Performance and Information: 
The Role of Organizational Demography*

Iris Bohnet† and Farzad Saidi‡

February 7, 2011

Abstract

Inspired by insights from organizational demography where the demographic mix determines the available set of similar others to learn from, we provide a theoretical framework and experimental evidence on how subsequent informational differences may translate into performance differences in competitive environments. Focusing on the gender imbalance in positions of leadership as our application, we employ binary information conditions with one group having a smaller sample of information available than the other. Such an informational disadvantage leads to a less clear understanding of how effort impacts tournament outcomes, and lower performance. A gap in performance due to differences in information conditions is compatible with labor market data, suggesting that gender pay gaps are particularly prevalent in the most competitive positions in male-dominated industries. (JEL C91, D81, J16, M50)

---

*We wish to thank David Cesarini, Armin Falk, Guillaume Fréchette, Xavier Gabaix, Andy Schotter, and Richard Zeckhauser, along with seminar audiences at Universitat Autònoma de Barcelona (MOVE Workshop) and Harvard Kennedy School, for helpful comments. Financial support from the Women and Public Policy Program and the Women’s Leadership Board at Harvard Kennedy School is gratefully acknowledged.

†Harvard University, Kennedy School of Government, 79 John F. Kennedy Street, Cambridge, MA 02138. iris.bohnet@harvard.edu.

‡New York University, Department of Economics, 19 W. 4th Street, New York, NY 10012. saidi@nyu.edu.
1 Introduction

People tend to relate to and learn from similar others. Role models and mentors typically have the same demographic characteristics as their mentees (e.g., Holmes and O’Connell 2007; Ibarra 1992, 1993; Ragins 1999 for a review). Thus, the demographic mix in an organization determines the available set of comparable others to learn from. Information derived from a smaller sample will likely be less precise, i.e., have a higher variance, than when derived from a larger sample, creating an informational disadvantage for members of smaller groups.

This paper shows how such informational differences can translate into differences in performance. Our theoretical model and experimental design are inspired by labor market reality where various groups are underrepresented in positions of leadership. Most noticeably, women only hold a small fraction of leadership positions in the corporate world (Bertrand and Hallock 2001). For example, at the Fortune 500 companies in 2010, 2.4 percent of the CEOs, 14.4 percent of the executive officers, and 15.7 percent of the board members were female according to a Catalyst Census.\footnote{http://www.catalyst.org/publication/132/us-women-in-business}

Informal accounts suggest that the scarcity of senior colleagues of the same sex puts female junior managers at a disadvantage: junior women “have inadequate information about acceptable (or successful) modes of behavior…” (Blau et al. 2005, p. 177). Similarly, Ibarra (1992: 67) argued that “organizational demography” constrains women’s available set of comparable others to learn from: “Women and minorities usually have a much smaller set of ‘similar others’ with whom to develop professional relationships based on identity-group homophily.”

In other words, the number of similar others is crucial for one’s understanding of the corporate surroundings because career-relevant information may be gender-specific. Furthermore, the career relevance of such gender-specific information is independent of whether the firm might discriminate in any form against any group, or whether returns to information vary between groups.

Based on the above premise, we offer a theoretical framework for why the gender imbalance in organizations might affect managers’ performance, test its implications in the laboratory,
and complement our analysis with suggestive labor market evidence. We focus on two key features of gender imbalance in organizations, namely that one group is in the minority and the other in the majority, and that imbalances are most pronounced in senior positions characterized by competitive work environments where managers are involved in promotion tournaments with substantial uncertainty about how effort translates into rewards. We argue that the extent of perceived uncertainty depends on the number of similar others available to learn from. Such an information disadvantage will lead women – or, more generally, members of smaller groups – to have a less clear understanding of how effort translates into tournament outcomes. The more a person attributes winning to randomness, the less effort she will exert and the worse she will perform in such competitive settings.

In our experiment, we abstract from gender or any other demographic characteristics, and assign information conditions randomly, thus creating a better informed group receiving information from a large sample of data and a worse informed group receiving information from a small sample. Performance is measured in a real-effort task, namely by the number of words found in a word find task where people are assigned to pairs and compete against their anonymous counterpart for a tournament prize. Tournament outcomes are determined by both individual effort and a random bonus component. Depending on the information condition, a person is well or poorly informed on the potential impact of the random component on tournament outcomes.

Experimental participants receiving less precise information on the effort-reward relationship indeed perceived the variance of their bonus component to be larger, and subsequently performed worse than better informed participants. Furthermore, the underperformance of the informationally disadvantaged group, if anything, became more pronounced when the competing group was better rather than equally well informed. This suggests that informational precision potentially affects performance through two channels: by affecting how well agents understand the effort-reward relationship and by creating informational differences that additionally discourage the disadvantaged from exerting effort.

This paper contributes to the theoretical question of how uncertainty impacts effort. Earlier work on the relationship between the information structure and individual performance in competitive environments is not conclusive. While in our setup informational differences
are inspired by managerial labor market reality where the majority group is likely better informed than the minority group, Bull et al. (1987) and Freeman and Gelber (2010) varied the amount of information on the past performance of competitors that their experimental subjects had available. For a hypothetical effort task, Bull et al. (1987) found that subjects who were informed of their counterparts’ decisions after each round exhibited more variance in their decisions, and exerted less effort than those who did not receive any information. In contrast, Freeman and Gelber (2010) who used a real-effort task (mazes) found that providing more information on the historical performance of competitors led to higher effort.

The core motivation of our paper is related to earlier work on the relative number of similar others in a group in sociology and political science, often referred to as “critical mass theory,” where it has been argued that the relative size of a group determines how it is perceived and behaves (Kanter 1977, Dahlerup 1988). Indeed, women’s performance evaluations have been found to correlate with the relative proportion of women in a group (e.g., Sackett et al. 1991, Pazy and Oron 2001). In addition to affecting perceptions and providing information, a larger share of women may affect the work environment directly and make it more responsive to women’s needs, thereby increasing women’s productivity (see, for instance, Bertrand et al. 2010). Furthermore, in recent experimental work in economics, women’s performance has been shown to depend on the gender composition of a group such that women were more willing to compete and cooperate with other women (e.g., Gneezy et al. 2003, Greig and Bohnet 2009, and, for a review, Croson and Gneezy 2009).

In our framework, results are not specific to gender differences in performance, but may speak to performance differences between any majority and any minority group as group affiliation determines the respective information conditions. The remainder of the paper is organized as follows. Section 2 offers a theoretical framework, Section 3 presents the experimental design, and the experimental results are reported in Section 4. In Section 5, we discuss some empirical evidence from labor market data, and Section 6 concludes.
2 Theoretical Framework

We model effort choices among competing workers using a tournament-theoretical framework (Lazear and Rosen 1981). We propose a setting in which two agents compete against each other. They belong either to the better informed, large group or the less informed, small group. Each agent can control the mean of the output distribution $\mu$ by means of effort exertion, which is costly ($C(\mu) > 0$). Furthermore, a stochastic luck component $\varepsilon$ is realized. This leads to the following observable output:

$$q_i = \mu_i + \varepsilon_i, i = L, S$$  \hspace{1cm} (1)

where $L$ and $S$ stand for two different agents who belong to the large and small group, respectively.

In this rank-order tournament, agent $L$ differs from agent $S$ in the perceived distribution of the luck component. The agents may or may not be aware of these differences. We distinguish two cases in both the model and the experiment. In the first case, agents are not aware of each other’s specific information conditions. In the second case, we allow for learning, and discuss the implications for further dynamics.

2.1 Heterogenous Beliefs

Our general setup is akin to that in Lazear and Rosen (1981). The specificity of our model lies in the beliefs of $L$ and $S$:

- Agent $L$ believes all luck components to be independent s.t. $\varepsilon_{L,L} \sim N(0, \sigma^2_L)$ and $\varepsilon_{S,L} \sim N(0, \sigma^2_L)$ where the variance reflects the precision of the perceived relationship between effort and pay, and $\varepsilon_{x,y}$ is agent $y$’s belief over agent $x$’s luck component.

- Agent $S$ believes the respective luck components to be independent s.t. $\varepsilon_{S,S} \sim N(0, \sigma^2_S)$ and $\varepsilon_{L,S} \sim N(0, \sigma^2_S)$ where $\sigma^2_L < \sigma^2_S$.

- However, the principal, i.e., the firm, believes that both agents perceive their luck components to be independent and normally distributed with zero mean and variance $\sigma^2_L$. This is a simplified way of saying that agent $L$’s beliefs are better aligned with the
actual effort-reward relationship whereas agent \( S \) underestimates the extent to which effort translates into rewards.\(^2\)

Furthermore, the following general assumptions apply:

- We consider a single time period.
- We assume that occupational sorting has already taken place, i.e., the agents have been with the firm for a while, and now the agents’ supervisor, the principal, is designing the performance pay scheme (a tournament).
- The cost function \( C(\cdot) \) is quadratic and not a source of agent heterogeneity.
- The firm acts in a perfectly competitive market.
- Denote the tournament prize spread by \( \Delta W \equiv W_1 - W_2 \) where \( W_1 \) and \( W_2 \) are the winner and loser prizes, respectively.

Before we move to the analysis of the game, we introduce the notion of a performance gap in this tournament.

**Definition**  A performance gap exists iff \( \mu^*_L \neq \mu^*_S \).

If equilibrium effort choices differ between agents, and \( \Delta W > 0 \), this results in a pay gap as different levels of effort exertion imply different probabilities of winning the tournament. The probability of winning the tournament is equal to:

\[
\text{prob}_i (q_i > q_j) = \text{prob}_i (\mu_i - \mu_j > \varepsilon_j - \varepsilon_i) \tag{2}
\]

where \( i \neq j \).

Given the above-mentioned distributional assumption, \( E[\varepsilon_j - \varepsilon_i] = 0 \), with the variance depending on the beliefs of the respective player (\( L, S \)). If \( \sigma^2_L < \sigma^2_S \), agent \( S \), belonging to the minority group, underestimates the impact of effort on actual pay due to limited peer information. As we shall see, the equilibrium investment in effort of \( S \) and \( L \) is a function

\(^2\)This approach seems similar to the model of Lundberg and Startz (1983) – however, we do not assume statistical discrimination on the part of the employer but heterogenous beliefs among employees depending on their group affiliation. Moreover, our model is different in that both agents believe the distribution of the luck component to be the same for everyone.
of prob\(_S\) (\(q_S > q_L\)) = g (\(\mu_S - \mu_L\)) and prob\(_L\) (\(q_L > q_S\)) = h (\(\mu_L - \mu_S\)). Here, \(g (\mu_S - \mu_L)\) and \(h (\mu_L - \mu_S)\) are the probability density functions of a normal distribution with zero mean and variance \(2\sigma^2_S\) and \(2\sigma^2_L\), respectively, so this is the channel through which the perceived variance of the luck component impacts effort choice.

From Lazear and Rosen (1981) we know that a decrease in the precision with which the agents understand the effort-reward relationship leads to reduced effort provision by risk averse agents. In the case of risk neutrality, however, this effect would be offset by an increase in the prize spread assuming homogenous agents. Hence, in Lazear and Rosen (1981), for risk neutral agents with homogenous beliefs about the error term, the optimum investment in effort does not vary with the variance of the luck component. Given that in our model we have two types of agents with heterogenous beliefs, this result does not hold. Prize spreads cannot be optimally adjusted for both groups at the same time, and thus, even under risk neutrality, we expect a worse understanding of the effort-reward relationship to result in less effort. Accordingly, we assume risk neutrality, and continue with the analysis of the game.

Unlike agent \(S\), agent \(L\) perfectly observes the variance (this assumption can be relaxed as we simply require \(L\)'s belief to be closer to reality than \(S\)'s belief), but both agents are unaware of each other's beliefs. In this setup, a performance gap follows from Lazear and Rosen (1981). The proof of the following proposition is in Appendix A.

**Proposition 1** If \(\sigma^2_L < \sigma^2_S\) and both agents are unaware of each other's beliefs, a performance gap exists s.t. \(\mu^*_S < \mu^*_L\).

Agent \(S\) does not invest efficiently and \(\mu^*_S < \mu^*_L\), i.e., the equilibrium investment in effort of \(L\) is greater than that of \(S\). This is due to the fact that \(S\) underestimates the responsiveness of pay to effort whereas \(L\) knows the correct distribution of the luck component. Hence, there is a performance gap in equilibrium, and \(S\) is less likely to win the tournament than \(L\).

Note that, for the above result to hold, we do not require the firm to adjust the prize spread \(\Delta W\) optimally for any one of the agents (cf. proof of Proposition 1): as can be seen in (A.11), a performance gap persists irrespective of the exact prize spread as long as the latter is the same for both agents. For our experiment, this implies that the theory can be tested using a fixed prize scheme for the tournament.
2.2 Stepwise Convergence of Beliefs

So far, we have assumed that the agents have different beliefs, but are unaware of these differences, which is equivalent to each of them assuming identical information sets. Now consider the case where only agent $S$ assumes the distribution of the luck component to be homogenous for both agents, i.e., agent $S$ still overestimates the variance of the luck component, but this time agent $L$ is aware of it. This assumption can be defended on the following grounds. Interpret $L$ and $S$ as gender depending on the gender composition at the firm. Suppose the principal as well as the majority of the firm’s employees are male, and they recognize women’s beliefs. If male agents are faster at incorporating this insight into their decision-making than the principal, for example, because the principal faces institutional constraints in designing contracts and thus cannot easily give up “gender-blindness,” then one may assume only the male agent $L$ knows both perceived distributions of the respective luck components, which are independent such that $\varepsilon_{L,L} \sim N(0, \sigma^2_L)$ and $\varepsilon_{S,L} \sim N(0, \sigma^2_S)$ where $\sigma^2_L < \sigma^2_S$. We now show that, given an upper bound on $W_2$ (the loser prize), a performance gap persists, the proof of which is in Appendix A.

**Proposition 2**  If $\sigma^2_L < \sigma^2_S$ and only $L$ is aware of $S$’s beliefs (but not vice versa), a performance gap exists s.t. $\mu^*_S < \mu^*_L$ as long as $W_2 < \sigma_L \left( 4\sigma_L \ln \frac{\sigma_S}{\sigma_L} - \sqrt{\pi V} \right)$, and both agents invest inefficiently.

The result that this time both agents do not invest efficiently is due to the fact that $L$ knows the correct distribution of the luck component, but also incorporates $S$’s belief in his best response. However, the actual performance gap has decreased compared to the baseline case in Proposition 1 as the distance between $h(x)$ and $g(x)$ is maximized for $x = 0$.

Finally, assume that $S$-players recognize that they have a noisier notion of the effort-reward relationship, so they understand that their estimate of the variance of the luck component is actually an upper bound. Then, denote by $\tilde{g}(\cdot)$ the pdf of a normal distribution with some lower variance, say $\tilde{\sigma}^2_S < \sigma^2_S$. Furthermore, assume that the corresponding upper bound on the loser prize holds, that is $W_2 < \sigma_L \left( 4\sigma_L \ln \frac{\tilde{\sigma}_S}{\sigma_L} - \sqrt{\pi V} \right)$. Now, if $L$-players are unaware of this learning process on the part of the $S$-players, the performance gap is char-
acterized by the difference $h(\mu_L^* - \mu_S^*) - \tilde{g}(0) > 0$ where $\tilde{g}(0) > g(0)$, so the performance gap remains but becomes smaller. In the course of learning dynamics (e.g., $L$-players could correct their estimate of $\mu_S^*$ in the next round and thus increase their effort), the tendency will be for the performance gap to shrink the more $L$-players and $S$-players learn about each other’s notions of the effort-reward relationship – until the agents’ beliefs converge and the performance gap eventually vanishes.

We conclude that, given heterogenous beliefs and all else equal, overestimating the variance of the luck component leads to underinvestment in effort and – in expectation – to reduced pay. Hence, the actual incentive effect of a competitive compensation scheme (in this case, a rank-order tournament) is deemed to be weak for agents who have less precise information on the relationship between effort and pay. This has the following experimentally testable implications:

**Implication 1** Informationally disadvantaged agents ($S$-players) perform worse than informationally advantaged agents ($L$-players).

**Implication 2** The performance gap between the informationally disadvantaged ($S$-players) and the informationally advantaged ($L$-players) group is due to differences in the perception of the variance of the luck component.

As shown in Proposition 2 and the subsequent discussion above, the more aware the agents are of the restrictions of their information sets, the smaller is the gap in effort exertion, as the informationally advantaged group will decrease and the informationally disadvantaged group will increase its effort. This leads to:

**Implication 3** The performance gap between the informationally disadvantaged ($S$-players) and the informationally advantaged ($L$-players) group is largest if the agents are not aware of the informational differences, i.e., if they assume identical information sets, and smallest if the agents are aware of these differences.
3 Experimental Design and Procedures

The goal of our experiment is to test Implications 1 to 3. In particular, we examine whether individuals who make inferences about the effort-reward relationship based on smaller samples perform worse in a tournament than their counterparts who make inferences based on larger samples. We also control for competing explanations of differential effort choices, such as risk preferences.

Our experiment involved a tournament where subjects competed in pairs of two in a word find task. Subjects were confronted with a matrix containing letters: most letters appeared in random order but some formed words. Provided with a list of 20 words, subjects could find these words in the matrix by marking sequences of letters horizontally, vertically, or diagonally. The subjects’ task was to mark as many words as possible in three minutes. For every correct word marked, subjects received 10 points. In addition to their task score, subjects received a bonus, their luck component. Bonuses were randomly drawn from a hat containing 18 different bonus values which, in turn, were drawn from a uniform distribution. In every round, subjects were informed of the lowest and the highest possible bonus value in the distribution, plus they were given a randomly drawn sample of actual bonus values in the hat (without replacement). The sample was either large (for $L$-players), namely 12 out of 18 possible bonus values, or small (for $S$-players), namely 3 out of 18 bonus values, mimicking the majority and minority groups’ respective experiences in the workplace, with members of the minority group having a less precise understanding of the effort-reward relationship than members of the majority group. The final score was equal to the sum of the task score and the value of the bonus. The subject with the higher final score in a pair won the tournament and received a large prize (10 tokens) whereas the subject with the lower score received a small prize (2 tokens). One token was worth $1.

To examine the role of individuals’ knowledge of one’s competitor’s information condition, we created two treatments: the baseline where subjects assumed identical information conditions, and a second condition where they were aware of informational differences. To assure that subjects indeed believed that information conditions were identical, we in fact created identical information conditions, and informed all subjects of that. Thus, in the baseline
conditions, each pair received information on either the small or the large sample of bonus values in the hat, and this was common knowledge. We refer to the baseline conditions as same-type tournaments. In the second condition, one person per pair received information on the large sample and the other person in the pair received information on the small sample, and this was common knowledge. We refer to this as mixed-type tournaments.

After an initial practice round, the task was repeated four times (with a different letter matrix and word list in every round). Subjects remained in the same pair for the duration of the experiment. In rounds 1 and 2, subjects were confronted with a wide range of possible bonus values from 0 to 100. In rounds 3 and 4, we decreased the range of bonus values by limiting them to be between 30 and 70. Performance is likely responsive to both experience with the task and the range of potential bonus values. At the end of each round, subjects were informed of their task score, their final score, their counterpart’s final score, and the tokens they won. They did not receive information on their counterpart’s task score, and were thus unable to determine with certainty whether they won/lost because of their counterpart’s performance or the randomly drawn bonus.

Besides the subjects’ performance on the word find task, we collected three additional pieces of information. After subjects had been informed of the highest and the lowest possible bonus values, and had seen their random draw of sample values, we asked them to guess the mean bonus value and report a 90% confidence interval for their guess. Reporting a larger confidence interval is equivalent to perceiving a larger variance. Furthermore, after the completion of the main experiment, we had subjects participate in a risky choice task to measure their risk preferences. We employed the task introduced by Holt and Laury (2002) with identical incentives. Finally, the study concluded with a short questionnaire collecting demographic information. Subjects were paid for their performance in all four rounds and, in addition, received their earnings from the risky choice task. Average earnings, including a $10 show-up fee, were about $36 for a study that lasted one hour.

We ran the experiments in the Harvard Decision Science Laboratory in the spring of 2010. 206 subjects participated in nine sessions with 22 or 24 subjects in each of them.
4 Experimental Results

We first report descriptive statistics. Then, we examine our central predictions, Implications 1 and 2: agents who are provided with a smaller sample of information on potential bonus values (S-players) perform worse compared to their counterparts with more precise information (L-players), and this effect is due to the perceived variance of the bonus component. Finally, we examine Implication 3, i.e., whether the performance gap varies depending on whether subjects assume identical information sets (same-type tournament), or are aware of informational differences (mixed-type tournament).

On average, subjects found 10.13 words out of a total of 20 words available in a given letter matrix (with a standard deviation of 3.88). Women and men differed slightly in their performance, with women marking 10.35 words correctly and men finding 9.78 words on average ($p < 0.05$). This difference was entirely driven by performance in the first round, and women and men did not differ at all in their performance in the remaining three rounds.

Figure 1 presents the distribution of the number of words people found in the pooled sample. Typical outcomes range from 5 to 16 words found per matrix. Four participants, i.e., roughly 2% of our subjects, found the maximum of 20 words in at least one round.

Examining Implication 1, we first review differences in the mean number of words found by L- and S-players. Table 1 reports the data pooled across both treatments (cf. first panel) and separately for each treatment condition (cf. second and third panels). Within each panel, in the first row we present performance levels aggregated over all four rounds, in the second for the wide-range rounds (rounds 1 and 2), in the third for narrow-range rounds (rounds 3 and 4), and in the last row for the rounds where people had already gained one round’s experience within a given range condition (rounds 2 and 4). L-players found about one word more than S-players on average ($p < 0.01$).

Looking at same-type and mixed-type tournaments separately in panels 2 and 3, the performance gap between L- and S-players is exacerbated in mixed-type tournaments. In
panel 3, L-players found 1.3 words more than S-players on average, which corresponds to an increase of one-third of a standard deviation. Experience increased the performance gap to 1.8 words in rounds 2 and 4. On average (cf. first row of panels 2 and 3), the performance gap was mainly driven by S-players who performed significantly worse when competing against L-players rather than against identically informed counterparts ($p < 0.05$). In contrast, the better informed group was not differentially affected by the two treatment conditions.

We make two more observations: people’s absolute performance improved over time, suggesting learning. In particular, performance levels increased as the bonus range decreased (as can be seen from comparing rounds 3 and 4 to rounds 1 and 2), which is in line with theory.

To more precisely measure these effects, and examine whether the L- and S- players’ differential effort choices were driven by their perceptions of the variance of the bonus component (Implication 2), we run regressions. We approximate perceived variance by subjects’ reported confidence intervals of their estimates of the mean bonus. As the range of possible bonus values differs between the first two and the last two rounds, we use a standardized measure of the reported confidence interval between 0 and 1, the relative range per round, to make it comparable across rounds. This yields Perceived range with a mean of 0.46 and a standard deviation of 0.25. On average, L-players perceived the range of bonus values to be 0.41 whereas S-players reported the range to be 0.50 (with perceived, normalized mean bonus values of 0.51 and 0.45, respectively). To more easily interpret the effect on performance, we include $1 - \text{Perceived range}$ in our regressions as the smaller a person perceives the variance to be, the more effort she should exert.

A simple test of the theory would be to estimate the impact of $1 - \text{Perceived range}$ on task performance, and we would hypothesize that impact to be positive. However, the reported confidence intervals used to construct Perceived range are the result of some intellectual effort, and are therefore subject to endogeneity with respect to task performance. Indeed, the $p$-values from Durbin-Wu-Hausman tests (reported at the bottom of Table 2b) indicate

\footnote{Note that, theoretically, different perceptions of the mean bonus value do not impact effort choice as all players know that they receive draws from the same urn and the values of the random components cancel each other out in the expected utilities (cf. (A.1) and (A.2)). Furthermore, all of the regressions presented in this paper are robust to including the perceived mean bonus values, which have no explanatory power for performance.}
that simple OLS estimates would not be consistent. Thus, we test Implications 1 and 2 using two-stage least squares regressions. A valid first stage in line with Implication 2 requires that the experimental treatment of receiving a large sample of potential bonus values, which is exogenous by design, have a strong, positive impact on \(1 - Perceived\ range\). Then, for the second stage, we hypothesize that the projected values of \(1 - Perceived\ range\) are associated with higher task scores, which is the essence of Implication 1. Tables 2a and 2b report the results of the first and the second stage, respectively.

[Insert Table 2a about here]

[Insert Table 2b about here]

Our results support Implications 1 and 2. As can be seen in Table 2a, receiving a larger sample of information implies attaching a lower variance to the relationship between effort and pay, which – as suggested by Table 2b – encourages effort and leads to higher task performance. The first stage is strong, and the effect of receiving a larger sample on \(1 - Perceived\ range\) is, on average, equal to two-fifths of a standard deviation.\(^4\) The results hold controlling for risk aversion, the corresponding coefficient of which has the predicted negative sign in the second stage (yet not significantly so).\(^5\)

Having shown that informational differences impact performance through the perceived variance of the bonus component, we now test Implication 3 that the performance gap is smaller when informational differences are public information. We thus hypothesize that the difference in perceived variance is smaller in the mixed-information condition than in

\(^4\)While the positive effect of the narrow range of possible bonus values on performance in rounds 3 and 4 in the second stage is in line with theory, the negative coefficient of the indicator for these rounds on \(1 - Perceived\ range\) in the first stage (cf. Table 2a) might be confusing. Subjects reported significantly smaller perceived absolute ranges in rounds 3 and 4 as opposed to rounds 1 and 2 (namely, 20.91 compared to 39.81). However, the perceived relative ranges reverse as we divide perceived ranges by the size of the actual given ranges to make the measures comparable. We cannot use the absolute range in all our variables, although the respective coefficient in unreported regressions suggests a significantly smaller perceived absolute range in rounds 3 and 4, as that correlation is partly mechanical. Still, the correlation between the perceived relative and absolute ranges is 0.692. We therefore use the relative range in all regressions to avoid mechanical correlations. Also, our results are virtually unchanged if one drops the indicator for rounds 3 and 4 from the right-hand side of the regression specifications.

\(^5\)The number of observations drops once we include the measure for risk aversion as some responses were either incomplete or invalid by reasonable standards. However, all results in this paper are robust to dropping these observations irrespective of whether we control for risk aversion or not.
the identical-information condition. To test this, we interact the indicators for the mixed-type tournament ("Mixed") and receiving a large sample in the last two columns of Table 2a. Rather than being negative, the interaction effect is not significant. As we have seen in Table 1, the descriptive results also cast doubt on the validity of Implication 3 and, if anything, suggest the opposite, namely that the performance gap widens when informational differences are public information. Against this background, we interpret the finding that the performance gap, in contrast to the theoretical prediction, is not reduced in the mixed-information condition as evidence that informationally disadvantaged parties are generally likely to underperform in competitive settings.

Our study was motivated by female performance in male-dominated environments. We argued that women as the minority are at an informational disadvantage if the number of similar others available to learn from determines the precision of their understanding of the effort-reward relationship. In the next section, we complement our analysis with labor market evidence that gender composition matters for the gender pay gap under performance pay schemes.

5 Some Labor Market Evidence – A Discussion

In this section, we examine whether the gender composition on the job impacts gender differences in performance as captured by pay, and return to a data set analyzed by Manning and Saidi (2010). We decompose the data further to focus on the conditions under which we expect such differences to occur based on our theoretical model and experimental results. Using British labor market data from the 1998 and 2004 editions of the Workplace Employment Relations Survey (WERS), we test whether gender gaps in pay under competitive compensation schemes are more pronounced in male-dominated industries, particularly in (senior) management positions.

Extensive information on the data and the empirical strategy were given in Manning and Saidi (2010), and we summarize the main components of our analysis – including the empirical strategy and identification assumptions – in Appendix B. We run quantile regressions to examine how the difference between female and male returns to working under performance pay
schemes develops for three different gender compositions on the job: male-dominated, mixed, and female-dominated. In Figure 2, we plot the respective difference-in-difference estimates along with the confidence intervals in the regressions for the male-dominated environments. We find that, for the most competitive jobs at the upper end of the wage distribution, the gender pay gap under performance pay schemes is most pronounced in male-dominated environments. In contrast, this effect diminishes in mixed competitions as women climb up the corporate ladder whereas in female-dominated environments there is basically no such pay gap.

We view these results as indicative of the importance of the gender balance in a group: the gender pay gap under performance pay schemes increases the more male-dominated the job environment becomes. Also, we do not find a reversal of the gender pay gap in female-dominated environments, suggesting that, in addition to informational differences between the majority and the minority group, other gender-specific factors contribute to wage gaps.

The performance pay gap in male-dominated compared to other environments is increasing across the wage distribution. This is in line with Proposition 2 where we have shown that an upper bound on the loser prize $W_2$ is a sufficient condition for a performance gap in our setup. In the data, the loser prize is likely decreasing across the wage distribution as tournament schemes tend to become harsher at the top of the wage distribution. An extreme example is the up-or-out system implemented by firms in very competitive industries (e.g., investment banking) such that candidates below a certain percentile in the performance ranking are dismissed ($W_2 = 0$) while the remaining employees are promoted. We conclude that, at least with respect to the development of the gender pay gap under performance pay schemes across the wage distribution, our theory — for which we have provided experimental evidence in this paper — seems to be in accordance with labor market reality.

6 Concluding Remarks

Our paper suggests a potential mechanism for why performance gaps in mixed-sex competitive environments might emerge: when people (have to) derive at least part of their
career-relevant information from similar others, women being in the minority in senior management positions are at a disadvantage. The size of the group of similar others determines how precise the information received is and, thereby, the effort group members exert. This in turn affects the likelihood that women succeed under performance pay schemes, and are eventually promoted in their organizations.

While we purposely created “gender” randomly – by focusing on key characteristics of people’s gendered experiences in the workplace, namely either being a member of a better informed majority or of a less informed minority group – the patterns of behavior found in our experiments are compatible with other field and laboratory observations based on men and women.

Examining the effectiveness of same-sex networks in a professional service firm where only a small minority of women held senior management positions, Ibarra (1993) reports that men reaped greater benefits from their larger same-sex networks than women did. Moreover, holding formal status in the organization constant, male senior corporate executives have been found to earn about twice as much from insider trading as their female colleagues, mainly due to an informational advantage through same-sex networks rather than dispositional factors (Bharath et al. 2009).

Gneezy et al. (2003) as well as Booth and Nolen (2009) present experimental evidence in support of a gender performance gap in mixed or male-dominated competitive environments as compared to same-sex competitions. In line with our findings, gender differences in performance were also driven by women – or, in our case, the informationally disadvantaged group – adjusting their behavior to the different environments: women performed better in same-sex than in mixed-sex competitions while men’s performance was not affected by the gender composition (Gneezy et al. 2003). A similar pattern has been found in performance evaluations in an organization where women were in the minority, namely among officers in the Israeli military: women were evaluated the more positively the larger their relative share in a group was whereas men’s evaluations were invariant to the gender balance in the group (Pazy and Oron 2001).6

6 Differences in the evaluations of men and women based on the gender composition of the group are compatible with a number of explanations, including psychological theories of implicit biases where “seeing is believing” and only a change in numbers can affect people’s stereotypical judgments (e.g., Dasgupta and
Clearly, the gender balance in organizations may affect men’s and women’s productivity through a multitude of channels. For example, a larger share of women in an organization might be correlated with a larger fraction of women in the talent pool of organizationally relevant professions, thereby increasing the firm’s economic benefits of adjusting working conditions to women’s needs (see Bertrand et al. 2010 for a discussion). Accordingly, we would also expect gender performance and pay gaps to narrow as the fraction of women increases – which is compatible with the empirical pattern observed in Section 5.

An increasing number of firms seem to care about the gender balance in their organizations – because they believe in the business case for gender diversity (e.g., Dezso and Ross 2008), fear legal sanctions or reputational harm. While most firms choose incremental strategies to expand the proportion of women in senior management, in 2010 Deutsche Telekom was the first major European company to introduce gender quotas for its upper and middle management, with the goal of having 30 percent female senior managers by 2015.

Our paper contributes to two literatures. It suggests an additional mechanism through which gender differences in performance, pay, and representation in leadership positions can emerge – informational differences due to the relative size of one’s group. Organizational demographics may thus be an important determinant of the productivity, promotion likelihood, and pay of an organization’s employees. In addition, our paper also speaks to the literature exploring the impact of noise in people’s perceptions of the effort-reward relationship on their performance. We provide a new mechanism to manipulate differences in noise building on the simple statistical truth that smaller samples provide less precise information than larger samples, and show that differential perceptions of noise indeed affect performance. This has broader implications. Differences in the sample sizes of role models or, more generally, task-relevant information likely affects a wide range of performance gaps, including in education, sports, politics, the military, and the arts.

Asgari 2004, Beaman et al. 2009), but also theories of statistical discrimination and information asymmetries in economics where the employer is less well informed about the productivity of the minority group than of the majority group (Coate and Loury 1993), or where the minority group has invisible abilities (Milgrom and Oster 1987).

The authors find that the gender balance in senior management is related to firm performance, with company success being positively correlated with the fraction of women in senior management. Note that the gender of the CEO was not related to company performance (Wolfers 2006, Dezso and Ross 2008).

http://news.bbc.co.uk/2/hi/8568066.stm
References


20


# Tables

## Table 1: Differences in Mean Scores

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Rounds</th>
<th>Large sample (N=102)</th>
<th>Small sample (N=104)</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>All</td>
<td>10.637</td>
<td>9.632</td>
<td>1.005***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[3.83]</td>
<td>[3.87]</td>
<td>[0.27]</td>
</tr>
<tr>
<td>All</td>
<td>1 &amp; 2</td>
<td>10.134</td>
<td>9.203</td>
<td>0.931***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[3.59]</td>
<td>[3.62]</td>
<td>[0.36]</td>
</tr>
<tr>
<td>All</td>
<td>3 &amp; 4</td>
<td>11.146</td>
<td>10.063</td>
<td>1.083***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[4.00]</td>
<td>[4.08]</td>
<td>[0.40]</td>
</tr>
<tr>
<td>All</td>
<td>2 &amp; 4</td>
<td>11.280</td>
<td>9.966</td>
<td>1.314***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[3.96]</td>
<td>[4.05]</td>
<td>[0.60]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Rounds</th>
<th>Large sample (N=46)</th>
<th>Small sample (N=48)</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same</td>
<td>All</td>
<td>10.793</td>
<td>10.138</td>
<td>0.656</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[3.85]</td>
<td>[3.90]</td>
<td>[0.40]</td>
</tr>
<tr>
<td>Same</td>
<td>1 &amp; 2</td>
<td>10.231</td>
<td>9.800</td>
<td>0.431</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[3.72]</td>
<td>[3.66]</td>
<td>[0.54]</td>
</tr>
<tr>
<td>Same</td>
<td>3 &amp; 4</td>
<td>11.375</td>
<td>10.479</td>
<td>0.896</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[3.92]</td>
<td>[4.11]</td>
<td>[0.60]</td>
</tr>
<tr>
<td>Same</td>
<td>2 &amp; 4</td>
<td>11.231</td>
<td>10.505</td>
<td>0.726</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[4.13]</td>
<td>[4.05]</td>
<td>[0.60]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Rounds</th>
<th>Large sample (N=56)</th>
<th>Small sample (N=56)</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mixed</td>
<td>All</td>
<td>10.509</td>
<td>9.205</td>
<td>1.304***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[3.82]</td>
<td>[3.81]</td>
<td>[0.36]</td>
</tr>
<tr>
<td>Mixed</td>
<td>1 &amp; 2</td>
<td>10.055</td>
<td>8.696</td>
<td>1.358***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[3.50]</td>
<td>[3.52]</td>
<td>[0.47]</td>
</tr>
<tr>
<td>Mixed</td>
<td>3 &amp; 4</td>
<td>10.964</td>
<td>9.714</td>
<td>1.249**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[4.08]</td>
<td>[4.04]</td>
<td>[0.55]</td>
</tr>
<tr>
<td>Mixed</td>
<td>2 &amp; 4</td>
<td>11.321</td>
<td>9.509</td>
<td>1.812***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[3.84]</td>
<td>[4.00]</td>
<td>[0.53]</td>
</tr>
</tbody>
</table>

**Notes:** In the first two columns, standard deviations are in parentheses. The third column indicates the results of a two-sided difference-in-means test (with standard errors in parentheses) where */**/*** denote significance at the 10%/5%/1% level, respectively.
Table 2a: Determinants of Perceived Variance of Bonus Component (First Stage)

<table>
<thead>
<tr>
<th></th>
<th>Dependent variable: 1—<em>Perceived range</em></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Large sample</td>
</tr>
<tr>
<td></td>
<td>0.091***</td>
</tr>
<tr>
<td></td>
<td>[0.03]</td>
</tr>
<tr>
<td></td>
<td>0.122***</td>
</tr>
<tr>
<td></td>
<td>[0.03]</td>
</tr>
<tr>
<td></td>
<td>0.100**</td>
</tr>
<tr>
<td></td>
<td>[0.03]</td>
</tr>
<tr>
<td></td>
<td>0.111***</td>
</tr>
<tr>
<td></td>
<td>[0.03]</td>
</tr>
</tbody>
</table>

# of observations  | 768 | 622 | 768 | 622 |
F-statistic         | 16.32 | 15.58 | 14.33 | 13.95 |

Notes: */**/*** denote significance at the 10%/5%/1% level, respectively. In the linear regressions, standard errors are given in parentheses, and are clustered at the pair level. Self-reported economic background is scaled from 1 (very poor) to 6 (very rich). Risk aversion is measured on a scale from 1 (very risk loving) to 10 (very risk averse). The mixed-information condition, under which informational differences are public information, is labeled as “Mixed.”
Table 2b: Determinants of Task Performance (Second Stage)

<table>
<thead>
<tr>
<th></th>
<th>Dependent variable: Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1—Perceived range</td>
<td>9.898*</td>
</tr>
<tr>
<td>(endogenous)</td>
<td>[5.72]</td>
</tr>
<tr>
<td>Mixed</td>
<td>-0.734</td>
</tr>
<tr>
<td></td>
<td>[0.54]</td>
</tr>
<tr>
<td>Female</td>
<td>0.559</td>
</tr>
<tr>
<td></td>
<td>[0.48]</td>
</tr>
<tr>
<td>Rounds 2 &amp; 4</td>
<td>1.019***</td>
</tr>
<tr>
<td></td>
<td>[0.22]</td>
</tr>
<tr>
<td>Rounds 3 &amp; 4</td>
<td>2.008***</td>
</tr>
<tr>
<td></td>
<td>[0.58]</td>
</tr>
<tr>
<td>Student</td>
<td>2.028***</td>
</tr>
<tr>
<td></td>
<td>[0.58]</td>
</tr>
<tr>
<td>Economic background</td>
<td>-0.484*</td>
</tr>
<tr>
<td>(1-6)</td>
<td>[0.29]</td>
</tr>
<tr>
<td>Risk aversion</td>
<td>-0.206</td>
</tr>
<tr>
<td>(1-10)</td>
<td>[0.18]</td>
</tr>
<tr>
<td>Constant</td>
<td>3.783</td>
</tr>
<tr>
<td></td>
<td>[3.50]</td>
</tr>
<tr>
<td># of observations</td>
<td>768</td>
</tr>
<tr>
<td>Durbin-Wu-Hausman test</td>
<td>0.003</td>
</tr>
<tr>
<td>(p-value)</td>
<td>622</td>
</tr>
<tr>
<td></td>
<td>0.003</td>
</tr>
</tbody>
</table>

Notes: */**/*** denote significance at the 10%/5%/1% level, respectively. In the linear regressions, standard errors are given in parentheses and clustered at the pair level. Self-reported economic background is scaled from 1 (very poor) to 6 (very rich). Risk aversion is measured on a scale from 1 (very risk loving) to 10 (very risk averse). The corresponding first-stage regressions are reported in the first two columns of Table 2a.
Figures

Figure 1: Histogram of Scores (Pooled)
Figure 2: The Gender Pay Gap under Performance Pay Schemes across the Hourly Wage Distribution Separated by Gender Composition on the Job
Notes: The regression specification is in (B.1), and includes controls for personal and job characteristics, as well as occupation. Personal characteristics are education, race, age, job tenure, marital status, and dependent children. Job characteristics are industry, log establishment size, and gender composition on the job. The lower panel plots the 90% confidence interval of the difference-in-difference estimates in male-dominated environments.
Appendix A

Proofs

Proof of Proposition 1 (follows from Lazear and Rosen, 1981) The optimum investment in effort $\mu^*_i$ ($i = L, S$) will be a function of $\Delta W$, the prize spread, and the variance of the net dose of bad luck ($\varepsilon_{L,S} - \varepsilon_{S,S}$ and $\varepsilon_{S,L} - \varepsilon_{L,L}$ for $S$ and $L$, respectively) which equals $2\sigma^2_S$ for $S$ and $2\sigma^2_L$ for $L$. Under risk neutrality, $S$’s and $L$’s expected utilities are:

$$W_2 + \text{prob}_S(q_S > q_L) \Delta W - C(\mu_S) \tag{A.1}$$

and

$$W_2 + \text{prob}_L(q_L > q_S) \Delta W - C(\mu_L) \tag{A.2}$$

Then, one yields the following FOCs for $S$ and $L$, respectively:

$$\Delta W g(\mu_S - \mu_L) = C'(\mu_S) \tag{A.3}$$

and

$$\Delta W h(\mu_L - \mu_S) = C'(\mu_L) \tag{A.4}$$

where $g(\mu_S - \mu_L)$ is the pdf of a normal distribution with zero mean and a variance of $2\sigma^2_S$, and $h(\mu_L - \mu_S)$ is the pdf of a normal distribution with zero mean and a variance of $2\sigma^2_L$.

Given normality of the error terms in the output equations, one knows that $g(\cdot)$ and $h(\cdot)$ are symmetric. Also, each agent believes that both contestants are homogenous, so female and male beliefs reflect the standard Nash-Cournot case s.t.:

$$g(0) < h(0) \tag{A.5}$$

Turning to the principal, he has the same beliefs as the majority group to which agent $L$ belongs, leading the principal to assume the same optimization problem for all agents:

$$\Delta W h(0) = C'(\mu) \tag{A.6}$$

Now that the firm makes zero profit in expectation, it holds that:

$$2V\mu^* = W_1 + W_2 \tag{A.7}$$

where $V$ is the marginal social return and $\mu^*$ is the solution to (A.6).
Also, the firm has to make the right choice with regard to the prize spread even to achieve zero profit in expectation. Hence, the principal chooses \( \Delta W \) such that (given his beliefs):

\[
\Delta W^* = \arg \max \{ \frac{1}{2} W_1 + \frac{1}{2} W_2 - C(\mu^*) \} = \arg \max \{ V \mu^* - C(\mu^*) \}
\]  

(A.8)

Recall that \( \mu^* \) depends on \( \Delta W \), so the corresponding FOC is:

\[
(V - C'(\mu^*)) \frac{\partial \mu^*}{\partial \Delta W} = 0 \iff V = C'(\mu^*)
\]  

(A.9)

Combining (A.6), (A.7), and (A.9), one obtains:

\[
\Delta W^* = \frac{W_1 + W_2}{2 \mu^* h(0)} = \frac{W_1 + W_2}{2 \mu^* h(0)}
\]  

(A.10)

where \( \mu^*_L < \mu^*_m = \mu^* \).

Given that \( C(\cdot) \) is quadratic, we can finally conclude from (A.3), (A.4), and (A.10) in conjunction with (A.5) that:

\[
C'(\mu^*_S) = \frac{(W_1 + W_2) h(0)}{2 \mu^* h(0)} < \frac{(W_1 + W_2) h(0)}{2 \mu^* h(0)} = C'(\mu^*_L) = C'(\mu^*) = V
\]  

(A.11)

Thus, we have that \( \mu^*_S < \mu^*_L = \mu^* \). ■

Proof of Proposition 2 First, we show that the upper bound on \( W_2 \) is equivalent to an upper bound on \( \mu^* \) and thus on \( \mu^*_S - \mu^*_L \):

\[
W_2 < \sigma_L \left( 4 \sigma_L \ln \frac{\sigma_L}{\sigma_S} - \sqrt{\pi} V \right) \iff \sigma_L \sqrt{\pi} V + W_2 < -4\sigma_L^2 \ln \frac{\sigma_L}{\sigma_S}
\]

\[
\iff \sqrt{\sigma_L \sqrt{\pi} V + W_2} < \sqrt{-4\sigma_L^2 \ln \frac{\sigma_L}{\sigma_S}} \iff \sqrt{\frac{V}{2h(0)} + W_2} \iff \frac{\Delta W^*}{2} + W_2 < -4\sigma_L^2 \ln \frac{\sigma_L}{\sigma_S}
\]

(A.12)

as the principal chooses the prize spread optimally (cf. (A.10)). Now, given a quadratic cost function, one knows from (A.9) and (A.10) that \( \mu^* = \sqrt{\frac{W_1 + W_2}{2}} \). So, the above inequality translates to \( \mu^* < \sqrt{-4\sigma_L^2 \ln \frac{\sigma_L}{\sigma_S}} \), from which one can infer that \( \mu^*_S < \mu^*_L < -\sqrt{-4\sigma_L^2 \ln \frac{\sigma_L}{\sigma_S}} \) because \( \mu^*_L < \mu^* \) due to (A.11) and the symmetry of \( h(\cdot) \) around zero.

In the next step, we show that the latter condition is equivalent to \( g(0) < h(\mu^*_L - \mu^*_S) \) \( \iff \mu^*_S < \mu^*_L < \mu^* \) as dictated by the logic of (A.11). Note that the alternative assumption solely impacts the decision problem of the \( L \)-player:

\[
\mu^*_L - \mu^*_S < \sqrt{-4\sigma_L^2 \ln \frac{\sigma_L}{\sigma_S}} \iff \mu^*_L - \mu^*_S < -\sqrt{-4\sigma_L^2 \ln \frac{\sigma_L}{\sigma_S}} \iff \mu^*_L - \mu^*_S < \sqrt{-4\sigma_L^2 \ln \frac{\sigma_L}{\sigma_S}}
\]

(A.13)

From this, it follows that \( \mu^*_S < \mu^*_L < \mu^* \). ■
Appendix B

Decomposition of Data Used in Manning and Saidi (2010)

Here we describe the empirical strategy and the data used in Section 5 of this paper.

Individual performance pay schemes

In both the 1998 and the 2004 WERS data sets, managers at the firms were asked to indicate what kinds of pay schemes were offered to specific occupations. We are interested in performance pay schemes as opposed to fixed pay schemes. Performance pay schemes can typically be separated into piece rates and merit pay (based on subjective performance evaluation). Although our theory and experiments are explicitly based on tournament schemes (merit pay), performance pay schemes labeled as merit pay in labor market data do not necessarily cover all pay schemes that are effectively tournaments. Even if employees report to work under a piece rate scheme, their measured performance under that scheme eventually affects internal performance rankings relevant for promotion decisions. Thus, we generate a performance pay variable that indicates whether either compensation scheme, piece rates and/or merit pay, is in place. Specifically, we interpret pay schemes that are related to individual or team performance/output, assessments by supervisors, and acquisition of skills/core competences as performance pay schemes.

For all employees, we assign the respective pay schemes through their occupations at the respective firms.

Wage measure

Given the banded nature of the data, we use the midpoints of the categories, and divide the latter by the actual hours worked by each employee. This yields the hourly wage the logarithm of which is used as the dependent variable in the regressions.

Gender composition on the job

Both questionnaires include information on the type of work employees are doing, and whether it is done: only by men, mainly by men, equally by men and women, or only
by women. We summarize categories 1 and 2 as jobs done predominantly by men, and categories 4 and 5 as jobs done predominantly by women. Thus, one can generate three indicator variables, namely for female-, male-dominated, and mixed job environments. Note that in the 2004 survey employees were also able to indicate that they are the only ones doing their job, so the respective observations do not fall into any of the above categories, and are thus not used in the regressions separated by gender composition on the job.

**Empirical strategy**

We test whether women earn less under performance pay schemes by regressing log hourly wages on the existence of a performance pay scheme as defined above. In order to run such a regression, we assume that working under a competitive pay scheme is exogenous conditional on job and other controls (very much in the fashion of Lemieux et al. 2009). Indeed, Manning and Saidi (2010) showed that there is no significant pre-sorting of workers into pay schemes beyond sorting into occupations. We estimate the following regression specification:

\[
\ln wage_i = \alpha + \beta_1 female_i + \beta_2 performance\ pay_i \\
+ \beta_3 performance\ pay_i \times female_i + \beta_4 X_i + \varepsilon_i
\]  

(B.1)

where \(wage_i\) is the hourly wage and \(X_i\) is a vector of control variables for personal characteristics, job characteristics, and occupation. The results for the difference-in-difference estimate \(\beta_3\) are presented in Figure 2.