasani. 2017. The morale effects of pay inequality. *The Quarterly Journal of Economics* **133**(2) 611–663.

Buell, Ryan W, Tami Kim, Chia-Jung Tsay. 2016. Creating reciprocal value through operational transparency. *Management Science* **63**(6) 1673–1695.

Burbano, Vanessa C. 2016. Social responsibility messages and worker wage requirements: Field experimental evidence from online labor marketplaces. *Organization Science* **27**(4) 1010–1028.

Cahuc, Pierre, Fabien Postel-Vinay, Jean-Marc Robin. 2006. Wage bargaining with on-the-job search: Theory and evidence. *Econometrica* **74**(2) 323–364.

Cain, Daylian M, Jason Dana, George E Newman. 2014. Giving versus giving in. *Academy of Management Annals* **8**(1) 505–533.

Cameron, A Colin, Jonah B Gelbach, Douglas L Miller. 2008. Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics* **90**(3) 414–427.

Campbell, Benjamin A, Russell Coff, David Kryscynski. 2012. Rethinking sustained competitive advantage from human capital. *Academy of Management Review* **37**(3) 376–395.

Card, David, Ana Rute Cardoso, Patrick Kline. 2015. Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women. *The Quarterly Journal of Economics* **131**(2) 633–686.

Card, David, Alexandre Mas, Enrico Moretti, Emmanuel Saez. 2012. Inequality at work: The effect of peer salaries on job satisfaction. *American Economic Review* **102**(6) 2981–3003.

Castilla, Emilio J. 2005. Social networks and employee performance in a call center. *American Journal of Sociology* **110**(5) 1243–1283.

Castilla, Emilio J. 2008. Gender, race, and meritocracy in organizational careers. *American Journal of Sociology* **113**(6) 1479–1526.

Castilla, Emilio J. 2015. Accounting for the gap: A firm study manipulating organizational accountability and transparency in pay decisions. *Organization Science* **26**(2) 311–333.

Castilla, Emilio J, Stephen Benard. 2010. The paradox of meritocracy in organizations. *Administrative Science Quarterly* **55**(4) 543–676.

Chadwick, Clint. 2017. Toward a more comprehensive model of firms' human capital rents. *Academy of Management Review* **42**(3) 499–519.

Chan, Tat Y, Jia Li, Lamar Pierce. 2014a. Compensation and peer effects in competing sales teams. *Management Science* **60**(8) 1965–1984.

Chan, Tat Y, Jia Li, Lamar Pierce. 2014b. Learning from peers: Knowledge transfer and sales force productivity growth. *Marketing Science* **33**(4) 463–484.

Chatman, Jennifer A. 2010. Norms in mixed sex and mixed race work groups. *Academy of Management Annals* **4**(1) 447–484.

Chen, Kay-Yut, Murat Kaya, Özalp Özer. 2008. Dual sales channel management with service competition. *Manufacturing & Service Operations Management* **10**(4) 654–675.

Chen, Tianqi, Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 785–794.

Coff, Russell W. 1997. Human assets and management dilemmas: Coping with hazards on the road to resource-based theory. *Academy of Management Review* **22**(2) 374–402.

Cook, Cody, Rebecca Diamond, Jonathan Hall, John List, Paul Oyer. 2018. The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers. Working paper, Stanford University.

Croson, Rachel, Uri Gneezy. 2009. Gender differences in preferences. *Journal of Economic Literature* **47**(2) 448.

Eckel, Catherine C, Philip J Grossman. 2001. Chivalry and solidarity in ultimatum games. *Economic Inquiry* **39**(2) 171–188.

Edelman, Benjamin, Ian Larkin. 2014. Social comparisons and deception across workplace hierarchies: Field and experimental evidence. *Organization Science* **26**(1) 78–98.

Elfenbein, Hillary Anger. 2015. Individual differences in negotiation: A nearly abandoned pursuit revived. *Current Directions in Psychological Science* **24**(2) 131–136.

Fehr, Ernst, Klaus M Schmidt. 1999. A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics* **114**(3) 817–868.

Fernandez, Roberto M, Emilio J Castilla, Paul Moore. 2000. Social capital at work: Networks and employment at a phone center. *American Journal of Sociology* **105**(5) 1288–1356.

Fernandez-Mateo, Isabel. 2009. Cumulative gender disadvantage in contract employment. *American Journal of Sociology* **114**(4) 871–923.

Festinger, Leon. 1954. A theory of social comparison processes. *Human Relations* **7**(2) 117–140.

Frey, Bruno S, Stephan Meier. 2004. Social comparisons and pro-social behavior: Testing" conditional cooperation" in a field experiment. *American Economic Review* **94**(5) 1717–1722.

Gartenberg, Claudine, Julie Wulf. 2017. Pay harmony? social comparison and performance compensation in multibusiness firms. *Organization Science* **28**(1) 39–55.

Gel, Yulia R, Vyacheslav Lyubchich, L Leticia Ramirez Ramirez. 2017. Bootstrap quantification of estimation uncertainties in network degree distributions. *Scientific Reports* **7**(1) 5807.

Gelfand, Michele J, Marianne Higgins, Lisa H Nishii, Jana L Raver, Alexandria Dominguez, Fumio Murakami, Susumu Yamaguchi, Midori Toyama. 2002. Culture and egocentric perceptions of fairness in conflict and negotiation. *Journal of Applied Psychology* **87**(5) 833.

Gino, Francesca, Lamar Pierce. 2009. Dishonesty in the name of equity. *Psychological science* **20**(9) 1153–1160.

Gino, Francesca, Lamar Pierce. 2010. Robin hood under the hood: Wealth-based discrimination in illicit customer help. *Organization Science* **21**(6) 1176–1194.

Groysberg, Boris, Linda-Eling Lee. 2009. Hiring stars and their colleagues: Exploration and exploitation in professional service firms. *Organization Science* **20**(4) 740–758.

Gubler, Timothy, Ian Larkin, Lamar Pierce. 2016. Motivational spillovers from awards: Crowding out in a multitasking environment. *Organization Science* **27**(2) 286–303.

Gunia, Brian C, Jeanne M Brett, Michele J Gelfand. 2016. The science of culture and negotiation. *Current Opinion in Psychology* **8** 78–83.

Haitao Cui, Tony, Jagmohan S Raju, Z John Zhang. 2007. Fairness and channel coordination. *Management science* **53**(8) 1303–1314.

Hamilton, Barton H, Jack A Nickerson, Hideo Owan. 2003. Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation. *Journal of Political Economy* **111**(3) 465–497.

Heilman, Madeline E, Michelle C Haynes. 2005. No credit where credit is due: attributional rationalization of women's success in male-female teams. *Journal of Applied Psychology* **90**(5) 905.

Huckman, Robert S, Gary P Pisano. 2006. The firm specificity of individual performance: Evidence from cardiac surgery. *Management Science* **52**(4) 473–488.

Huckman, Robert S, Bradley R Staats. 2011. Fluid tasks and fluid teams: The impact of diversity in experience and team familiarity on team performance. *Manufacturing & Service Operations Management* **13**(3) 310–328.

Huckman, Robert S, Bradley R Staats, David M Upton. 2009. Team familiarity, role experience, and performance: Evidence from indian software services. *Management science* **55**(1) 85–100.

Huseman, Richard C, John D Hatfield, Edward W Miles. 1987. A new perspective on equity theory: The equity sensitivity construct. *Academy of Management Review* **12**(2) 222–234.

Ibanez, Maria R, Jonathan R Clark, Robert S Huckman, Bradley R Staats. 2017. Discretionary task ordering: Queue management in radiological services. *Management Science* .

John, Leslie K, George Loewenstein, Scott I Rick. 2014. Cheating more for less: Upward social comparisons motivate the poorly compensated to cheat. *Organizational Behavior and Human Decision Processes* **123**(2) 101–109.

Jung, HeeJung, Balagopal Vissa, Michael Pich. 2017. How do entrepreneurial founding teams allocate task positions? *Academy of Management Journal* **60**(1) 264–294.

KC, Diwas, Bradley R Staats, Francesca Gino. 2013. Learning from my success and from others' failure: Evidence from minimally invasive cardiac surgery. *Management Science* **59**(11) 2435–2449.

KC, Diwas S, Christian Terwiesch. 2009. Impact of workload on service time and patient safety: An econometric analysis of hospital operations. *Management Science* **55**(9) 1486–1498.

Kopelman, Shirli, Ashley E Hardin, Christopher G Myers, Leigh Plunkett Tost. 2016. Cooperation in multicultural negotiations: How the cultures of people with low and high power interact. *Journal of Applied Psychology* **101**(5) 721.

Larkin, Ian, Lamar Pierce. 2015. Compensation and employee misconduct: the inseparability of productive and counterproductive behavior in firms. *Cambridge University Press: Cambridge, UK* 1–27.

Larkin, Ian, Lamar Pierce, Francesca Gino. 2012. The psychological costs of pay-for-performance: Implications for the strategic compensation of employees. *Strategic Management Journal* **33**(10) 1194–1214.

Lee, Jeong-Yeon, Daniel G Bachrach, Denise M Rousseau. 2015. Internal labor markets, firm-specific human capital, and heterogeneity antecedents of employee idiosyncratic deal requests. *Organization Science* **26**(3) 794–810.

Lee, Margaret, Marko Pitesa, Madan M Pillutla, Stefan Thau. 2017. Male immorality: An evolutionary account of sex differences in unethical negotiation behavior. *Academy of Management Journal* **60**(5) 2014–2044.

Lepak, David P, Scott A Snell. 1999. The human resource architecture: Toward a theory of human capital allocation and development. *Academy of management review* **24**(1) 31–48.

Liaw, Andy, Matthew Wiener. 2002. Classification and regression by randomforest. r news, 2 (3), 18–22. *URL: http://CRAN. R-project. org/doc/Rnews* .

Malhotra, Deepak, Max H Bazerman. 2008. Psychological influence in negotiation: An introduction long overdue.

Mas, Alexandre, Enrico Moretti. 2009. Peers at work. *American Economic Review* **99**(1) 112–145.

Nickerson, Jack A, Todd R Zenger. 2008. Envy, comparison costs, and the economic theory of the firm. *Strategic Management Journal* **29**(13) 1429–1449.

Obloj, Tomasz, Todd Zenger. 2017. Organization design, proximity, and productivity responses to upward social comparison. *Organization Science* **28**(1) 1–18.

Oettl, Alexander. 2012. Reconceptualizing stars: Scientist helpfulness and peer performance. *Management Science* **58**(6) 1122–1140.

Pierce, Lamar, Jason Snyder. 2008. Ethical spillovers in firms: Evidence from vehicle emissions testing. *Management Science* **54**(11) 1891–1903.

Pollak, Robert A. 1985. A transaction cost approach to families and households. *Journal of Economic Literature* **23**(2) 581–608.

Prendergast, Canice. 1999. The provision of incentives in firms. *Journal of Economic Literature* **37**(1) 7–63.

Rabin, Matthew. 1993. Incorporating fairness into game theory and economics. *American Economic Review* 1281–1302.

Roels, Guillaume, Xuanming Su. 2013. Optimal design of social comparison effects: Setting reference groups and reference points. *Management Science* **60**(3) 606–627.

Sarsons, Heather. 2017. Recognition for group work: Gender differences in academia. *American Economic Review* **107**(5) 141–45.

Shaw, Jason D, Michelle K Duffy, Eric M Stark. 2001. Team reward attitude: Construct development and initial validation. *Journal of Organizational Behavior* **22**(8) 903–917.

Small, Deborah A, Michele Gelfand, Linda Babcock, Hilary Gettman. 2007. Who goes to the bargaining table? the influence of gender and framing on the initiation of negotiation. *Journal of Personality and Social Psychology* **93**(4) 600.

Snijders, Tom AB, Stephen P Borgatti. 1999. Non-parametric standard errors and tests for network statistics. *Connections* **22**(2) 61–70.

Solnick, Sara J. 2001. Gender differences in the ultimatum game. *Economic Inquiry* **39**(2) 189–200.

Song, Hummy, Anita L Tucker, Karen L Murrell, David R Vinson. 2017. Closing the productivity gap: Improving worker productivity through public relative performance feedback and validation of best practices. *Management Science* .

Staats, Bradley R, Hengchen Dai, David Hofmann, Katherine L Milkman. 2016. Motivating process compliance through individual electronic monitoring: An empirical examination of hand hygiene in healthcare. *Management Science* **63**(5) 1563–1585.

Sterling, Adina D, Roberto M Fernandez. 2018. Once in the door: Gender, tryouts, and the initial salaries of managers. *Management Science* .

Stevenson, Betsey, Justin Wolfers. 2006. Bargaining in the shadow of the law: Divorce laws and family distress. *The Quarterly Journal of Economics* **121**(1) 267–288.

Strobl, Carolin, Anne-Laure Boulesteix, Achim Zeileis, Torsten Hothorn. 2007. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC bioinformatics* **8**(1) 25.

Tan, Tom, Serguei Netessine. 2017. At your service on the table: Impact of tabletop technology on restaurant performance .

Tan, Tom, Serguei Netessine. 2018. When you work with a superman, will you also fly? an empirical study of the impact of coworkers on performance .

Tan, Tom, Bradley R Staats. 2016. Behavioral drivers of routing decisions: Evidence from restaurant table assignment .

Tinsley, Catherine H, Jeanne M Brett. 2001. Managing workplace conflict in the united states and hong kong. *Organizational Behavior and Human Decision Processes* **85**(2) 360–381.

Tinsley, Catherine H, Kathleen M O'Connor, Brandon A Sullivan. 2002. Tough guys finish last: The perils of a distributive reputation. *Organizational Behavior and Human Decision Processes* **88**(2) 621–642.

Tinsley, Catherine H, Madan M Pillutla. 1998. Negotiating in the united states and hong kong. *Journal of International Business Studies* **29**(4) 711–727.

Walters, Amy E, Alice F Stuhlmacher, Lia L Meyer. 1998. Gender and negotiator competitiveness: A meta-analysis. *Organizational Behavior and Human Decision Processes* **76**(1) 1–29.

Wuchty, Stefan, Benjamin F Jones, Brian Uzzi. 2007. The increasing dominance of teams in production of knowledge. *Science* **316**(5827) 1036–1039.

Yu, Qiuping, Gad Allon, Achal Bassamboo. 2016. How do delay announcements shape customer behavior? an empirical study. *Management Science* **63**(1) 1–20.

Appendix

A. Simulation on Bargaining Outcomes

In order to understand what mechanisms drive the results in Figure 7, we conduct a large-scale simulation. In particular, we observe two important moment conditions in Figure 7:

1. Panel (a) shows that in same-gender groups the average standard deviation of female-to-female splits is much smaller than that of male-to-male splits. In other words, females are more likely to split around 50 percent.

2. Panel (b) shows that in mixed-gender groups male workers' splits are much larger than those of female workers.

Therefore, we design a simulation method by varying the proposociality of male and female workers (i.e., $\alpha_{male}, \beta_{male}, \alpha_{female}, \beta_{female}$) and the bargaining power of male and female workers (i.e., the proposing probability of male workers).[5] In this simulation, we have five parameters:

1. Prosociality of male workers: $\alpha_{male} = \beta_{male} \sim Uniform(0, x_{male})$
2. Prosociality of female workers: $\alpha_{female} = \beta_{female} \sim Uniform(0, x_{female})$
3. Proposing probability of male workers: $p_{male} = 1 - p_{female}$
4. Outside options of male workers: $r_{male} \sim Uniform(0, \bar{r}_{male})$
5. Outside options of female workers: $r_{female} \sim Uniform(0, \bar{r}_{female})$

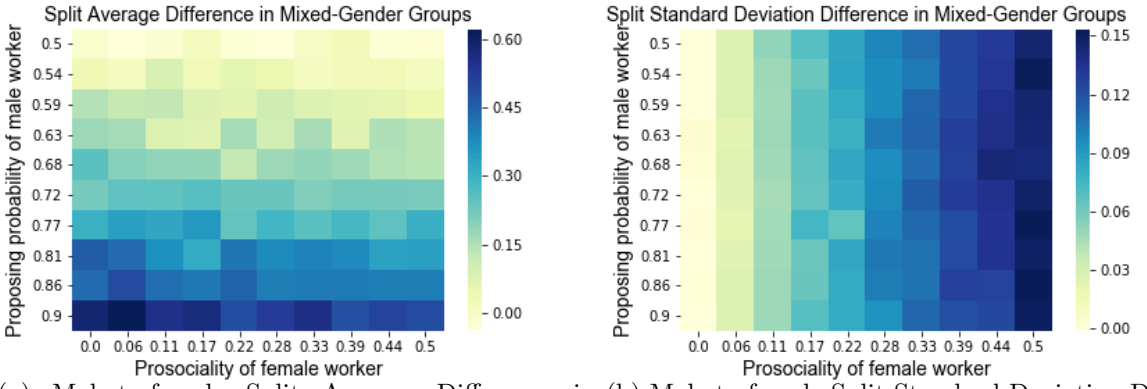
In the simulation, we assume that $\bar{r}_{male} = \bar{r}_{female} = 0$ (i.e., both female and male workers have outside options 0) and $x_{male} = 0$ (i.e., male workers are always not prosocial).[6] We vary x_{female} from 0 to 0.5. We also vary p_{male} from 0.5 (i.e., male and female workers have same bargaining power) to 0.9 (i.e., male workers have much higher bargaining power). For each set of parameters (i.e., $(x_{male}, x_{female}, p_{male}, \bar{r}_{male}, \bar{r}_{female})$, we randomly draw $1,000$ samples of $(\alpha_{male}, \alpha_{female}, p_{male}, r_{male}, r_{female})$ and report the average split of male workers in mixed gender groups, the standard deviations of splits in all-male and all-female groups. Figure A.1 shows the simulation results.

Panel (a) in Figure A.1 shows the difference between male splits and female splits in the mixed-gender groups with respect to female workers' prosociality and male workers' bargaining power (i.e., the probability of proposing for male workers). The upper-left corner of the graph shows that, if female and male workers are equally prosocial and they have the same proposing probability, female and male workers' split difference is precisely at 0. The vertical direction of the graph shows that the difference of average splits becomes larger and larger when the male workers have higher proposing probability (i.e., higher bargaining power). However, the horizontal direction of the graph shows that the difference of average splits does not change much if females become more prosocial. This suggests that, in order for us to rationalize the data where male workers have higher bargaining outcomes than female workers, we need male workers to have higher bargaining power.

[5] Bargaining power can also be represented by the outside options of female and male workers (r_{male} and r_{female}). Our simulation results hold if we use outside options to represent bargaining power instead of the proposing probability.

[6] Our simulation results hold qualitatively if we use different outside option levels or different prosociality levels for male workers.

Figure A.1 Simulation of Male and Female Bargaining Outcomes



(a) Male-to-female Split Average Difference in Mixed-Gender Groups

(b) Male-to-female Split Standard Deviation Difference in Same-Gender Groups

Note: This figure presents the average splits between male and female workers in mixed-gender groups as well as the average difference between standard deviations of male and female splits in same-gender groups for different prosociality levels as well as different proposing probabilities. Each point represents $1,000$ simulations.

Panel (b) in Figure A.1 shows the difference between the standard deviations of male workers' and female workers' splits in the same-gender groups with respect to female workers' prosociality and male workers' bargaining power. Again, the upper-left corner in the heat map is 0, suggesting that, when female and male workers have the same bargaining power and prosociality, they have the same standard deviation of splits in the same-gender groups. The vertical direction of the graph shows that the difference of standard deviations does not change with respect to one gender's bargaining power. However, the horizontal direction of the graph shows that this difference changes dramatically when one gender becomes more prosocial. This suggests that, in order to match the data that female workers have much lower standard deviation of splits in the same-gender groups, we need female workers to be more prosocial than male workers.

In summary, this simulation numerically shows that female workers need to have lower bargaining power and higher prosociality to rationalize the two aforementioned observations in Figure 7.

B. Proof of Proposition 1:

Let us first analyze the decision by the responder given an offer s. If $s \geq 0.5$, the utility of accepting s for the responder is $u_r(s) = s - \beta_r(2s-1)$. If $\beta_r < 0.5$, $u_r(s)$ is always greater than or equal to 0.5, which means that $u_r(s)$ is always greater or equal to r_r. If $\beta_r \geq 0.5$, we need $s \leq \frac{r_r - \beta_r}{1-2\beta_r}$. Since $s \leq \frac{r_r-\beta_r}{1-2\beta_r}$ is bounded above by -0.5 for $\beta_r \in [0.5, 1]$ and $r_r \in [0.0, 0.5]$, there is always no trade when $\beta_r \geq 0.5$.

Similarly, if $s < 0.5$, the utility of the responder is $u_r(s) = s - \alpha_r(1-2s)$, which is greater than the outside option if $s - \alpha_r(1-2s) \geq r_r$, or $s \geq \frac{r_r+\alpha_r}{1+2\alpha_r}$. This means that the responder only accepts the offer if the offer is big enough.

Given the responder's reaction, we can compute the proposer's exact optimal strategy since this is a complete-information game. First notice the fact that $\beta_s \leq 1$ requires that proposer to at most offer $s = 0.5$. This is because if $\beta_s \leq 1$, the proposer will always benefit by offering $s = 0.5$ instead of $s > 0.5$. Moreover,

notice that if $s > 0.5$ will be accepted by one responder, then $s = 0.5$ will always be accepted by the same responder. Therefore, when the proposer is prosocial enough $\beta_s > 0.5$, he prefers offering $s = 0.5$.

If the proposer is not prosocial enough, he will offer $s < 0.5$ and his utility is $u_s = (2\beta_s - 1)s + (1-\beta)$. Since the proposer can only offer when his share from the game is greater than his outside option, he will only offer s if $(2\beta_s - 1)s + (1-\beta_s) \geq r_s$, which is $s \leq \frac{1-\beta_s - r_s}{1-2\beta_s}$. This means the offer has to be small enough for the proposer to benefit from the game. Therefore, combing this with the responder's action, the proposer will always offer the lowest possible offer, which is $\frac{r_r+\alpha_r}{1+2\alpha_r}$ if the offer is lower than the proposer's break-even point $\frac{1-\beta_s-r_s}{1-2\beta_s}$. In other words, if $\frac{1-\beta_s-r_s}{1-2\beta_s} \geq \frac{r_r+\alpha_r}{1+2\alpha_r}$, $s = \frac{r_r+\alpha_r}{1+2\alpha_r}$. Note that $\frac{1-\beta_s-r_s}{1-2\beta_s} \geq 0.5$ for $\beta_s < 0.5$ and $\frac{r_r+\alpha_r}{1+2\alpha_r} \leq 0.5$, therefore, the condition is always satisfied.

In summary, since $u_s = (2\beta_s - 1)s + (1-\beta)$ is decreasing in s if $\beta_s < 0.5$. This means the proposer will offer the smallest possible $s < 0.5$ when $\beta_s < 0.5$, which is $\frac{r_r+\alpha_r}{1+2\alpha_r}$. If $\beta_s \geq 0.5$, the proposer will offer $s = 0.5$, and this offer is only accepted by the responder if $\beta_r < 0.5$. ■

C. Proof of Lemma 1:

We first prove that agent i's payoff from the game is weakly increasing in p_i, i.e., his probability of being the proposer. This is intuitive since if an agent is a proposer, he will always get 0.5 ($\beta_s \geq 0.5$) or above ($\beta_s < 0.5$) when there is a trade. This is because $\frac{r_r+\alpha_r}{1+2\alpha_r}$ is bounded above by 0.5 for all feasible primitives. However, if an agent is a responder, when there is a trade he will at most get 0.5. Therefore, an agent's average payoff increases when he is proposer compared to a responder. The weak increase happens when there is no trade regardless whether the agent is a proposer or a responder.

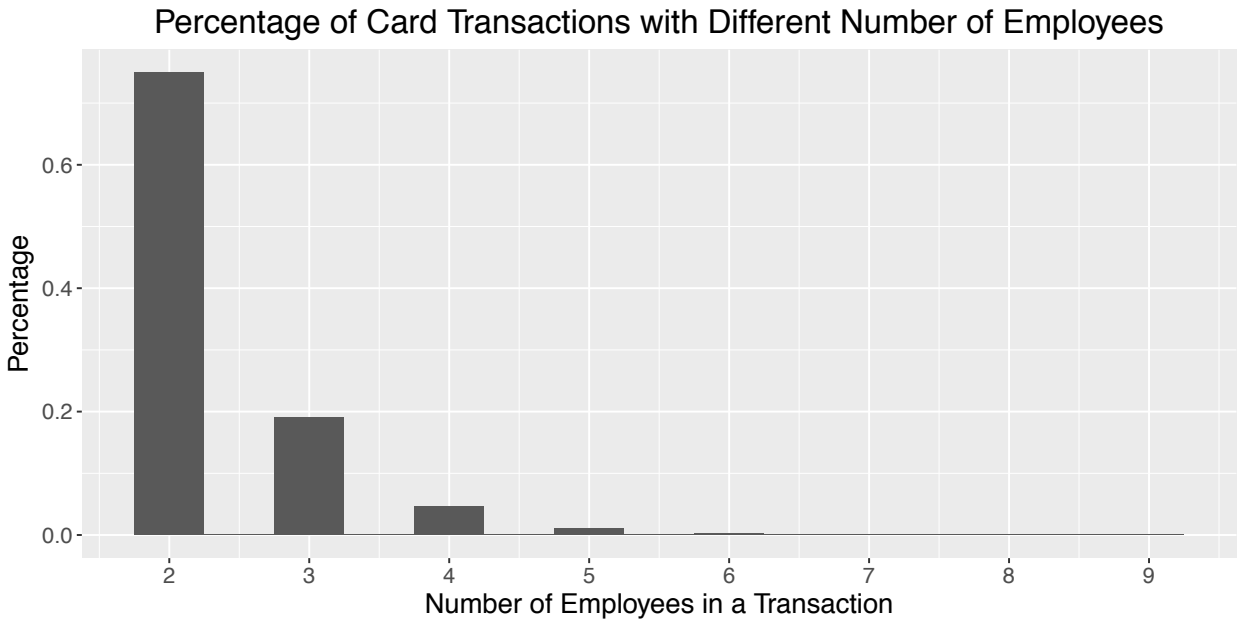
Second, let us prove that if an agent i has higher outside option r_i, he will receive a higher outcome when bargaining with a pool of agents. This is also intuitive. When the agent i is bargaining with other agents, regardless whether he is a proposer or a responder, his split is bounded below by r_i. Therefore, when r_i increases, the lower bound of the bargaining outcome increases. Since r_i does not change the bargaining outcome except setting the lower bound, the agent will get a higher split when r_i increases.

Third, let us show that the average outcome is weakly increasing in α_i. Notice that when agent i is the proposer, his payoff does not depend on α_i. If the agent is a responder, his average payoff is increasing in α_i since he will be willing to only accept a higher offer when α_i increases.

Last, let us show that the average outcome is weakly decreasing in β_i. Notice that when one agent i is a proposer, he will get on average less payoff, if $\beta_i \geq 0.5$ versus $\beta_i < 0.5$. If the person is a responder, he will have some probability to get 0.5 if $\beta_i > 0.5$. And he will not get 0.5 if $\beta_i \geq 0.5$. Since 0.5 dominates all of agent i's payoffs regardless of the opponent when the agent is a responder, the agent's average payoff decreases when β_i increases. ■

D. Auxiliary Graphs and Tables

Figure D.1 Distribution of Card Team Size



Note: This figure shows the distribution of card team size. As we note in the text, we use only those card teams with two workers.