UNIVERSITY OF SOUTHAMPTON

FACULTY OF SOCIAL, HUMAN AND MATHEMATICAL SCIENCES

Department of Economics

Management Practices: Incentives, Team Formation, and Mentoring

by

Xiaocheng Hu

Thesis for the degree of Doctor of Philosophy

in the

Faculty of Social, Human and Mathematical Sciences

Department of Economics

November 2018

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SOCIAL, HUMAN AND MATHEMATICAL SCIENCES
DEPARTMENT OF ECONOMICS

Doctor of Philosophy

MANAGEMENT PRACTICES: INCENTIVES, TEAM FORMATION, AND MENTORING

by Xiaocheng Hu

The utilisation of incentives is essential for the success of organisations. Given the complexity of the real business world and the scarcity of high quality data, the academic literature on the impacts of incentives remains scant. This Thesis contributes to the literature by studying three different incentives including remuneration for multitasking group leaders, team formation rules, and determinants of manager-employee mentoring relationships. I map my economic analysis into three unique datasets to examine the incentive effects. I conducted a field experiment in two Chinese factories. I introduced rank incentives and monetary prizes to group leaders regarding their organisational behaviours. As a result, I find a significant increase in group performance. In another study, I designed a laboratory experiment with two stages of a real effort task to test the possible dynamic effects of optimal team composition. The results show that pairing the worst performing individuals with the best yields 20% lower first stage effort than random matching. Pairing the best with the best, however, yields 5% higher first stage effort than random matching. Last but not the least, I study both non-monetary and monetary determinants of mentoring relationships between managers and employees in British firms by using data from the Workplace Employment Relations Survey. In particular, I focus on the role of a manager's gender and the use of managerial incentive schemes. Past literature suggests a significant association between a manager's gender and mentoring behaviour. However, using longitudinal data this paper finds that the significant relationship disappears once firm fixed effects are included. The results also show a positive but weak association between managerial incentive schemes and managers' mentoring behaviour. Widespread mentorships are more likely to be found in firms where managers' payments are linked to organisational profits.

Contents

List of Figures

List of Tables

Research Thesis: Declaration of Authorship

I, Xiaocheng Hu , declare that the thesis entitled and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- none of this work has been published before submission.

Signed:...

Date:..

Acknowledgements

Summarising my transformation towards a PhD in a few lines is challenging. As the most well-known story ever written by Zhuangzi: [*O*]*nce, Zhuang Zhou dreamed he was a butterfly, a butterfly flitting and fluttering about, happy with himself and doing as he pleased. He didn't know that he was Zhuang Zhou. Suddenly he woke up and there he was, solid and unmistakable Zhuang Zhou. But he didn't know if he was Zhuang Zhou who had dreamt he was a butterfly, or a butterfly dreaming that he was Zhuang Zhou.* (Zhuangzi, chapter 2, translated by Watson (2003)). But, it is crystal clear that I can never reach such achievement alone. First and foremost, I thank my supervisors Thomas Gall and Michael Vlassopoulos, whose dedications transcended any border of professional obligation. I am honoured of the time and effort they patiently spent in guiding my unsteady thoughts into compelling propositions. This dissertation would not have been possible without their encouragements along all dimensions involved in my PhD. I also owe an enormous debt of gratitude to Brendon McConnell, Emmanouil Mentzakis, Giulio Seccia, Hector Calvo Pardo and Zacharias Maniadis for their advice and input along the way. Along this very line, I would like to thank the person who enlightened my path to pursue my PhD, Rocco Macchiavello. I also thank Spyros Galanis, Christos Ioannou, Corrado Giulietti, Carmine Ornaghi, Emanuela Lotti, Arnau Valladares-Esteban, Jan Podivinsky and the other scholars who promoted my personal and research development along these years. I appreciated the real sense of collegiality by integrity of persons I am proud to call friends: those in my cohort with whom I directly shared the passions and the traumas of the PhD, with particular regards to Abbas Gillani and Burak Demirtas. Those who preceded my experience - Michele Tuccio, Andrea Giovannetti, Panos Giannarakis, Dafni Papoutsaki, Jana Sadeh, Chuhong Wang and Michael Kerns and those who followed it - Abu Siddique, Larissa Marioni, Marius Strittmatter and Richard Kima among the many. Friends in other departments with whom I shared other fragments of PhD - Wentao Fu - had likewise importance. I also thank *ESRC* and *China Scholarship Council* for the faith society had put on my research.

Lastly, and most importantly, I would like to thank my wife Lin. Her support, encouragement, patience and unwavering love were undeniably the bedrock upon which the past few years of my life have been built. Her tolerance of my occasional impulsive moods is a testament in itself of her unyielding devotion and love. I would like to pay homage to my mother and my father, Yang and Hong, the pillars of my life. Also, I thank Lin's parents, Julian and Jianping, for their unending encouragement and support.

Nomenclature

$C(.)$	cost function
DiD	difference-in-difference estimation
$E\cdot$	expectation of a random variable
max	maximisation problem
EQ	employee questionnaire
FE	fixed effect estimation
HRM	human resource management
MQ	management questionnaire
M-W	Mann-Whitney-U
NAM	negative assortative matching
OLS	ordinary least squares
P(A)	abolish a payment scheme
P(I)	introduce a payment scheme
PAM	positive assortative matching
PERFM	performance related pay
PROFT	profit related pay
RAM	random matching
R&I	random matching and early stage incentives
RMB	Chinese Yuan
SME	small and medium-sized enterprise
SSEL	Social Sciences Experimental Lab
W	experimental week
WERS	Workplace Employment Relations Surveys

To the loves of my life:

My mom, Yang, and
My wife, Lin

Chapter 1

Introduction

The disparity in organisation performance has become a focus of empirical and theoretical interest throughout the social sciences, including economics. These enormous inequality in between-firm productivity are highly persistent, contributing to the radical divergence in economic performance over time and across countries.

Since the 1970s, personnel economics started to forge distinctively from the traditional labour economics as a unique field. Unlike labour economists, personnel economists no longer treat the organisational behaviour as a "black box". Instead, it focuses on the human resource management (HRM) problems in firms and how the solutions to these HRM problems can shape the performance of the firm.

As suggested by the seminal work of Ichniowski et al. (1997) the application of advanced HRM practices is a crucial determinant of organisation performance. About half of the difference in average productivity between organisations in the U.S. and Southern EU countries, which studied in Bloom et al. (2016), can be explained by an index of advanced practices. A micro level example can also be found in Bloom et al. (2013) who show a significant causal role for management practices in Indian textile factories.

While some management practices can directly impact organisation productivity, many others - like HRM practices: mentoring, teamwork, and incentive schemes - are mediated through employee decision-making and effort investment. If HRM practices are complementary with higher-ability employees, as seems plausible, then one would expect firms that use these practices to systematically adapt both the skill composition of their workforce and the structure of their payment system. Therefore, to unravel what drives the disproportion in organisation performance it is vital to understand the impact of the HRM practices.

HRM conducts a wide range of activities. Teamwork and manager-employee relationships have, in particular, caught my attention. Collaborative work with others is prominent today. It is almost present in every context, such as industries, sports, schools, academics, and health services. However, we are yet to unlock the full potential of teamwork. The two main problems of teamwork are the free-rider problem and interpersonal conflicts. The free-rider problem has been vastly developed in the "Public Goods" literature, while the interpersonal conflicts are mostly seen in organisations that facilitate teamwork in a hierarchical context. Managers at the top sometimes appear to be ingrained individualistic and have troubles to engage in collaborative work, such as mentoring employees, organising resources, and allocating tasks, which contributes to poor teamwork and inefficiency. Therefore, I study three aspects of organisational economics that are related to these issues, including: managerial incentives, team formation rules, and mentoring relationships between managers and employees.

Incentives are employed by the firm to encourage employees to perform tasks, recruit talents, and facilitate wage flexibility. The use of incentive schemes is vital for the success of organisations. In the UK the British Workplace Employment Relations Surveys (WERS) report that 41% of UK firms had used use at least one incentive pay scheme in 1984, and the figure rose to 55% in 2011. Similar figures can also be found in the US.

It is well established that incentives do have an impact on behaviour in organisations. Lazear (2000) studies the use of pay-for-performance incentive scheme for workers at Safelite Glass. He found that a change from hourly rate to piece rate pay led to a 44 % increase in workers' productivity. Even though the firm opted for a suboptimal compensation scheme, both workers' wage and the firm's profit had increased after the implementation of the pay-for-performance scheme. Similarly, Shearer (2004) found that performance-related pay induced a 20% increase in workers' productivity by using data from a field experiment which shifted the incentive scheme from a fixed pay to a performance-related pay in British Columbia.

In the above studies, performance-related pay appears to increase employee's performance in the measured dimension. However, as the multi-tasking literature emphasises that there may be important-but-unquantifiable aspects of performance such as quality of job and creativity that are ignored as employees focus on the measured-performance dimensions (Oyer and Schaefer, 2010). The empirical evidence using experimental data are mixed. Kishore et al. (2013), Al-Ubaydli et al. (2015), and Hong et al. (2018) find evidence to support the standard theoretical prediction (Holmstrom and Milgrom, 1991), while Shearer (2004), Bandiera et al. (2005), Hossain and List (2012), and Englmaier et al. (2017) do not find evidence that incentives focusing on the measurable dimension (e.g. productivity) affected the performance in the unmeasurable dimension (e.g. quality of production).

There have been very few empirical studies on how multitasking issues affect the structure of incentive schemes, one notable exception is Slade (1996). Slade studied the multitask-agency problem using incentive contracts between private, integrated oil firms and their service stations and tested how variations in the feature of one task affect the structure of the incentive scheme for another. The results showed that higher-powered incentives were offered for one task while the other task is not a strong complement.

When a group-based pay is given for producing a public good such as team cooperation and cohesion free-riding becomes the new problem (Griffith and Neely, 2009; Bartel et al., 2017).

Hence, an important and complex problem arises when output has different dimensions and varies in externality and quantifiability which is the focus of Chapter 2. To the best of my knowledge, Chapter 2 is the first to provide field evidence on the productivity effect of incentivising group leaders regarding their (hard-to-measure) organisational inputs under a subjective performance evaluation system and its impact on the group members' performance.

Mentoring is another mechanism that managers can use to improve employees' performance. Mentoring programs are widespread in academia, industry, and sports. In the business world, career mentoring takes place when an experienced manager (mentor) helps an inexperienced worker to achieve his/her career goals. Organisations have recognised the substantial impact of mentoring on business performance. Mentoring fosters positive behaviours among workers and improves social connections and behavioural outcomes within the organisation (Allen et al., 2004; Ragins and Cotton, 1999; Ragins et al., 2000). Given the rapidly-evolving demographic composition of the workforce and subsequent organisational needs, there is a renewed interest in utilising mentoring in the workplace. A study reports that 93 per cent of the 11,000 small and medium-sized enterprises (SMEs) surveyed in 17 countries across the world acknowledge that mentoring helps employees to succeed. However, only 28 per cent of them make use of mentors.

To understand this perplexing gap, it is important to revisit the factors that can predict provision or receipt of mentoring in the workplace. Traditionally, organisational psychologists identified the following factors as the determinants of mentoring relationships: the personal characteristics of mentors and mentees, perceived similarities between mentors and mentees, and the type of mentoring used in organisations. In particular, it includes mentors' and mentees' big five personality traits, gender, age, race, education, tenure, and the formality and similarity in mentoring dyads (e.g. Allen and Eby, 2004; Armstrong et al., 2002; Eby et al., 2013; Godshalk and Sosik, 2003; Kammeyer-Mueller and Judge, 2008; Liang and Gong, 2013; O'Brien et al., 2010).

The literature on gender and leadership also shows the potential relationship between mentor gender and mentoring supports. Mentors and leaders are indifferent in many forms (Scandura and Schriesheim, 1994). They are typically more experienced seniors responsible for juniors (McManus and Russell, 1997). Theories of the relationship between gender and leadership (See Powell (2010) for a review) suggest that female leaders are higher than male leaders in individualised consideration, defined as a focus on developing and mentoring subordinates and attending to their personal needs (Eagly et al., 2003). Moreover, Sosik and Godshalk (2000) finds that individualised consideration have been positively linked to the amount of support mentors provide. Taking together, female managers who are higher in individualised consideration in general should have a stronger focus on meeting employees' career development needs through mentoring supports.

On the other side, organisational culture and norms also play a vital role in promoting mentoring relationships. HRM practices can provide powerful incentives for mentoring by, for instance, rewarding managers for investing their time to support their employees' careers. Managers might be more motivated to provide mentoring support and employees might be more willing to seek mentoring support if they perceive that the organisation values such behaviours. In contrast to the extensive literature focusing on easily observable individual traits, little research has determined the underlying role the firm plays.

Previous research has shown that the share of female managers at higher levels of the firm has high potential to affect female workers' career outcomes such as promotion (e.g. Bell et al., 2008; Matsa and Miller, 2011). A similar relationship has been established in other settings, such as higher education (e.g. Neumark and Gardecki, 1998; Bettinger and Long, 2005). In Chapter 4, I test the relationship between the share of female managers and the percentage of employees perceives managers' mentoring supports in the workplace.

Moving beyond easily observable manager characteristics, Bandiera et al. (2009) study a change in managerial incentives in a British fruit plant. When managers are paid using fixed wages, they tend to allocate resources to those who are socially connected to them as measured by shared nationality, living quarters, or whether the manager and worker arrived at the plant at the same time. However, when managerial pay is changed to bonuses based on the overall group output, managers change their managerial behaviour. They start to allocate resources to the most productive workers regardless of the social connections. As another example of team-based compensation, Chan et al. (2014) also shows that group-based payment schemes increase employees' incentives to help others. In Chapter 4, I investigate the impact of team-based incentives on the managers' mentoring behaviours.

Team formation rules are also important when one contemplates the supportive and caring relationships between employees. There is a large literature that has focused on the effects of employees on their colleagues and teammates (Kandel and Lazear, 1992; Falk and Ichino, 2006; Mohnen et al., 2008; Mas and Moretti, 2009; Babcock et al., 2015), the way that social connections among teammates affect performance (Bandiera et al., 2005), on whether having team members with different gender, age or ethnic background can be disruptive (Charness and Villeval, 2009; Apesteguia et al., 2012; Hjort, 2014), on the impact of financial incentives on task assignment within teams (Burgess et al., 2010), and on how the interaction of team composition and incentive schemes may affect performance (Hamilton et al., 2003; Bandiera et al., 2013; Delfgaauw et al., 2013; Chan et al., 2014). While much thought has gone into the analysis of how to optimally organise teams and how team composition may affect individual and aggregate performance in teams, there has been less emphasis, however, on possible effects on individual behaviour before teams are formed.

For instance, some firms assign workers into teams that are heterogeneous in ability, by partnering strong managers with weaker employees, in order to facilitate learning or to provide role models that lead to productivity gains (Hamilton et al., 2003). This practice may limit an individual's desire to exert effort at an earlier stage, i.e., there is an equity-efficiency trade-off. Conversely, if the best managers are grouped with better employees for optimal team outcomes it will provide additional incentives for effort at an earlier stage. Depending on the degree of production complementarity (Franco et al., 2011) and the strength of incentives (Bandiera et al., 2013), this pattern can be the outcome when employees are allowed to choose their own teammates since employees will tend to match positively assortatively in ability. Of course, team formation may also be left to chance, for instance, if the assignment is by the sequence of arrival, follows a rotation system or is guided by alphabetic order of names (e.g. Bartel et al., 2014).

The outcome of teams may be important, but in except in a few industries, like salesman, where the best salesman is usually reassigned to another territory on an annual basis to help the weak ones for customer acquisition and these salesmen's bonuses are linked to the overall acquisition of their territory, the choice before the reassignment is likely to be as or more important than that after the assignment. Good salesmen may subtly pretend to be average performers to avoid being assigned to weaker teams and underdeveloped territories in which they need to invest more efforts to achieve the same outcome as they would now. At a minimum, this remains a void that needs to be filled. Chapter 3 is the first paper to do so, and it provides us with a window into individuals' reactions to different matching policies even before the team is assigned, rather than focusing solely on the in-team effort for a single matching policy.

The model which guides this study is related to the more general theoretical framework in Gall et al. (2006, 2012), which points out that access to better peers may affect early stage investment incentives suggesting that matching policy which pairs the best individuals into teams (i.e. positive assortative matching) will increase subjects' effort and those are not may decrease subjects' effort. The results of Chapter 3 are consistent with this theory. Several other studies consider investments before matching under asymmetric information (e.g. Bidner, 2010; Hopkins, 2012; Hoppe et al., 2009), mainly focusing on wasteful signalling, but not considering rematch policies.

Finally, Chapter 3 is also related to a small but growing literature recognising that admissions policies shape incentives for pre-college human capital accumulation. Francis and Tannuri-Pianto (2012) estimated an overall negative effect of a policy change which applied only to top universities in Rio De Janeiro. Bodoh-Creed and Hickman (2015a) estimated a structural empirical model of the U.S. college admissions market based on the theoretical foundation of Bodoh-Creed and Hickman (2015b). This counterfactual analysis of admissions, investment, and welfare under alternative affirmative action policies found evidence that minority students invest more in human capital in response to improved opportunities afforded them by affirmative action policy, and in turn, their graduation rates and economic outcomes improve. Chapter 3 frames this dynamic in a more general setup.

In this Thesis, I study incentives, mentoring, and team formation rules. An outline of this work can be summarised as follows. In Chapter 2 I estimate the impact of incentivising multitasking group leaders regarding their hard to measure inputs such as organising teams on the group performance by using the data which I collected in a field experiment that is conducted in two

Chinese manufacturing factories. I introduced a rank incentive with monetary prizes to the production line foreman (group leader of the production line) in one factory while keeping the other factory as constant. This incentive scheme is linked to the foremen's organisational activities which are subjectively evaluated by senior line managers. In Chapter 3 I conducted a laboratory experiment with two stages of a real effort task to test the possible dynamic effects of optimal team composition: anticipating that team formation is based on individuals' prior performance will affect individual prior performance. Student subjects first work individually without monetary incentives and are then assigned to teams of two where compensation is based on team performance. Lastly, in Chapter 4 I test the possible determinants of mentoring relationships between managers and employees in British workplaces by using the WERS. In particular, I focus on two firm characteristics: the share of female managers and the type of managerial incentive schemes used in the workplace.

Chapter 2

Subjective Performance Evaluation in a Multi-tasking Environment: a Firm-level Experiment in China

Abstract. We examine a multitasking problem where one task is to produce private goods while the other is to create public goods which is hard to measure. Such problems can be found in organisations that make use of multitasking leaders. Group leaders take responsibility for organising teams (public goods) and contribute as a member (private goods). Presenting evidence from a natural field experiment, we shed light on the impact of a high-powered remuneration system regarding leaders' organisational behaviours. In particular, we designed a monitoring system which subjectively evaluates leaders' organisational inputs, and we offered each leader a new bonus scheme that is depending on her relative performance in organising teams among other group leaders within the factory. Using individual daily production records, we find an overall 6% increase in workers' productivity, excluding the leaders. In line with our theoretical model, strengthening incentives on organising teams does not necessarily have a negative impact on leaders' production performance. We show that leaders' production performances increase as they invest more time on the job.[1]

[1] I warmly thank Thomas Gall, Michael Vlassopoulos and Brendon McConnell for their valuable comments and suggestions. This paper also greatly benefited from comments and suggestions received from seminar participants at Southampton and the East Anglia Economics PhD Workshop. Financial support is gratefully acknowledged from $ESRC$.

2.1 Introduction

In the modern business world, performing more than one task is virtually unavoidable: at the office, during the regular meeting, or in the boardroom. The general problem is how to incentivise multitasking employees on multiple dimensions. When incentives fail to encourage employees to meet the employer's goals and objectives, the outcomes are inefficient. For instance, a fixed payment system may be extremely inefficient as it overcompensates poor performers and under-compensates top performers while a performance-based compensation scheme can encourage workers to work harder on the measured dimension to earn a higher wage but the quality of job and creativity which are also important but hard to measure are usually disregarded. When a group-based pay is used for producing or improving a public good - such as team cooperation and cohesion - free-riding comes down the pipeline.[2] Hence, an important and complex problem arises when outputs have different dimensions and vary in externality and quantifiability.

In organisations, this problem is particularly daunting at the first line management level such as team leaders. A team leader usually works within a group, as a member, carrying out the same roles as others but with additional 'leader' responsibilities - such as providing supervision, motivation, and maintaining the quality and quantity of production and/or service - as opposed to higher level management who often have a separate job role altogether. A 2016 Deloitte study reported that only 26% of 7,000 large companies from 130 countries are functionally organised, 90% of these companies cite leadership as a significant issue, and 70% of them rate it as urgent.

As multitasking models predict, leaders will not invest effort in tasks that are not covered by the pay-for-performance scheme. For example, leaders in a US law firm shifted their billable hours to non-billable hours after the firm reduced the individual-based billable performance pay and introduced a bonus scheme which is determined by objective and subjective evaluation of both billable and non-billable activities (Bartel et al., 2017). However, this study provided limited insights for how efficiency can be achieved within a firm by employing a compensation system that monitors multiple dimensions as the complexity of their remuneration design.

Understanding how multidimensional incentive schemes affect leaders' effort choice in individual performance (private good) and subordinates' performance (public good) is essential for the contract design of team leaders. In particular, our study focuses on two dimensions: perfectly quantifiable performance (production output) and hard-to-measure inputs (organising teams). When performance is hard to measure, firms tend to use subjective performance

[2] Griffith and Neely (2009) find that moving from a group-based incentive - profit-related pay - to a Balanced Scorecard system which includes a variety of measures increased the organisational performance. They argue that the information content on the Balanced Scorecard is the key for the improvement, not the incentive per se.

evaluations to solve the multitasking problem.[3] Information about the effort leaders invest in organising groups can be gathered from the senior management.[4] However, while subjective evaluations may enable the firm to reward leaders' organisational inputs and bring efficiency to the team, they may have some pitfalls. First, subjectively evaluating and rewarding leaders' organisational inputs may be costly; the firm needs to invest enormous resources, such as time, training, and money. Second, the employee being evaluated will have an incentive to invest effort in developing a good rapport with those evaluating them, rather than spending time on the job. Third, they can have a counterproductive impact on employee morale and satisfaction, as employees may question the fairness and accuracy of the evaluation process. We developed a novel system to evaluate leaders' organisational behaviours while circumventing these issues.

This study presents evidence from a natural field experiment designed to measure the effect of an incentive system which subjectively evaluates forewomen (group leaders) regarding their hard to measure effort - organising teams. By taking advantage of an opportunity that a Chinese manufacturing company was avid for increasing production efficiency, we conducted a field experiment in a natural setting with factory workers. This collaboration enables us to explore how forewomen respond to multidimensional incentives in a multitasking environment. We choose work where the forewomen perform the same manufacturing task as the workers (group members), and the quantity of output is carefully recorded.

We constructed an incentive scheme and trained the production management to implement it, which helped the company to minimise the cost of introducing a new evaluation system. The new scheme is composed of two elements: rank incentives and monetary prizes, both are linked to the subjective evaluation results of forewomen's organisational activities. In doing this, we can assess how multitasking forewomen respond to an incentive, which based on partially observable inputs - organising teams. We decompose the total effect on team productivity into that caused by changes in workers' productivity (as the group becomes more organised) and that caused by changes in forewomen's production effort (due to multitasking).

The firm we study has two independent factories located on the northern side of Jiangxi, a south-eastern province of China with a high concentration of manufacturers of medical products. The field experiment was designed and implemented in collaboration with the production management in both factories, and they allowed us to introduce the treatment in one of the factories and left the other one operationally unchanged throughout (the control group). Ours is not a randomised controlled trial, namely we do not randomly assign teams to treatment and control groups. The choice of between-factory experimental design is determined by the fact that teams from the same factory can easily observe each other and react to the incentives

[3] A 2015 Global Partner Compensation System Survey of the processes law firm partnerships use to determine partner compensation found that 52% of US and Canadian law firms use a purely subjective system while 38% use a modified subjective or combination system. See https://www.edge.ai/2018/05/edge-global-partner-compensation-system-survey/ for details.

[4] Evaluation can also be given by those who are evaluated if the interaction is not repeated and collusion can be prevented.

offered to their colleagues so that the comparison of contemporaneously assigned treatment and control groups would yield biased treatment effect estimates.

Importantly, before our intervention, the salary structure of forewomen in both factories is a flat monthly payment for organising teams, an individual piece rate for manufacturing, and other fixed bonus schemes. Our experiment offers every forewoman in the treated factory an opportunity to compete for an extra bonus per month. The senior management ranks the forewomen regarding their organisational behaviours. The higher the ranking, the higher the reward. The monetary prize for the lowest ranked forewoman is the same size as the payment she receives from the company for organising the team, while the highest ranked forewoman receives a prize which is more than twofold of the flat rate paid by the company.[5]

During the treatment period, the ranking is public information and updated on a weekly basis in the treated factory, whereas in the controlled factory both forewomen and workers were not aware of the evaluation process and the results. Publicly educating the treated forewomen regarding the evaluation system and their weekly ranking results alleviate the concern of fairness and accuracy.[6] Moreover, in each factory, more than one manager was assigned to evaluate the forewomen. This can also prevent the manager's misreport due to personal perceptions. Hence forewomen are less likely to question the fairness and accuracy. Employing multiple assessors can increase forewomen's costs for collusion as well. Last but not least, this also allows the more productive forewomen to motivate other forewomen through contagious enthusiasm or through embarrassment over the unfavourable direct performance comparison. Peer pressure may force forewomen to internalise their spillovers. If pressure is sufficiently strong, it could push forewomen toward higher organisational performance, as illustrated in Kandel and Lazear (1992).

By comparing the output quantity of workers and forewomen between these two factories, we can estimate the treatment effects of our interventions. To the best of our knowledge, this paper is the first field experiment designed to evaluate the effect of subjective evaluations along with monetary incentives in a multitasking environment. However, our experimental design does not allow us to separately identify the effect of rank incentives from the effect of monetary prizes. In the literature, empirical studies show that the provision of relative performance feedback can have mixed effects on individual performance. Blanes i Vidal and Nossol (2011), Bradler et al. (2016), Delfgaauw et al. (2013), and Kosfeld and Neckermann (2011) find that it can induce higher performance, while Bandiera et al. (2013), Barankay (2012), and Eriksson et al. (2009) obtain an opposite result.[7]

[5]The field evidence on tournament incentives tests whether individual behaviour changes with various schemes can be found in Erev et al. (1993), Bandiera et al. (2013), and Delfgaauw et al. (2013, 2015).

[6]Rankings can also effectively provide information on relative earnings and which is a key determinant of happiness (Layard, 2011; Kahneman et al., 2006).

[7]Theorists highlight that relative performance feedback has an impact on individual performance if individuals have concerns for their relative status (Lizzeri et al., 2002; Ertac, 2005; Moldovanu et al., 2007; Besley and Ghatak, 2008; Ederer, 2010), and supported by research in psychology (Kluger and Denisi, 1996), and neuroscience (Fliessbach et al., 2007).

We develop a simple theoretical framework in which a forewoman's manufacturing effort and organising effort are complements because of the team efficiency spillover. The rationale underlying this assumption is that a forewoman brings efficiency to the workers by organising the team production, which can make the forewoman more productive as efficient workers are less likely to create problems on the production line relative to the inefficient ones. The model makes precise how the intervention affects the forewoman's and the worker's performance.

It is expected that forewomen can shape team productivity by facilitating mutual learning or by influencing the group production norm. Mutual learning suggests that forewomen may be able to encourage more able workers (e.g., those who are more productive under individual piece rates) to teach the less able workers to be more productive, thereby enhancing team productivity. On the other hand, peer pressure may be used to achieve a productive group norm, as modelled in Kandel and Lazear (1992). A forewoman maximises but takes into account the effect of her actions on the views of her line workers, which enter the forewoman's utility function. Norms can be established as the equilibrium outcome of a process where deviations from any given (say, mean) level of effort result in direct or indirect sanctions. When a forewoman departs from the team norm and other workers impose disutility on her for the extent of her departures, peer pressure arises, and the equilibrium effort is higher than it would be without peer pressure.

Using more than 5,000 individual-daily observations across 59 workers and 13 forewomen, we first show that a sufficiently large increase in incentive power motivates forewomen to exert more efforts to organise the team, and this increases both the average production output and productivity of workers by 8% and 6% respectively. Second, we show that the average production output of forewomen unambiguously increases while the effect on their average productivity is ambiguous. On the one hand, this can be explained by the fact that the positive spillover effect may vary across production lines. As precisely predicted in our model, a forewoman's output would fall when this spillover effect is particularly small. On the other hand, average forewomen's productivity can decrease if they frequently organise the team production. This is because switching between different tasks can slow them down, and multitasking can sometimes cause unnecessary stresses.[8] Another reason that we are unable to measure the precise change in forewomen's productivity is we do not observe the exact time a forewoman invested in the manufacturing task.

More subtly, the introduction of subjective evaluations and monetary prizes lead to significant changes in forewomen's working time. Relative to the forewomen worked in the controlled factory the treated forewomen spent roughly an extra 30 minutes a day on the job during the treatment period. In fact, this result is consistent with our expectations as the monthly prize for the highest (lowest) ranked forewomen is equivalent to the product between the highest (lowest) piece-rate wage per hour and the average number of days worked in a month. It was the senior management's aspiration to have the forewomen's engagements in organisation and management for about an hour in a day.

[8] Psychologists find that multitasking can lead to distraction and stress (e.g. Drews et al., 2008; Mark et al., 2012).

Furthermore, we find that the ranking of forewomen's organisational performance does positively correlate with workers' productivity but not perfect, suggesting that workers do benefit from a better-organised forewoman and the senior management was not simply taking the workers' production outcome into account when evaluating the forewomen. For the forewomen, the ranking of their organisational performance is positively associated with their productivity, suggesting that a better-ranked forewoman is also more productive as the production line is better-organised and efficient which is consistent with our assumption on team efficiency spillover.

Finally, we take the difference between forewoman's weekly organisational ranks to help us understand the underlying mechanism of the treatment effects. We find that forewomen had to increase the rankings by marginally sacrificing their productivity. This is consistent with juggling multiple tasks slows the forewomen down. Nevertheless, due to a small sample size this result is not statistically significant at the conventional levels.

The results from the scant literature that analyses the effect of incentives in a multitasking environment using experimental data are mixed. Shearer (2004), Bandiera et al. (2005), Hossain and List (2012), and Englmaier et al. (2017) do not find that incentives focusing on one dimension (e.g. productivity) affected the performance in the other dimension (quality). On the other hand, Kishore et al. (2013) find that multitasking concerns are modest when workers reached their targets and they are paid bonus-based incentive schemes. Similarly, Al-Ubaydli et al. (2015) and Hong et al. (2018) find that workers under a piece-rate wage produce high-quality work while workers under a flat wage rate do not. In this paper, we do not investigate the quantity-quality trade-off but a group leader's choice between individual performance and group performance. There have been very few empirical studies on how multitasking issues affect the structure of incentive schemes, one notable exception is Slade (1996). Similarly, Manthei et al. (2018) shows that workers' efforts are distorted towards the more profitable tasks when managers have no access to objective measures but assess worker's performance subjectively. Once the managers have access to objective performance measures, both worker's and firm's performance increase significantly. To the best of our knowledge, this paper is the first to provide field evidence on the productivity effect of incentivising group leaders regarding their (hard-to-measure) organisational inputs under a subjective performance evaluation system.

Using field experimental data, Bandiera et al. (2013), Casas-Arce and Martínez-Jerez (2009), and Delfgaauw et al. (2013) have studied tournaments among fruit pickers and retailers. They find a positive effect of tournament incentives on performance, but none of them varies the prize spread. Lim et al. (2009) varies both the number and the distribution of prizes in contests among fundraisers. They find that performance is higher in tournaments with multiple prizes in comparison with single-prize tournaments, but there is no further effect on performance by differentiating prizes by rank.

There is a large literature, both theoretical (see Kandel and Lazear, 1992) and empirical (see Falk and Ichino, 2006; Mas and Moretti, 2009) have studied the effects of workers on their peers and team members (see Ichniowski and Shaw, 2003, for teams and complementarities). Peer effects may be important, but the relationship with one's superior is likely to be as important as or more important than that to any other worker. Using data from a service firm, Lazear et al. (2015) find that a higher quality manager increases the output of the supervised team over that supervised by a lower quality manager by about as much as adding one member to the team. Different from the forewomen studied in our experiment these managers are not multitasking.

A similar relationship can be found in other settings, such as education. Kremer et al. (2010) and Muralidharan and Sundararaman (2013) conduct experiments to show that performance pay to teachers increases student performance in the dimensions along which teachers are incentivised, and there are no adverse effects in the unrewarded dimensions. If one assumes that students do not know their production functions, adverse effects may be found for poor-performing students (see Fryer et al., 2012).

The remainder of the paper proceeds as follows. The next section outlines the theoretical framework. Section 2.3 describes the setting and our experimental design. Sections 2.4 and 2.5 discuss our main results and the underlying causes of these results, respectively. Finally, section 2.6 concludes.

2.2 Theoretical Framework

We develop a simple model to demonstrate the impact of changing incentive schemes for multitasking foremen regarding their hard-to-measure inputs - organising team production. In the context of our experiment, the firm hires two types of employees: a worker and a foreman. The worker only performs production task while the foreman is responsible for both production and organisation tasks. The firm observes employees' production output and offers them a piece-rate payment scheme. While the piece-rate remains constant, the firm replaces a fixed bonus scheme which compensates foreman's organisational activities with a new bonus depending on the foreman's relative position in the leadership ranking within the firm.

We first derive the results of employees' optimal effort choices when the firm offers the foreman a fixed bonus for organising the team. We then derive the results in the context of the new bonus scheme where the foreman receives a bonus depending on her relative rank in the subjective organisational input distribution as perceived by the firm management. By comparing these results, we are able to illustrate the effect of introducing the new bonus scheme on the worker's effort provision and the foreman's. We interpret effort choices in our model as intensity. It is important to notice that we are not aiming to derive an optimal incentive scheme from the firm's or the social planner's perspective.

A. Basic Model

Production Function and Team Efficiency Spillover.— First and foremost, the production function of worker w who produces individual output $y_w(e_w)$ can be written as follows:

$$y_w(e_w) = e_w(1 + \lambda g_f),$$

where $y_w(e_w)$ is depending on individual production effort e_w and foreman f's organisational input g_f. In this production function, the worker's production effort and the foreman's organisational input are complements, meaning that the worker benefits from the outcome of the foreman's organisational activities only if she expends production efforts. That is to say, the return of production effort is increasing in foreman's organisational input, and the greater this increase, the more important the team efficiency spillover captured by the parameter $\lambda > 0$. The worker benefits from the positive team efficiency when she exerts more effort since we define effort as intensity, meaning that working harder yields a better outcome when the team is efficient.[9]

On the other side, the multitasking foreman f who produces individual output $y_f(e_f, g_f)$ and organises the team according to the following production function:

$$y_f(e_f, g_f) = e_f(1 + \lambda g_f),$$

where this is to say the foreman's production output $y_f(e_f, g_f)$ is depending on individual production effort e_f and organisational input g_f. The complementarity assumption between the foreman's organisational input and production effort implies that the foreman benefits from the outcome of organising the team production only if she exerts more production efforts. The idea is that when the foreman invests more organisational inputs the worker becomes more efficient. The foreman, therefore, earns herself time to concentrate on her own production task, which provides her with a higher wage, as she encounters less disturbance (e.g. informal conversations) from the worker than if the worker is not working very hard. For simplicity, we assume that the team efficiency spillover here is identical to the parameter λ in the worker's production function, and it is complementary to the foreman's production effort. In the general case, the effect of team efficiency may be different across employees. Removing this assumption does not affect our results.

[9] As in existing studies, this formulation abstracts from the dynamic implications of contemporaneous spillover through support and cooperation between the worker and the foreman on the job. The underlying rationale is that workers with well-organised foremen are more productive because they have much better access to resources than if they were in an unorganised group. In addition, this equation assumes that worker's production effort is required to "unlock" the potential of team efficiency. This assumption of complementarity between team efficiency spillover and production effort provision is one of the drivers of why team efficiency translates into worker's wage in our model: workers in efficient teams exert higher effort, for which they are compensated with higher wages.

Cost of Efforts.— Because exerting effort is costly to the foreman regardless the work type, the cost of effort function is quadratic in both production and organisational inputs: $C(e_f, g_f) = \frac{(e_f+g_f)^2}{2}$. In this cost function, the positive cross derivative with respect to e_f and g_f implies that increasing effort in one dimension increases the marginal cost in the other. When the foreman increases her effort in organising the team, it leads to some negative externality on her production effort. For the worker, the cost of production effort is $C(e_w) = \frac{(e_w)^2}{2}$.

B. Pre-Intervention: Firm Offers the Foreman a Fixed Bonus for Organising Teams

The Forman's and The Worker's Maximisation Problems Before Interventions.— During the pre-intervention period, production output is compensated by piece rate w and the firm offers the foreman a constant bonus b for organising the team to optimise its profit. We assume that both the firm and the employees are risk neutral. This assumption simplifies our analysis without being a necessary condition for our general argument. Because of risk neutrality, the foreman maximises her expected wage minus the combined cost of effort:

$$\begin{aligned}\max_{\{e_f,g_f\}} U_f(e_f,g_f) &= W_f(e_f,g_f) - C(e_f,g_f) = wy_f(e_f,g_f) + b - C(e_f,g_f)\\ &= w[e_f(1+\lambda g_f)] + b - \frac{(e_f+g_f)^2}{2},\end{aligned}$$

This leads to the first order condition with respect to e_f:

$$e_f = w + (\lambda w - 1)g_f,$$

and the first order condition with respect to g_f:

$$g_f = (\lambda w - 1)e_f.$$

For the worker w, she maximises her expected wage minus the cost of effort:

$$\begin{aligned}\max_{\{e_w\}} U_w(e_w) &= W_w(e_w) - C(e_w) = wy_w(e_w) - C(e_w)\\ &= w[e_w(1+\lambda g_f)] - \frac{(e_w)^2}{2}.\end{aligned}$$

the first order condition with respect to e_w is given as follows:

$$e_w = w(1 + \lambda g_f).$$

Solving these first order conditions we get the optimal effort levels of the foreman:

$$e_f^* = \frac{w}{1-(\lambda w-1)^2}, \tag{2.1}$$

$$g_f^* = \frac{w(\lambda w-1)}{1-(\lambda w-1)^2}. \tag{2.2}$$

For the worker, the optimal production effort level is given by:

$$e_w^* = \frac{\lambda w^2}{1-(\lambda w-1)^2}. \tag{2.3}$$

Wage Contracts and The Firm's Maximisation Problems Before Interventions.— Now, we move on to solve the firm's profit maximisation problem. Taking into account both the worker's and the foreman's optimal effort levels, the firm chooses the piece rate w to maximising its expected profit, $E\pi = (p-w)(y_f(e_f, g_f) + y_w(e_w)) - b$. In line with the context of our experiment that the firm's marketing team usually sets the price of products at the beginning of each year, we assume that the market price for per unit of output is given exogenously at $p > 0$.

As detailed in Appendix A.1.1, we should assume $0 < \lambda w < 2$ and there are two distinct solutions for the firm's profit maximisation problem. When $\lambda w \leq 1$, $\overline{g}_f = 0, \overline{e}_f = \overline{e}_w = w$. If $\lambda w > 1$, g_f^*, e_f^*, and e_w^* are expressed as in equations 2.1, 2.2, and 2.3. We can then derive the firm's first order condition and an expression of the piece rate for each case. By comparing the firm's expected profits across these two cases, we predict that for a given λ there exists a p^* such that the profit maximising firm will choose w^* which yields g_f^*, e_f^*, and e_w^* and a higher profit than choosing $\overline{w}$ for $p > p^*$. On the other hand, the firm prefers $\overline{w}$ which yields $\overline{g}_f = 0, \overline{e}_f = \overline{e}_w = \overline{w}$ for all positive p that $p < p^*$. These results are reported in Lemma 2.1 (see Appendix A.1.2 for details).

Lemma 2.1. *For a given λ, the firm facing a market price where $p > p^*$ will set $w = w^*$ when it maximises its expected profit, and the foreman responds to it by choosing $g_f = g_f^*$. However, if $p < p^*$ the firm favours another piece rate scheme $\overline{w}$ in which the foreman exerts 0 effort in organising team production.*

Lemma 2.1 implies that when the market price is dramatically low selling products is not profitable for the firm. Thus, the firm would not value the foreman's organisational behaviour and sets the piece rate at $\overline{w}$. The foreman, therefore, has no intention to organise the team production. On the other hand, when the market price is sufficiently high producing products is beneficial to the firm. Hence, the firm attempts to expand its production by offering a piece rate w^* which also encourages the foreman to organise the team due to the complementarity between her production effort and organisational input imposed by the production technology.

C. Post-Intervention: Firm Offers the Foreman a Performance-Related Bonus for Organising Teams

Wage Contracts and The Maximisation Problems After Interventions.— When the management starts to evaluate the foremen's organisational activities subjectively, the foreman receives a bonus depending on her rank in the subjective organisational input distribution as perceived by the manager. If the management believes that the foreman is more engaged in organising the team's production relative to her counterparts, the foreman receives a higher rank which provides her with a higher bonus. For simplicity, we opt for a random variable $B(g_f)$ to capture the incentive scheme which is based on an individual's relative position in the firm (see Lazear and Rosen (1981) for details of modelling rank-order tournament incentives). We impose a standard set of conditions on $B(g_f)$ as below:

- $B(g_f)$ is strictly convex and is continuously differentiable on its domain and
- $B'(g_f)$ is strictly positive where the superscript denotes the derivative with respect to g_f.

Thus, a foreman expects a higher bonus that is paid for her organisational activities as she increases effort in organising team production.

Furthermore, we assume that the market price in our case is high (i.e. w^* is offered to the employees and $1 < \lambda w^* < 2$) because the firm intends to increase its compensation paid for foreman's organisational activities. Since the firm does not readjust the piece rate scheme or re-maximise its expected profit after introducing the new incentive scheme, individual piece rate is given at w^*.[10] Thus, the foreman's maximisation problem is now changed to:

$$\begin{aligned}\max_{\{e_f,g_f\}} EU_f(e_f,g_f) &= E[B(g_f) + w^* y_f(e_f,g_f) - C(e_f,g_f)] \\ &= EB(g_f) + w^*[e_f(1+\lambda g_f)] - \frac{(e_f+g_f)^2}{2}.\end{aligned}$$

This leads to the first order condition with respect to e_f:

$$e_f = w^* + (\lambda w^* - 1)g_f,$$

and the first order condition with respect to g_f:

$$g_f = \frac{dEB(g_f)}{dg_f} + (\lambda w^* - 1)e_f,$$

[10] The management told us that it is extremely difficult to adjust the piece rate in the workplace, especially decrease the rate, as the employees are very defensive about changes in their performance-related compensations.

For the worker w, her maximisation problem is unchanged which implies that the first order condition with respect to her effort choice is given as below:

$$e_w = w^*(1 + \lambda g_f).$$

Solving these first order conditions we get the optimal effort levels of the foreman:

$$\widehat{e}_f = \frac{(\lambda w^* - 1)}{1 - (\lambda w^* - 1)^2} * \frac{dEB(\widehat{g}_f)}{dg_f} + \frac{w^*}{1 - (\lambda w^* - 1)^2}, \tag{2.4}$$

$$\widehat{g}_f = \frac{1}{1 - (\lambda w^* - 1)^2} * \frac{dEB(\widehat{g}_f)}{dg_f} + \frac{w^*(\lambda w^* - 1)}{1 - (\lambda w^* - 1)^2}. \tag{2.5}$$

For the worker, the optimal production effort level is given by:

$$\widehat{e}_w = \frac{\lambda w^*}{(1 - (\lambda w^* - 1)^2} * \frac{dEB(\widehat{g}_f)}{dg_f} + \frac{\lambda (w^*)^2}{1 - (\lambda w^* - 1)^2}. \tag{2.6}$$

C. Optimal Effort levels: Pre-Intervention vs Post-Intervention

In this subsection, we aim to show whether the firm does successfully increase the optimal effort levels of both the foreman and the worker by subjectively evaluating the foremen's organisational activities and offering the foreman a bonus depending on her relative position within the firm. As the piece rate scheme is constant throughout, taking the differences of the optimal effort levels between the pre-intervention and the post-intervention gives us:

$$\begin{aligned}\widehat{e}_f - e_f^* &= \frac{(\lambda w^* - 1)}{1 - (\lambda w^* - 1)^2} * \frac{dEB(\widehat{g}_f)}{dg_f} + \frac{w^*}{1 - (\lambda w^* - 1)^2} - \frac{w^*}{1 - (\lambda w^* - 1)^2} \\ &= \frac{(\lambda w^* - 1)}{1 - (\lambda w^* - 1)^2} * \frac{dEB(\widehat{g}_f)}{dg_f},\end{aligned} \tag{2.7}$$

$$\begin{aligned}\widehat{g}_f - g_f^* &= \frac{1}{1 - (\lambda w^* - 1)^2} * \frac{dEB(\widehat{g}_f)}{dg_f} + \frac{w^*(\lambda w^* - 1)}{1 - (\lambda w^* - 1)^2} - \frac{w^*(\lambda w^* - 1)}{1 - (\lambda w^* - 1)^2} \\ &= \frac{1}{1 - (\lambda w^* - 1)^2} * \frac{dEB(\widehat{g}_f)}{dg_f},\end{aligned} \tag{2.8}$$

$$\begin{aligned}\widehat{e}_w - e_w^* &= \frac{\lambda w^*}{(1-(\lambda w^*-1)^2} * \frac{dEB(\widehat{g}_f)}{dg_f} + \frac{\lambda (w^*)^2}{1-(\lambda w^*-1)^2} - \frac{\lambda (w^*)^2}{1-(\lambda w^*-1)^2} \\ &= \frac{\lambda w^*}{(1-(\lambda w^*-1)^2} * \frac{dEB(\widehat{g}_f)}{dg_f}.\end{aligned} \tag{2.9}$$

The assumptions $1 < \lambda w^* < 2$ and $\frac{dEB(\widehat{g}_f)}{dg_f} > 0$ imply that optimal effort levels are higher when the foreman is offered a performance-related incentive concerning her organisational activities. As this higher-powered incentive increases the foreman's organisational inputs, production efforts of the foreman and the worker also increase because of the team efficiency spillover λ.

It is important to note that the first order condition for the foreman's maximisation problem with respect to her organisational inputs g_f has an extra positive term $\frac{dEB(g_f)}{dg_f}$ during the post-intervention period relative to the pre-intervention period. This implies that the foreman's organisational inputs would increase from zero to above zero after the intervention even in the case of $\lambda\overline{w} < 1$ provided that $\frac{dEB(g_f)}{dg_f} + (\lambda\overline{w} - 1)\overline{w} > 0$. However, the foreman's production effort would then fall below $\overline{w}$ in this regard as it is equal to $\overline{w} + (\lambda\overline{w} - 1)g_f$ and the second term is negative. In the case of $\lambda\overline{w} = 1$, the foreman's organisational input is guaranteed to increase while her production effort remains unchanged.

Recall that the production function is increasing in g_f, e_f, and e_w. If the foreman and the worker increase their effort levels, the production output grows after the intervention. In the case that the foreman's g_f increases while e_f decreases, the foreman's output would fall if $\lambda\overline{w} + (\lambda\overline{w} - 1)(1 + \lambda g_f) < 0$ where $\lambda\overline{w} < 1$.[11]

Furthermore, taking the difference between equation 2.7 and equation 2.9, we get:

$$\begin{aligned}\widehat{e}_w - e_w^* - (\widehat{e}_f - e_f^*) &= \frac{\lambda w^*}{(1-(\lambda w^*-1)^2} * \frac{dEB(\widehat{g}_f)}{dg_f} - \frac{(\lambda w^*-1)}{1-(\lambda w^*-1)^2} * \frac{dEB(\widehat{g}_f)}{dg_f} \\ &= \frac{1}{(1-(\lambda w^*-1)^2} * \frac{dEB(\widehat{g}_f)}{dg_f} > 0,\end{aligned}$$

this indicates that the worker increases more production effort than the foreman. This result is rather trivial since it is directly imposed by the production technology. The first order conditions for both the foreman's and the worker's maximisation problems with respect to their production effort indicate that an increase in the foreman's organisational inputs has a larger impact on the worker's production effort (multiplied by λw^*) than the foreman's (with multiplier equals to $\lambda w^* - 1$).

[11]When $\lambda\overline{w} < 1$, the foreman produces $\overline{w}$ in the pre-intervention period as $\overline{g}_f = 0$ and $\overline{e}_f = \overline{w}$. After the introduction of the new bonus scheme, g_f becomes positive if $\frac{dEB(g_f)}{dg_f} + (\lambda\overline{w} - 1)\overline{w} > 0$ and the foreman produces $\overline{w} + [\lambda\overline{w} + (\lambda\overline{w} - 1)(1 + \lambda g_f)]g_f$ given that $\overline{w}$ remains constant. Therefore, there would be a decrease in the foreman's output if $\lambda\overline{w} + (\lambda\overline{w} - 1)(1 + \lambda g_f) < 0$.

Proposition 2.2 (Predictions)**.** *To summarise the theoretical model would predict the following:*

(i) The introduction of a performance-related bonus scheme regarding foreman's organisational activities (weakly) increases the foreman's g_f, holding w constant, and strictly so for some types of bonuses.

(ii) When the market price and λ are sufficiently high, introducing a higher-powered incentive scheme concerning the foreman's organisational activities increases the foreman's e_f even though she is multitasking. However, if the market price or λ is extremely low and the slope of the new bonus scheme is large enough, the foreman's choice of e_f would decrease.

(iii) e_w increases when g_f increases.

(iv) For a given w, the increase in g_f as a result of the new bonus scheme has a larger impact on e_w than on e_f.

(v) Production output increases after the introduction of the new incentive scheme if effort level increases. The foreman's output may fall if she increases g_f while reducing e_f.

2.3 The Firm and Experimental Design

2.3.1 Production Setting

We conducted an experiment in two sister medical-device companies between June 2017 and September 2017. Both companies are located on the northern side of Jiangxi, a southeastern non-coastal province of China. Each sister company has its own personnel and branding, they are not closely related and have limited interactions with each other below the top-level management. One company is located in Fuzhou prefecture while the other is in Fengcheng county, the driving distance between Fuzhou and Fengcheng is about seventy-five miles. For simplicity, we refer the former company as Fuzhou and the latter as Fengcheng below.

In our sample, 70 regular employees (all females, 27 from Fuzhou and 43 from Fengcheng respectively) produce disposable infusion sets in an assembly-line fashion but noncontinuous (see Online Appendix).[12][13] Each production line is composed of numerous workers and one foreman. Seven lines (consist with one foreman and averagely four workers per line) operate regularly in Fuzhou, while six lines (consist with one foreman and averagely six workers per line) operate regularly in Fengcheng. Their backgrounds are mostly local farmers. The manufacturing task for both workers and foremen is a supporting work, product packaging, which

[12]This sample excludes newly hired workers because their compensation schemes are different from those who work more than three months, a few workers who we do not have records during the status quo hence they are not valid for our difference in difference estimation, and some workers who left before the experiment ended as they may respond differently to our treatment.

[13]The disposable infusion set is a major source of revenue for this company, accounting for approximately 50% of its total revenue based on the data in 2016.

requires relatively little training or human capital. Salary schemes for such task are identical to the compensation schemes offered to other tasks within the same production unit such as assembling, leak testing, or pressure testing. Thus, it is unlikely that workers in our experiments have sorted themselves out to a specific kind of base salary structure by their choice of profession.

In addition to the manufacturing job, the foreman in each line is also responsible for monitoring workers' performance, organising and distributing materials, and assisting production managers on production matters. According to the production managers, foremen are internally promoted only, and a successful candidate should be able to demonstrate her loyalty to the company, reliability, and modest leadership. However, qualitative evidence from interviewing the workers and forewomen reveals that no one craves the foreman position because it requires more effort, sidetracks them from the primary task, and the corresponding compensation is relatively low. All existing forewomen have worked in the company for more than two years. They had established a good rapport with the production managers over the years. They accepted the foreman appointment mainly because they ran out of excuses to reject it again. This suggests that we would not expect certain types of worker deliberately stand out to be a foreman. The sorting effect is negligible.

Both factories offer an individual-based multiple piece rate payment scheme (i.e., producing more outputs yields higher rates) to employees for packing the products. In each month, in addition to the piece rate, their base salary comprises some other bonuses and vary across factories. Table 2.1 summaries the various piece rates and bonuses by factory. This salary structure is consistent throughout our experiment.

Table 2.1: Summary of Wage Structure by Factory

Daily Average Output	Piece Rate (per unit) (1)	Performance Bonus (2)	Attendance Bonus (3)	Tenure Bonus (4)	Lunch Subsidy (5)	Foreman Subsidy (6)
Panel A. Fengcheng						
Less than 2,400	.0195	200	30	50	42	90
2,400 - 2,600	.0205	200	30	50	42	90
2,600 - 2,800	.0210	200	30	50	42	90
2,800 - 3,000	.0225	200	30	50	42	90
3,000 - 3,200	.0230	250	30	50	42	90
3,200 - 3,400	.0235	250	30	50	42	90
More than 3,400	.0240	300	30	50	42	90
Panel B. Fuzhou						
Less than 3,100	.0188	60	40	65	60	40
3,100 - 3,500	.0193	80	40	65	60	40
More than 3,500	.0196	100	40	65	60	40

A forewoman is given an additional flat rate on a monthly basis for her services regardless of outputs. This rate is higher (more than doubled) in Fengcheng than in Fuzhou. In particular, it equals 40 RMB ($\approx$ 6 dollars) in Fuzhou as shown in Column 6 of Table 2.1, which is roughly 2% of a forewoman's monthly income. In Fengcheng, a forewoman receives an extra 90 RMB ($\approx$ 13.5 dollars) per month, which is about 3% of her monthly income on average. Similarly, the piece rates offered in Fengcheng are higher than those in Fuzhou. A fast-packaging worker, who can make more than 3,500 units averagely in a day, is given 0.0044 RMB more per unit in Fengcheng in comparison to Fuzhou. The daily average output is calculated by dividing the total production output in a month by the number of days worked during that month. This implies a difference of 430 RMB ($\approx$ 65 dollars) in 28 working days. Therefore, either the worker or the forewoman working at the Fengcheng factory earns 20 percent more income than in Fuzhou.

These differences do not necessarily raise a concern. The two factories are independently operated and organised. The decision, with regard to the rates, made by each factory manager was unassociated. It is mainly determined by the condition of the local labour market. Furthermore, employees work at the production level in one factory barely know any information about the other factory. They do not choose which factory to work.

In a manufacturing setting like ours where employee turnover rate is high, it is unlikely that the management allocates workers into groups randomly. In general, there are two types of relocations of workers. One type of relocations only applies to the newly hired employees. The management usually separates the newly hired employees from the regular workers and assigns them to work on a different line as they call "probation line". Based on the turnover rate of regular workers, the management assigns these newly hired ones to fill the vacancies in the regular operating production lines. Some new workers may stay at the probation line for more than 5 months while others may be relocated to the regular operating lines in 1- or 2-months' time. The other type of relocations is an extreme case and takes place when there are no new employees at stock and the turnover rate is high (e.g. before the Chinese New Year), or when there is a technical breakdown at the downstream of the production. The management may disband one regular operating line and randomly assign these workers or allow them to self-select into other lines.

Nevertheless, the forewoman of each line does not change in general unless the management decides to disband a production line for more than a month and the forewoman of the dismantled line is assigned to work on another line. This didn't occur throughout our experiment. A change in forewomen also occurs if a forewomen decides to quit the job or leave for a long time because of sickness, in this case a new forewoman will be appointed. This type of forewoman change did occur during our experimental period, in our analysis we exclude the observations of workers who had worked with the new forewoman.

In the first relocation scenario which I described above, it should not raise concerns for our identification strategy. First of all, workers who were hired during our experimental period or three months (probation period) before our experiment started were excluded in our analysis. Secondly, according to the managers, new workers are very unlikely to affect group norms (if any group specific norms are present due to line composition) as they are already taking a highly demanding production task. It is unlikely that they have time and energy to change the group norm or establish a rapport with the line foreman in a short time. Lastly, all forewomen in our sample have at least one-year experience of being the group leader. Any relationships/norms that have been established between the regular line workers and the forewomen should be captured by the individual fixed effects.

For the extreme case, no regular operating lines had been disbanded by the management during our experimental period. But, there were a couple of incidences took place at the downstream of the production lines during the experiment, and some workers were assigned to other lines. Nevertheless, these relocations do not decrease the effectiveness of organisation. The relocated workers only changed their working locations. They continued to report daily outputs or personal issues to their initial line forewoman. Therefore, such relocations can potentially have an impact on workers' productivity (because of the location of lines, some production lines may have easier access to raw materials or better illumination and temperature conditions), but there is no impact on the effectiveness of forewomen's organisation. Each forewoman continued to be responsible for her line workers who changed working locations. This may increase the cost of organisation, but managers would have taken this into account when they were evaluating the forewomen. Hence, we include an indicator which captures the incidences of workers changing their locations to control for these effects.

2.3.2 The Field Experiment

During our experiment (between 7^{th} June 2017 and 30^{th} September 2017), both workers and forewomen performed their tasks individually within their natural work environment. They were unaware of an experiment was taking place. The treatment was introduced to them by the production manager via the internal communication channel in Fengcheng factory. The reason Fengcheng was selected for implementing the treatment is that the factory manager in Fuzhou had unexpectedly submitted his resignation in June. This makes the Fuzhou factory a natural setting as a control group since the workplace condition is most likely to remain constant during the transition period. Most importantly, the board of the company shared the same notion and agreed to introduce the treatment in Fengcheng while keeping Fuzhou as constant for this period of time. Therefore, we denote our experiment as a field experiment, following the terminology of Harrison and List (2004).

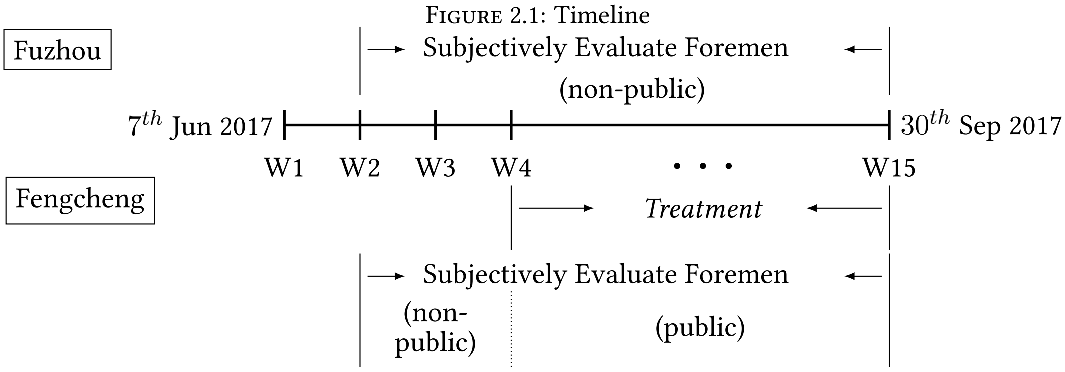
2.3.2.1 Timeline

The experiment was designed to generate exogenous variation in the trade-off between performing the perfectly measurable task and partially quantifiable job. The timeline of the experiment is shown in Figure 2.1. Starting from 7^{th} June 2017 individual daily production records were collected and monitored by our research team.[14] During the first experimental week (W1), production managers from both factories were trained to use the evaluation system we

[14] The factories were not recording worker's daily productivity before our intervention. They only collected the output data. Therefore, historical production data from these factories cannot be used in this study.

designed to assess the forewomen subjectively (discussed in the next subsection). On the last day of the second week (W2), forewomen's organisational activities during W2 were subjectively evaluated by the production managers in both factories while neither the workers nor the forewomen were aware of this evaluation. In each following week, the production manager performed a weekly independent evaluation of the forewomen's organisational activities. Notice that the evaluation process and the corresponding results were never made public in Fuzhou throughout the experiment. Whereas both workers and forewomen in Fengcheng were well informed about the evaluation process, how the forewomen are evaluated, and their evaluation results after the treatments were introduced (weeks 4-15). For instance, ranking results for W4 were posted on the factory floor at the end of the fourth week.

FIGURE 2.1: Timeline



Notes: W denotes the experimental week.

On the last day of week 3 (30^{th} June 2017), the production manager in Fengcheng had a regular monthly meeting with all workers and forewomen from the packaging unit. The manager ventilated production issues raised during that month and outlined plans for the upcoming months including our treatment. We instructed the manager to announce our treatment as follows. The production managers will subjectively evaluate forewomen's organisational activities each week. The evaluation starts from 1^{st} July 2017. Four criteria regarding management and organisation will be assessed. At the end of each week, the ranking for each evaluated criterion and the weekly overall ranking will be updated on the whiteboards located next to the production lines. All weekly rankings within a month are important as they will be used to compute the ranking of the month. On the last day of each month, every forewoman will receive a pecuniary reward in cash based on her monthly ranking.[15] A higher ranking yields a higher payment. The monthly ranking is then reset at the beginning of next month. In other words, both weekly rankings and monthly rankings are intra-independent.[16]

A detailed instruction was handed to each forewoman after the meeting. It illustrated the four criteria assessed with brief examples, detailed the incentive scheme, and outlined other

[15]Even though the employees in this firm are paid one month in arrears, receiving prizes in advance is not fresh. There were policies rewarded employees in advance such as bringing in new recruits.

[16]There is a reason for introducing the treatment by the end of June. Individual piece rate is not constant. It is determined by the worker's monthly output. The more they produce, the higher the piece rate. Hence it affects workers' manufacturing behaviour over time. A worker may feel sluggish at one point when she realises that she can no longer reach the next higher piece rate, and vice versa. Therefore, the production manager advised us to not interfere in the middle of a month.

organisation-related information. Notably, we explicitly stated that this monetary prize is independent of the current foreman subsidy. Hence the forewomen would not consider this as a replacement of the current subsidy.

2.3.2.2 Subjective Evaluation

In each factory, more than one manager was asked to perform the evaluation task. Two managers in Fengcheng were assigned to perform the evaluation task, while there are three evaluators in Fuzhou. The reason for appointing at least two direct managers to assess the forewomen subjectively was threefold. It is consistent with other evaluations that were organised in the factories, such as the 5S system which was assessed by five managers.[17] This prevents manager's personal perceptions and biases to influence the evaluation results. Hence forewomen are less likely to question the fairness and accuracy. Moreover, employing multiple examiners increases forewomen's costs for collusion. Last but not the least, on some occasions one manager can still provide reliable evaluations when the other is absent due to illness or on holiday.

We consulted the management to list all the organisational activities they demand the forewomen to perform. The management proposed and exemplified four criteria which we embedded in the evaluation system, including: maintain an efficient production process (e.g. make sure the raw materials are sufficiently and unerringly distributed on the line for workers to work); increase the productivity of the line workers (e.g. manage the team effectively so that workers work efficiently, such as talk to the workers and motivate them to focus on working); reduce line defect rates (e.g. constantly remind workers to use standardised operating procedure in order to reduce the number of faulty products); and team building (e.g. provide support and communication to foster a friendly and positive work environment). Each indicator along with its associated example is clearly elucidated in the instructions given to the forewomen.

For efficiency purposes, we designed a novel spreadsheet to minimise the time required for the production managers to perform the subjective evaluation (see Figure B3 in the appendix as an example). Production managers were asked to provide relative performance evaluation by positioning sliders instead of giving exact scores to avoid a tie. We underlined that the positions of sliders within each assessed criterion should be unalike across forewomen since these are relative measurements. After positioning the sliders under each criterion, the overall ranking of each forewoman is automatically calculated and displayed. The examiners were then asked to verify whether the overall rankings are authentic. If not, they were instructed to repositioning the sliders without altering the ranking of each criterion until the valid overall rankings were reached.

In each week, both the ranking for each evaluated criterion and the overall week-ranking were posted on the Fengcheng factory floor in the form of a scoreboard and displayed in descending

[17] 5S is a workplace organisation system designed to improve manufacturing efficiency. For details, see https://en.wikipedia.org/wiki/5S_(methodology).

order. The management was instructed to put up this ranking board on the wall next to the production lines as shown in the Online Appendix. For consistency, the scoreboard only provides information for each week. At the end of each month, a hard copy which summarises four weekly rankings and the aggregated rankings of that month was posted next to the scoreboard.

2.3.2.3 Monetary Prizes

The management insisted to reward all forewomen rather than the highest ranked one(s). They do not have pleasant experiences with only rewarding the best employee(s) and find this type of incentive structure divisive. Therefore, we designed an incentive scheme which rewards every forewoman as presented in Table 2.2. Both Table 2.2 and the reasoning of this design (as we discuss below) were documented in the instructions given to the forewomen.

Table 2.2: Monetary Prizes

	Initial Foreman Fee (RMB/M) (1)	Tournament Reward (RMB/M) (2)	Difference from the next lower rank (3)	Change in Total Foreman Fee (%) (4)
#.1 ranked forewoman	90	205	45	228%
#.2 ranked forewoman	90	160	25	178%
#.3 ranked forewoman	90	135	15	150%
#.4 ranked forewoman	90	120	10	133%
#.5 ranked forewoman	90	110	10	122%
#.6 ranked forewoman	90	100	10	111%
#.7 ranked forewoman	90	90		100%

Notes: 0 RMB will be paid to the forewoman if she is eliminated by the management. RMB/M denotes Chinese yuan per month.

The lowest ranked forewoman is paid 90 RMB per month, which is identical to the amount of money the company paid to the forewoman for her leadership role. To further determine the amount of payment given to the highest ranked forewoman, we first calculate the highest piece-rate wage per hour a forewoman can possibly get. For a forewoman to be eligible for the highest piece rate .024 as shown in Table 2.1, she has to produce at least 3,400 units every day. It implies that the hourly output is 310 units for a forewoman works 11 hours a day. Therefore, packing products for an hour earns her 7.44 RMB. This suggests an opportunity cost of spending one hour per day on organising teams for 28 working days is 208 RMB.[18] Similarly, for the least productive forewomen her opportunity cost of spending one hour per day on organising teams for 28 working days is 94 RMB since our data indicate that the mean of the daily output of forewomen in Fengcheng is 2,600 units, the fastest forewoman can produce 4,400 units while the slowest forewoman only produces 1,900 units in a single day.[19] Notice that before our experiment not more than two forewomen could get the highest piece rate.

[18] In general, workers from Fengcheng factory work 11 hours per day and 28 days per month.
[19] The corresponding piece rate for 1,900 daily output is .0195, hence $94 = 1900 \div 11 \times .0195 \times 28$.

By intentionally making the top three ranks more attractive, the highest ranked forewoman receives 205 RMB per month as shown in Table 2.2. In particular, the highest ranked forewoman is paid 45 RMB more relative to the second highest ranked forewoman, 70 RMB more than the third highest ranked forewoman, and 85 RMB more than the fourth highest ranked forewoman. For the lower ranked forewomen (rank 4-7), the payment differences between individuals adjacent in rank are parallel which equals 10 RMB. In the case of a tie, a standard competition ranking is applied, i.e. all forewomen will be paid 205 RMB if they share the same score.

Therefore, the incentive for spending one hour per day to organise the team is as attractive as the hourly piece-rate wage. The highest ranked forewoman in the monthly tournament receives 205 RMB while the most productive forewoman is paid 208 RMB for 28 hours (i.e. one hour per day for 28 days). As for the lowest ranked forewoman, if she also turns out to be unproductive, spending one hour in a day to organise the team for 28 days gives her 90 RMB which is identical to the amount of payment for packaging the products for 28 hours - 94 RMB. Since the ranking is based on subjective evaluation results and forewomen's relative performance, the corresponding cost is perceived to be lower than packing products as reviewed in the interviews which we conducted with the forewomen after the experiment. If the most productive forewoman is given the lowest rank, she will have a strong motivation to invest more efforts in organising the team in the following weeks for a better rank. Therefore, by providing equivalent incentives and incrementally increasing the prizes when the ranking increases we believe that this scheme can offer sufficient incentives to the forewomen to spend around one hour per day to perform organisational tasks as the management craves.

Last but not the least, forewomen were also informed that the management was given the authority to eliminate their eligibility for the prizes. If the direct managers conclude that a forewoman had exerted zero efforts on any of the assessed criteria (i.e. maintain an efficient production process, increase the productivity of the line workers, reduce line defect rates, or team building), the forewoman will not be given the bonus in that month. Nonetheless, no forewoman was eliminated for the prize throughout the experiment.

2.4 Empirical Analysis

To test whether the introduction of an incentive scheme regarding foremen's organisational behaviours affects either the workers' or the foreman's productivity, we exploit the fact that workers and forewomen from both factories are observed over time. We estimate the following Difference-in-Difference (DiD) specification for individual i in factory f and day t:

$$log(Y)_{i,f,t} = \beta \mathit{FACTORY}_f + \gamma \mathit{POST}_t + \eta' Z_{i,f,t} + \rho I_i + \delta D_{f,t} + \epsilon_{i,f,t}, \tag{2.10}$$

where $log(Y)_{i,f,t}$ is the logged production-related outcomes of individual i (we analyse workers and forewomen separately) in factory f and day t, in particular, we are interested in production output and productivity. The productivity in our case is defined as a measure of the output per hour. The line forewomen are responsible for recording the data of every individual from the same production line including the daily output, time work started, and time left the factory. These figures are further scrutinised by the production manager with little measurement error and used to compute the daily productivity.

$FACTORY_f$ and $POST_t$ are dichotomous variables indicating the treatment factory f and the treatment period t, respectively. To take account of the natural trends in production process we control for the time-varying determinants $Z_{i,f,t}$: (i) the production line individual i worked on day t (i.e. line fixed effects), workers are assigned to work on the other production line when the downstream of her own line is disordered, to capture the variation in team efficiency spillover between production lines and account for unobserved and permanent differences in productivity across lines (e.g. distance to raw materials); (ii) an indicator variable for whether individual i is recorded sick or if there is an organisational error, which may cause negative sentiments and therefore reduce performance; (iii) a vector of variables captures the time effects including experimental weeks and the day of each week (e.g. Monday), to control for the time trend which influences individual i's productivity. Individual fixed effects I_i account for unobserved and time-invariant heterogeneities in productivity among individuals.

Finally, $D_{f,t}$ is the interaction of $FACTORY_f$ and $POST_t$ which equals to 1 if individual i is working at the factory where the treatment is already taking place. Therefore, δ is the coefficient of interest. It estimates our treatment effect. The disturbance term $\epsilon_{i,f,t}$ is individual specific. We present estimates under the fixed effects framework while the results are robust qualitatively or quantitatively under the random effects specifications. Heteroskedasticity-robust standard errors clustered at the individual level are used in all regression specifications. This allows us to address the concern that observations for an individual are not independent over time.[20]

The most critical assumption of the DiD is that the treatment and control factories have pre-treatment parallel trends in the outcome. In principle, the treatment factory and control factory are a good match. They are two sister companies which share the same board and corporate culture. Workers' incentive structure does not differ qualitatively but varies quantitatively. The marginal variations in quantity are mainly driven by the condition of the local labour market which is exogenous to the workers' outcomes. Therefore, working patterns in these two factories should be comparable.

[20] In general, the variation in worker production outcome over time across workers should be independent. There are cases where workers are likely to be dependent when the upstream production unit is short-handed. This is because all lines acquire manufacturing materials from the same upstream, the faster a productive worker (or a line collectively) can consume the materials the more likely the less productive ones have to wait for materials. This wastes the less productive workers' time on production and leads to a fall in productivity. For robustness, we applied the wild cluster bootstrap (see Cameron et al., 2008, for details) while clustering at both individual and line levels. The main results remain unchanged qualitatively.

2.4.1 Descriptive Evidence

To form a consistent sample throughout our analysis, we exclude workers who were new recruits and those who were working in their three-month probation period when the experiment started.[21] Employees under their probation period are offered a different payment scheme - hourly rate - in comparison to regular workers who are paid by piece rate. This implies that employees who were hired after the first day of March 2017 are excluded from the sample. A few workers from the treated factory were on holidays when we started the experiment in June, and they returned to work after the treatment was introduced. They are also not valid for the difference-in-difference estimation. Hence we leave them out of the analysis. Moreover, there are some workers left the factory during the experiment. It is reasonable to assume that they may not respond to our treatment because they intend to leave. Therefore, we exclude them from the analysis and opt for a clear measure of the treatment effect. In total, 70 workers constitute the final sample with 27 workers from the controlled factory Fuzhou and 43 in the treated factory Fengcheng.

In addition, the types of the infusion sets packed are different across factories. Products sold in the domestic market are easier to pack than those sold in the international market. When Fuzhou factory mainly focuses on the local market, Fengcheng factory produces goods for both markets. Luckily, the management of Fengcheng factory has developed methods to calculate piece rates for different types of products based on the level of difficulty. We used the same technology to standardise individual outputs in Fengcheng factory so that figures are comparable across factories.

Table 2.3 presents summary statistics for each factory during the pre-treatment period (June) and the post-treatment period (July, August, and September), including number of employees, number of production lines (which is also the number of forewomen, as there is only one forewoman assigned to each line), worker's daily output, worker's productivity (output per hour), forewoman's daily output, and forewoman's productivity.

Hourly productivity is the ratio of daily output to the total hours worked in that day. The total number of hours worked per day is derived by the difference between the time when the individual started her work and the time when she left the production line. We do not observe the precise time an individual had spent on the manufacturing task. Nevertheless, when we use the difference-in-difference estimation and assume that individual's work behaviour is constant, the treatment effects estimated are valid.

As shown in column 4 of Table 2.3, in August, one worker in factory Fuzhou was absent for the whole month because of illness, and one worker in factory Fengcheng was assigned to another production unit which is not included in our sample. In September (column 5), one forewoman from factory Fengcheng got sick and left the job for two weeks whose foreman responsibility was soon succeeded by a line supervisor. However, a line supervisor does not

[21] All forewomen had more than two years working experience in the factories.

Table 2.3: Summary Statistics

	June	Jul-Sep	Jul	Aug	Sep
	(1)	(2)	(3)	(4)	(5)
Panel A. Fengcheng					
Number of Employees	43	41.10	43	42	38
Number of Lines	6	6	6	6	5
Worker Daily Output	1,125.9	1,176.8	1,179.4	1,161.3	1,190.5
	(242.2)	(248.6)	(236.0)	(268.8)	(239.0)
Worker Hourly Productivity	91.74	96.86	93.94	97.25	99.67
	(17.87)	(17.45)	(16.57)	(18.08)	(17.22)
Forewoman Daily Output	1,027.8	1,093.9	1,082.0	1,092.4	1,109.6
	(210.4)	(250.2)	(222.7)	(274.0)	(252.3)
Forewoman Hourly Productivity	85.65	89.98	86.67	90.40	93.34
	(20.03)	(20.10)	(19.24)	(20.52)	(20.13)
Panel B. Fuzhou					
Number of Employees	27	26.71	27	26	27
Number of Lines	7	7	7	7	7
Worker Daily Output	1,082.9	1,049.7	1,039.4	1,032.7	1,072.4
	(221.9)	(242.4)	(228.5)	(284.8)	(216.6)
Worker Hourly Productivity	93.04	91.98	90.95	91.94	92.96
	(16.68)	(13.60)	(14.24)	(14.23)	(12.39)
Forewoman Daily Output	1,121.4	1,087.3	1,067.1	1,073.5	1,119.5
	(172.9)	(197.7)	(179.6)	(237.7)	(171.4)
Forewoman Hourly Productivity	94.72	94.97	93.51	95.00	96.39
	(12.62)	(10.35)	(10.07)	(11.08)	(9.809)

Notes: Productivity is a measure of the output per hour. June indicates the pre-treatment period and Jul-Sep implies the post-treatment period. The top number in each cell denotes the mean and the number in parentheses denotes the standard deviation.

perform manufacturing tasks. We have neither production records to determine her productivity nor the relative performance in organising teams. Observations of workers from this line and the newly appointed line supervisor are not included in our sample.[22]

The contrasting patterns in Table 2.3 are that performance was increasing from June onwards in Fengcheng but decreasing in Fuzhou. In particular, workers' daily output, workers' hourly productivity, and forewomen's daily output in Fuzhou were all lower in July-September in relation to June, and forewomen's hourly productivity in July is lower than their productivity in June. According to the company, this decline in Fuzhou is a normal pattern that applies to both factories. It is mainly due to the weather. Temperature reaches the peak of each year during July and August. Because the factories produce medical appliances, the workplaces are sterile, clean, and purified plants. Workers must wear impervious gowns in the workplace to reduce the risk of contamination. When the temperature gets high, the environment becomes too uncomfortable to work, and the production performance falls sharply around this time of the year. Therefore, in general, performance is expected to fall in both factories during this

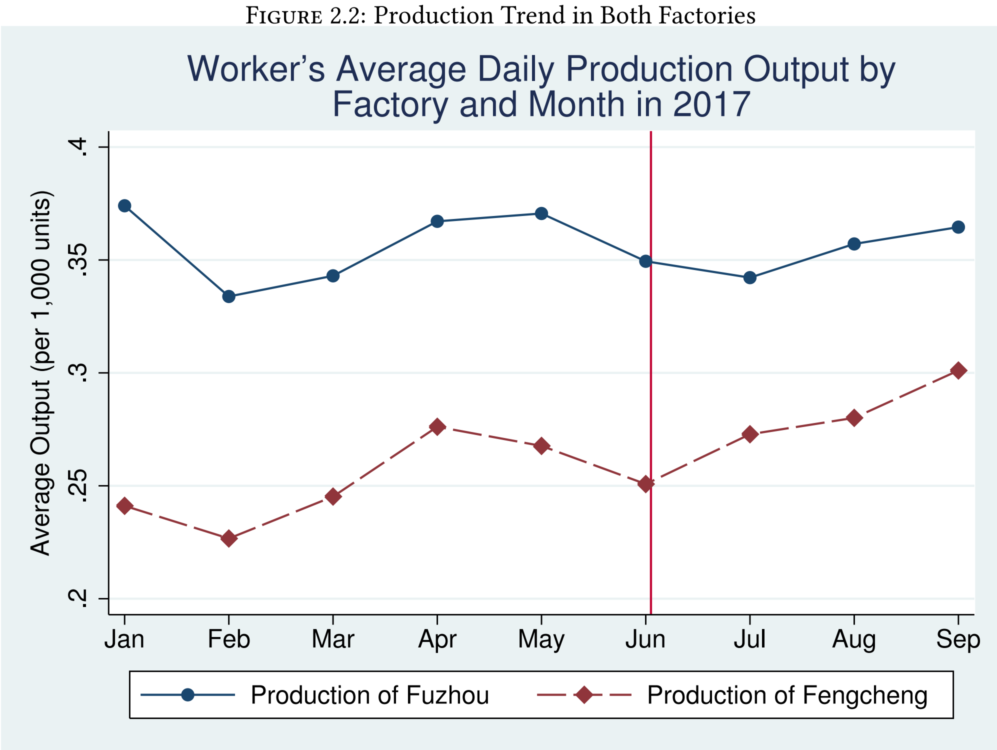
[22]Summary statistics for employees' other characteristics which we collected after the experiment in factories Fengcheng and Fuzhou are reported in Tables C1 and C2 in the appendix, respectively.

period. The fall in Fuzhou is following this normal trend, while the contrasting increase in Fengcheng is due to the introduction of our treatment.[23]

Furthermore, as the number of days a worker worked in each month vary across workers and months, the summary statistics of monthly production outcomes aggregating observations from all workers are not particularly informative. We discuss the results with the assistance of other analyses in the next subsections.

We are not able to show the trends in workers' productivity beyond our intervention because the company only has workers' output data to determine their salary. But, we are able to show the evidence of a parallel pre-trend based on the firm's administrative data. Figure 2.2 shows the average daily production output of workers for both the treatment and the control factory in each month in 2017 before our experiment ended (by September).

FIGURE 2.2: Production Trend in Both Factories



Notes: The vertical line indicates the introduction of our treatment in Fengcheng.

It indicates that there is a comparable trend in workers' daily output before our treatment was introduced in July 2017, with the exception of January. The variations between January and February are subject to the Chinese New Year. Depending on the condition of local labour markets and the turnover rate of workers, for instance, the output would fall deeper in February if the factory experienced a hard time to retain its workers and to hire new ones (such as Fuzhou). Therefore, the first trustworthy data point is February. The movements of these

[23]In the Online Appendix, I provide the evidence from WorldWeatherOnline.com which shows the maximum temperature in both cities is above 35-degree Celsius between July and August.

two factories between February and June are indeed parallel. This is not surprising because, as I discussed above, both factories share the same company culture, and their remuneration systems are qualitatively identical.

Furthermore, this administrative data can also illustrate our treatment effect. While the daily output of workers in Fuzhou (the solid line) fluctuates at its normal interval, the figures in Fengcheng (the dashed line) started to rise after the introduction of our treatment (in July) and further exceeded the peak of the year (on April 2017) since August.

2.4.2 Performance of Workers

First, we graphically compare worker's hourly productivity in the treated factory with the one in the controlled factory before and after the treatment was introduced. Recall that the treatment was introduced in Fengcheng by the end of the third experimental week, and it lasted from week 4 to the last week of our experiment. For each factory, we calculated the average output per hour of all workers in each week. Figure 2.3 depicts the mean of worker's productivity, averaged across all workers, in each experimental week. The area between two dashed lines corresponds to the 95% confidence intervals. The figure suggests that worker's productivity increased marginally during the first four weeks of treatment while worker's productivity in the controlled factory is somewhat flat. The treatment effect started to rise dramatically in the second treatment month. It did not decrease over time albeit there was a decline during the last few experimental weeks. This drop is mainly due to an unexpected incident took place in the treated factory which we will discuss below. On the other hand, in the status quo (the first three experimental weeks), we do observe a similar trend between factories. This suggests that the parallel trends assumption for the difference-in-difference estimation is satisfied.[24] [25]

To present formal evidence on the effect of the introduction of subjective evaluations and monetary prizes regarding forewomen's organisational behaviours on worker's performance, we estimate specification 2.10.

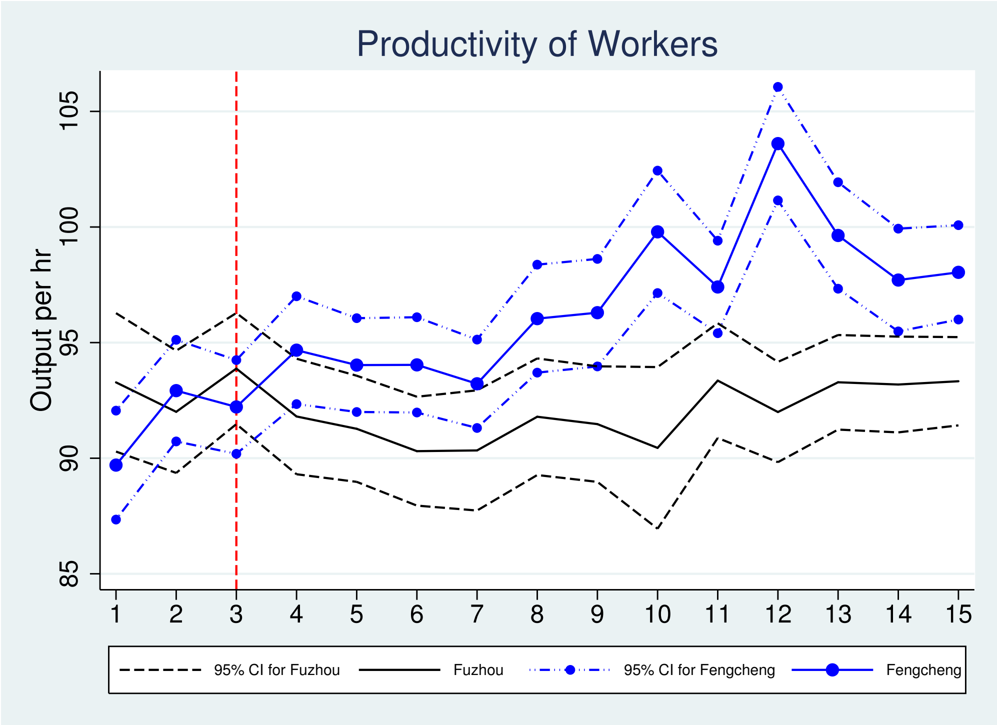
Columns (1) and (5) of Table 2.4 shows that compared to workers who had no additional incentives to perform organisational tasks in Fuzhou factory the subjective evaluations and monetary prizes increased workers' production output and productivity in Fengcheng factory by 9% and 7% respectively. This improvement is statistically significant at 1% level.[26]

[24]One way to formally test the parallel trends assumption is to drop the *POST* from specification 2.10 and augment it with the experimental week variable and its interaction with the treatment indicator. To allow for weekly fluctuations and variations, we put every two weeks into a group. Regression results for both productivity and production output are shown in Table C3 and Table C4 in appendix, respectively. An alternative way is to include linear factory specific time trends among other regressors in specification 2.10. Our conclusion does not change. Results are available upon request.

[25]Figure 2.3 does not change qualitatively if we substitute productivity with worker's production output. However, outputs have more noises as workers were late for work or left work early sometimes (e.g. due to sickness).

[26]Robust standard errors clustered at the individual level are reported in brackets below the estimates.

FIGURE 2.3: Worker Mean Productivity



Notes: The vertical dashed line indicates the last week before our treatment was introduced in Fengcheng.

Table 2.4: The Treatment Effect on Worker's Performance

	Log(Output)				Log(Productivity (output per hour))			
	Jun-Sep	Jun	Jun	Jun	Jun-Sep	Jun	Jun	Jun
	Jun-Sep	vs. Jul	vs. Aug	vs. Sep	-Sep	vs. Jul	vs. Aug	vs. Sep
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fengcheng	-0.426***	0.299***	0.043	0.069**	-0.286***	0.208***	-0.094***	-0.055**
	(0.101)	(0.009)	(0.062)	(0.030)	(0.043)	(0.008)	(0.033)	(0.023)
Post	-0.038**	-0.056***	-0.086***	-0.018	0.027	0.003	0.031**	0.022
	(0.018)	(0.017)	(0.022)	(0.021)	(0.018)	(0.015)	(0.015)	(0.020)
Fengcheng*Post	0.091***	0.099***	0.105***	0.079***	0.066***	0.044***	0.060***	0.082***
	(0.017)	(0.016)	(0.019)	(0.022)	(0.017)	(0.014)	(0.017)	(0.020)
Observations	5,655	2,770	2,647	2,712	5,655	2,770	2,647	2,712
Clusters	57	57	57	57	57	57	57	57
R^2	0.528	0.670	0.483	0.610	0.780	0.811	0.761	0.791
Controls	YES	YES	YES	YES	YES	YES	YES	YES

Notes: The unit of observation is worker i. The dependent variables in Columns 1-4 and Columns 5-8 are the log of worker's daily output and the log of worker's productivity, respectively. Columns 1 and 5 show the results for the full sample includes observations from June 7th until September 30th while Columns 2-4 and 6-8 compare the observations from the pre-treatment period (June) to each post-treatment month separately. Productivity is a measure of the output per hour. Worker fixed effects, factory-line fixed effects, week fixed effects, day of the week fixed effects (e.g. Monday), and an indicator variable for whether the worker is recorded sick or if there is an organisational error are included in all regressions. Robust standard errors clustered at the worker level are reported in brackets below the estimates. *** Significant at 1% level, ** significant at 5% level, * significant at 10% level.

All remaining workers in our sample had more than three months of experience on the job. Thus, learning by doing should not be significant during our experimental period. Nevertheless, as the incentive scheme for the forewomen was new and the forewomen exerted more efforts in organising the production line, the workers may still need to learn how to work efficiently on a better-organised line. On the other hand, foreman-worker ties might strengthen as time passes when the forewoman invests more time in organising the production line. As a result, this may develop team cohesion and further increase workers' productivity. To test this, we divide the post-treatment period into three months. We can then estimate whether the treatment effects during the first, second, and third post-treatment month vary. Columns (6), (7), and (8) show the results regarding productivity for July, August, and September, respectively. The estimates are indeed increasing significantly over time. The subjective evaluations and monetary prizes for forewomen increased worker's productivity significantly by 4.4% in July and the figure further increased to 8.2% in September. Thus, we conclude that a high-powered incentive scheme regarding foreman's organisational performance had a significant impact in increasing worker's productivity.

Furthermore, Column (5) indicates that the treatment failed to further increase worker's output in September. The tumble was due to the fact that a large number of defective products were returned from the market. Workers participated in our experiment were responsible for unpacking these products for remaking. This task was not incentivised monetarily. Hence, workers who spent some time on this unpaid job were displeased, and their outputs were diminished. However, this is not reflected in workers' productivity figures as shown in Column (8). This is because the management requested the forewomen to record the time workers worked on this task, which enables us to deduct the time the workers spent on this task when we calculate their productivity.[27]

By design, we can observe the real scores the managers gave to each forewoman regarding their organisational performance, although these scores are unobservable to both the managers and forewomen. Hence, we can further add the subjective evaluation scores into the DiD specification 2.10 to test whether subjective evaluation scores are correlated with workers' productivity. Because the evaluations took place on a weekly basis, the sample is now aggregated to the week level. Some individual-level controls are no longer available. We only control for the forewoman and week fixed effects here. Table 2.5 shows that the coefficient of *FACTORY* * *SCORES* * *POST* in Column (1) is indeed positive and statistically significant at 5% level, where *SCORES* indicates the logged evaluation scores of the forewomen. It indicates that a one percent increase in a forewoman's evaluation score is associated with a 0.24% increase in the (week) average productivity of workers during the treatment period. This positive association can be found in all three treatment months as shown in Columns (2)-(4). It reached its peak at 0.4% in August, and the coefficient becomes statistically significant at 10% level in September.

[27] Recording the time workers spent on this particualr task was not too intricate for the forewoman because, in general, workers who were assigned to the unpacking job did not return to their regular job on the same day.

Table 2.5: Subjective Evaluation Scores of Forewomen and Workers' Productivity

	Productivity (output per hour)			
	Jun-Sep	Jun vs. Jul	Jun vs. Aug	Jun vs. Sep
	(1)	(2)	(3)	(4)
Fengcheng	1.065*	0.784	0.846	0.975
	(0.537)	(0.511)	(0.589)	(0.713)
Scores	0.053	0.054	0.064	0.063
	(0.063)	(0.066)	(0.065)	(0.068)
Post	0.450	0.521	0.411	0.470
	(0.344)	(0.426)	(0.359)	(0.387)
Fengcheng*Post	-1.436**	-1.444**	-2.366**	-1.565*
	(0.523)	(0.555)	(0.908)	(0.814)
Post*Scores	-0.077	-0.095	-0.070	-0.083
	(0.061)	(0.075)	(0.064)	(0.069)
Fengcheng*Scores	-0.185*	-0.147	-0.157	-0.175
	(0.088)	(0.084)	(0.096)	(0.116)
Fengcheng*Post*Scores	0.247**	0.245**	0.400**	0.273*
	(0.086)	(0.092)	(0.150)	(0.133)
Observations	174	76	74	72
R^2	0.878	0.900	0.893	0.833
Experimental Week FE	YES	YES	YES	YES
Forewoman FE	YES	YES	YES	YES

Notes: The unit of observation is forewoman i per week. The dependent variables are the log of forewoman's productivity. Scores indicates the logged evaluation scores of the forewomen. Column 1 shows the results for the full sample includes observations from June 7th until September 30th while Columns 2-4 compare the observations from the pre-treatment period (June) to each post-treatment month separately. Forewoman fixed effects and week fixed effects are included in all regressions. Robust standard errors clustered at the forewoman level are reported in brackets below the estimates. *** Significant at 1% level, ** significant at 5% level, * significant at 10% level.

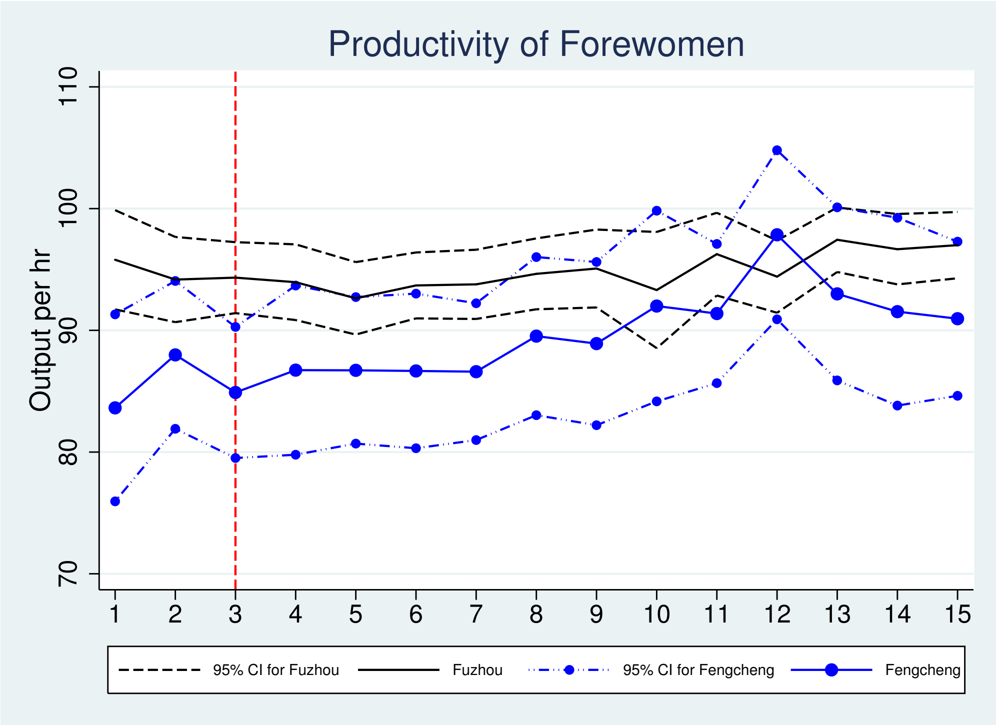
2.4.3 Performance of Forewomen

Similarly, we analyse how the subjective evaluations and monetary prizes affect forewomen's production performance as we did with the workers above. This is a formal test of the prediction we derived from the theoretical model. First, we test whether forewoman's productivity in the treated factory changed after the introduction of the treatment. Figure 2.4 shows the mean of forewomen's weekly productivity, averaged across all forewomen, in each experimental week. Forewomen's productivity started to increase gradually in the treated factory after week 7 but descended in the last three weeks, while forewomen's productivity in the controlled factory is exceptionally consistent throughout the experiment.

We also observe a significant difference in forewomen's productivity between the treated factory and the controlled factory since the status quo. This can be mainly driven by the individual-specific heterogeneity as we only have 7 forewomen in the controlled factory and 6 in the other. Nevertheless, there is no reason to expect that the average employee performance must be identical across factories since the two factories are independent of each other. Since the parallel trends assumption holds our difference-in-difference estimations are valid.[28]

[28]Formal tests are used to confirm the parallel trends assumption. Regression results for forewomen's productivity and production output are shown in Table C5 and Table C6 in appendix, respectively.

Figure 2.4: Forewomen Mean Productivity



Notes: The vertical dashed line indicates the last week before our treatment was introduced in Fengcheng.

One explanation for the convergence later in the experiment is that we exclude a forewoman, who was replaced by a line supervisor, from the sample after week 9. Therefore, the aggregated observations are not informative. If we plot Figure 2.4 and exclude this forewoman who was replaced in the mid of the experiment, the convergence is moderated considerably.

Columns (5-8) of Table 2.6 confirms this result by regressing log of forewoman's productivity on the regressors of specification 2.10. The coefficients for the variable of interest (Fengcheng*Post) in Columns (5-8) are positive but statistically insignificant. On the other hand, Columns (1-4) show that the treatment leads to an increase in forewomen's production outputs. In particular, the subjective evaluations and monetary prizes statistically significantly increased forewoman's output by 8% overall, while the coefficients are extremely large during the first two months as shown in Columns (2) and (3). It increased forewoman's output by around 10% in July and 9.8% in August, and the estimates are statistically significant at the 1% level and 5% level respectively. However, the coefficient for September is relatively small (by 5.8%) and statistically insignificant. Like the workers, forewomen were sometimes assigned to work on the products returned from the market during September which can somewhat explain why the statistical significance disappeared.[29]

[29] As the number of clusters in our case is small, we perform the wild cluster bootstrap as suggested in Cameron et al. (2008). With more than 200 replications, the results remain unchanged qualitatively.

Table 2.6: The Treatment Effect on Forewoman's Production Performance

	Log(Output)				Log(Productivity (output per hour))			
	Jun-Sep Jun-Sep (1)	Jun vs. Jul (2)	Jun vs. Aug (3)	Jun vs. Sep (4)	Jun-Sep -Sep (5)	Jun vs. Jul (6)	Jun vs. Aug (7)	Jun vs. Sep (8)
Fengcheng	0.180***	0.168***	0.194***	0.160***	0.258***	0.252***	0.259***	0.238***
	(0.024)	(0.014)	(0.023)	(0.023)	(0.017)	(0.012)	(0.012)	(0.014)
Post	-0.056*	-0.088**	-0.122***	-0.028	0.027	-0.003	0.030	0.031
	(0.031)	(0.039)	(0.039)	(0.032)	(0.026)	(0.030)	(0.027)	(0.028)
Fengcheng*Post	0.082**	0.100***	0.098**	0.058	0.034	0.023	0.042	0.038
	(0.031)	(0.028)	(0.040)	(0.043)	(0.021)	(0.023)	(0.023)	(0.025)
Observations	1,312	644	621	621	1,312	644	621	621
Clusters	13	13	13	13	13	13	13	13
R^2	0.350	0.491	0.314	0.497	0.845	0.873	0.832	0.840
Controls	YES	YES	YES	YES	YES	YES	YES	YES

Notes: The unit of observation is forewoman i. The dependent variables in Columns 1-4 and Columns 5-8 are the log of forewoman's daily output and the log of forewoman's productivity, respectively. Columns 1 and 5 show the results for the full sample includes observations from June 7th until September 30th while Columns 2-4 and 6-8 compare the observations from the pre-treatment period (June) to each post-treatment month separately. Productivity is a measure of the output per hour. Forewoman fixed effects, factory-line fixed effects, week fixed effects, day of the week fixed effects (e.g. Monday), and an indicator variable for whether the forewoman is recorded sick or if there is an organisational error are included in all regressions. Robust standard errors clustered at the forewoman level are reported in brackets below the estimates. *** Significant at 1% level, ** significant at 5% level, * significant at 10% level.

Overall, the evidence indicates that strengthening incentives in the organisational dimension encouraged forewomen to exert more effort in organising teams as workers became more productive which affirms the efficiency improvements among team members. On the other hand, the production outputs of forewomen also increased after the introduction of subjective evaluations and monetary prizes. Both changes are in line with the predictions listed in Proposition 1 that the introduction of a new bonus scheme increases worker's production efforts, and forewomen's organisational inputs and production efforts. As a result, their outputs increase.

2.5 Additional Evidence

2.5.1 Team Efficiency and Forewoman's Production Performance

The underlying assumption of our model is that there is a positive spillover effect among employees (including the forewoman) on the production line. For instance, in our case, one of the assessed organisational criteria is to maintain the efficiency of the production process, whereas forewomen have to invest effort and time in allocating raw materials effectively for workers to use. This effort increases not only workers' production performance but also the forewomen's because when the resources on the production line are systematically organised the forewoman is unlikely to be interrupted by the workers regarding this issue frequently. Therefore, a well-organised forewoman is able to work on her own rhythm and maintain productive. To check for this, we use the average production performance of workers, excluding the forewoman, and test its association with forewoman's production performance by estimating the following specification:

$$log(Y)_{i,l,f,t} = \beta log(\overline{Y})_{-i,l,f,t} + \eta' Z_{i,l,f,t} + \rho I_i + u_{i,l,f,t}, \tag{2.11}$$

where $log(Y)_{i,l,f,t}$ denotes the logged production-related outcomes (either output or productivity) of forewoman i from production line l in factory f and day t. $\overline{Y}_{-i,l,f,t}$ is the logged average performance of other coworkers (excluding the line forewoman) $-i$ from the same production line l in factory f and day t and all other variables are as previously defined.

Table 2.7: The Spillover Effect of Team Efficiency

	Log(Output)	Log(Productivity (output per hour))		
	(1)	(2)	(3)	(4)
Log(Ave. production of coworkers)	0.962***			
	(0.043)			
Log(Ave. productivity of coworkers)		0.589***	0.602***	
		(0.088)	(0.088)	
Log(Ave. experience of coworkers)			-0.054	
			(0.032)	
Assigned to a new line				-0.017
				(0.012)
Observations	1,305	1,305	1,305	3,276
Clusters	13	13	13	66
R^2	0.741	0.888	0.844	0.801
Controls	YES	YES	YES	YES
Sample	Only Foremen	Only Foremen	Only Foremen	Workers & Foremen

Notes: The unit of observation is forewoman i in Columns 1-3 while Column 4 uses the sample includes both forewomen and workers. The dependent variables in Columns 1 is the log of individual daily output while the dependent variables of Columns 2-4 are the log of individual productivity. Productivity is a measure of the output per hour. Individual fixed effects, factory-line fixed effects, week fixed effects, day of the week fixed effects (e.g. Monday), and an indicator variable for whether the forewoman is recorded sick or if there is an organisational error are included in all regressions. Robust standard errors clustered at the individual level are reported in brackets below the estimates. *** Significant at 1% level, ** significant at 5% level, * significant at 10% level.

Columns (1) and (2) of Table 2.7 reports the estimates from specification 2.11 for the subsample of forewomen who are observed throughout the experiment. In line with our assumption we find a positive association between forewoman's production performance and her coworkers', and the coefficients are statistically significant at 1% level. The spillover effect of team efficiency is extremely large: a 10% increase in a forewoman's coworkers' average production (productivity) will lead to a 10% (5%) increase in her own production (productivity).

While the productivity of workers can be linked to the team efficiency spillover, it can also be dependent on the experience of workers in a fabricating setting like ours. Experienced workers may have developed advanced techniques to perform the task, which they can share with their teammates. Hence any sorting into teams based on individuals' experience or productivity might lead to overestimating the positive spillover effect of team efficiency proxied by worker's productivity. To provide evidence on this, Column (3) tests the relationship between forewoman's productivity and her coworkers' seniority. Experience is defined as the number of days the worker worked in the factory. The association is statistically indistinguishable from zero. Furthermore, to test whether the management's decision to assign a worker to a production line depends on the worker's productivity, we use the sample includes both forewomen and workers and adjust specification 2.11 by regressing individual productivity (including both forewomen and workers) on a dummy variable for individuals who were relocated to another production line during our experiment. In line with the management's statement, the estimates reported in Columns (4) provide no evidence that the management's decision of relocating workers is associated with worker's productivity.[30]

2.5.2 Forewomen's Trade-off

Taken together, our findings in Section 2.4.3 and Section 2.5.1 indicate that incentives which encourage forewomen to undertake the organisational task indeed improve the production efficiency on the assembly line, for instance by rearranging the raw materials so that workers have easier access to the resources. While workers become more productive, the spillover effect of team efficiency also increases forewomen's production performance. The introduction of subjective evaluations and monetary prizes to forewomen regarding their organisational performance increased both workers' and forewomen's output by 9% and 8%, respectively. However, we do find evidence that workers' productivity had been increased by roughly 7%, the effects on forewomen's productivity cannot be economically and statistically distinguished from zero. A quick answer to this is the sample size. There are only 13 forewomen work in these factories comparing to 57 workers. Thus, the standard errors reported in Table 2.6 may be not informative.

Another way to explain an increase in the daily output but not output per hour is that the number of hours worked has also increased. This is consistent with the results reported in Table 2.8 which presents regressions using specification 2.10 with the number of minutes worked on the job per day as the dependent variable.

The results illustrate that the introduced incentive scheme increases the time forewomen spent on the job by roughly 28 minutes per day during the post-treatment period, but it is not significantly different from zero. If we decompose the treatment effects into each post-treatment month, in Column (2), we find that the incentive scheme has a significantly positive impact on the time forewomen invested per day during July. Forewomen from Fengcheng factory spent an additional 51 minutes per day on the job after the introduction of the rank incentive and monetary prizes. This is consistent with the design of our monthly prizes which is precisely aimed to motivate the forewomen to spend one hour organising the team instead of packing products. It suggests that forewomen exert more effort in production by extending their hours

[30] Performing the wild cluster bootstrap does not change our results qualitatively.

Table 2.8: The Treatment Effect on Forewoman's Working Time

	Number of Minutes Worked in a Day			
	Jun-Sep	Jun vs. Jul	Jun vs. Aug	Jun vs. Sep
	(1)	(2)	(3)	(4)
Fengcheng	-56.179**	-58.518***	-50.622**	-53.720**
	(18.471)	(11.001)	(18.044)	(17.638)
Post	-50.008***	-49.066**	-60.461***	-36.427**
	(13.999)	(22.448)	(16.016)	(15.044)
Fengcheng*Post	27.716	50.698**	28.924	12.794
	(23.804)	(18.852)	(32.352)	(31.925)
Observations	1,312	644	621	621
Clusters	13	13	13	13
R^2	0.316	0.419	0.295	0.371
Controls	YES	YES	YES	YES

Notes: The unit of observation is forewoman i. The dependent variables in Columns 1-4 are the working time (number of minutes) a forewoman worked in a day. Column 1 shows the results for the full sample includes observations from June 7th until September 30th while Columns 2-4 compare the observations from the pre-treatment period (June) to each post-treatment month separately. Forewoman fixed effects, factory-line fixed effects, week fixed effects, day of the week fixed effects (e.g. Monday), and an indicator variable for whether the forewoman is recorded sick or if there is an organisational error are included in all regressions. Robust standard errors clustered at the forewoman level are reported in brackets below the estimates. *** Significant at 1% level, ** significant at 5% level, * significant at 10% level.

of work when they invest more time in coordinating with workers. Columns (3) and (4) suggest that the treatment effects fade away over time. This is not particularly surprising because, according to the management, after the introduction of the subjective evaluations and monetary prizes forewomen are more frequently engaging in organisational tasks. This helped them to develop different styles of leadership and further equipped them with a variety of organisational skills. With more organisational experience forewomen were able to organise the workers more efficiently. An organisational task which costs the forewoman half an hour in July might only take ten minutes in September.[31]

Therefore, these findings suggest that multitasking forewomen spent more time on the job when they were given a higher-powered incentive on the organisational dimension.[32] Nonetheless, the impact of subjectively evaluating forewomen's organisational behaviour and monetary prizes on their productivity remains ambiguous since we do not observe the precise time forewomen invest in either manufacturing task or organising task.

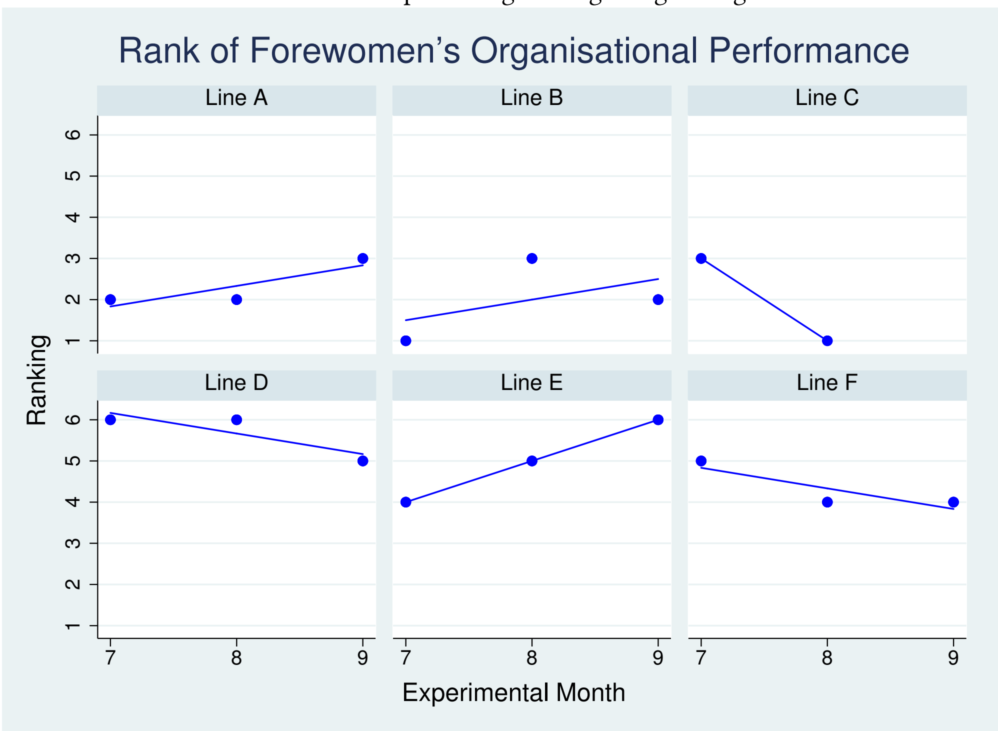
[31]The number of clusters in our case is small. Therefore, we perform the wild cluster bootstrap. Results do not change qualitatively.

[32]For the workers, they have also increased their time on the job during the treatment period as shown in Table C7 in the appendix. But, the effect is relatively small, the coefficients are more than 50% smaller than the ones for the forewomen.

2.5.3 Forewomen's Ranking

Finally, we use information on forewomen's rankings to provide evidence on whether the effect of subjective evaluations and monetary prizes are heterogeneous across forewomen's productivity distribution when the firm offers multiple piece-rate schemes (higher individual productivity yields a higher piece rate). In line with the complementarity assumption made in our production function that in Fengcheng during the treatment period the ranking of forewomen's organisational performance positively associate with their productivity ranking (correlation coefficient = 0.621 and statistically significant at 1% level), suggesting that a forewoman is more productive if she is ranked higher in leadership. It is also worth noting that the ranking distribution is heterogeneous as shown in Figure 2.5. Three forewomen (from Line A, Line B, and Line C) remained at the bottom of the ranking distribution, while other three forewomen competed for the top three rankings throughout the experiment.

FIGURE 2.5: Forewomen's Leadership Ranking in Fengcheng during the Treatment Period



Given the nature of the data, we are unable to identify how much time the forewoman have spent on organising teams in each day. However, it is still informative to compare the shift in a forewoman's weekly organisational ranking with the change in her weekly productivity because a forewoman who invests more time on organisation may increase her ranking but hurt productivity. The result suggests a negative correlation but relatively small and statistically insignificant (correlation coefficient = -0.077).

The evidence also suggests that in the treated factory Fengcheng during the treatment period, the forewoman's overall ranking regarding her organisational performance is positively correlated with the ranking of workers' average productivity but not perfect (coefficient = 0.651).

It indicates that the management evaluated the forewomen as we instructed. They considered workers' productivity as one of the criteria to determine forewoman's ranking but not entirely rely on it. This is crucial as the ranking incentive and monetary prizes provided to forewomen are entirely dependent on how the management subjectively evaluate the forewomen. It also implies that workers indeed benefit from a well-organised forewoman. However, the correlation between forewoman's organisational ranking and workers' productivity ranking is extremely weak and negative (coefficient = -0.047) when the evaluation is private information (in the controlled factory and during the pre-treatment period in the treated factory). This does not worry us from measuring the treatment effect because both forewomen and workers were not aware of this evaluation, it would not affect their performance. But, this result may raise a concern to the firm that the quality of assessment is poor when examiners are not monitored by either the firm or the examinees. Transparency is vital for the success of subjective evaluations.

Taken together, these results are consistent with our assumptions that forewoman's organisational inputs and production efforts are complements, and there exists a positive team efficiency spillover effect. Workers are more productive when the forewoman is a better organiser, and an efficient team results in a more productive forewoman.

2.6 Conclusions

Group leaders are usually responsible for organising the groups and contribute to the goal as a member. In the workplace, when one dimension of output is perfectly observable and quantifiable and the other is not, the classic multitasking theory applies. We address the issue by providing empirical evidence on the effects of multidimensional incentives and subjective evaluations. Through our interactions with managers at two Chinese factories who are struggling with this problem, we implement a natural field experiment to evaluate the impact of subjective evaluations and monetary prizes regarding foremen's organisational performance on two outcomes: worker's production performance, and foremen's production performance. When the former should be undoubtedly benefited from a better-organised group, the latter faces trade-offs. Specifically, by introducing the treatment in one factory while keeping the other factory as constant for three months, we provide a clean difference-in-difference test of the effects in a natural setting. This is important given the increased popularity of teams in industries such as manufacturing, academia, and healthcare.

Our results provide some meaningful insights: first, as we incent foremen to invest their time (not more than one hour per day) in organising the production process by introducing subjective evaluations and monetary prizes (which is equivalent to the hourly wage loss from manufacturing), their organisational inputs increase. As a result, the workers become more productive. We also find that a shift in a foreman's organisational ranking is negatively associated with the change in her productivity. A policy implication is that an incentive scheme which is based on the subjective evaluation results of group leaders' organisational activities is able to encourage leadership behaviour. But, there is a caveat to this: we do not find a positive correlation between foremen's organisational ranking and workers' productivity ranking when the subjective evaluation is not public information. This suggests that a subjective evaluation system may be ineffective when it is not under public scrutiny.

Second, we find that foremen's daily production output does not fall even when they spend more time on organising teams. This is because forewomen increased their working time on the job. We further show that there is a strong and positive spillover effect of team efficiency in the workplace, and foremen's organisational inputs and manufacturing efforts are complements. Nonetheless, we do not observe the specific time foremen invest in either production or organisation. The change in foremen's productivity is ambiguous.

Further, it is possible that peer pressure also plays a role in our setting as we do find a positive association between a forewoman's coworkers' productivity and her own productivity. The peer pressure which Chinese usually refer to as "Face" represents a person's reputation and feelings of prestige in the workplace. It may force a comparison of oneself versus her colleagues' performance. In our context, a forewoman whose productivity falls behind of line workers, or falls short of the local norm, may feel disgraced or dishonoured. This may propel them to increase efforts (e.g. Mas and Moretti, 2009). Absent such peer pressure, we might expect a relatively weak spillover effect of team efficiency, i.e. smaller λ, so that firms should be cautious when they are considering increasing the compensation for leaders' organisational activities to achieve a substantial effect like ours as this might lead to an adverse effect as shown in the model.

With all that being said, it provides a broad research agenda to learn about leadership and how multitasking leaders respond to multidimensional incentives.

Chapter 3

Dynamic Incentive Effects of Assignment Mechanisms: Experimental Evidence

Abstract. Optimal assignment and matching mechanisms have been the focus of exhaustive analysis. We focus on their dynamic effects, which have received less attention, in particular, in the empirical literature: anticipating that assignment is based on prior performance may affect prior performance. We test this hypothesis in a lab experiment. Participants first perform a task individually without monetary incentives; in a second stage, they are paired with another participant according to a pre-announced assignment policy. The assignment is based on first-stage performance and compensation is determined by average performance. Our results are consistent with theory: pairing the worst performing individuals with the best yields 20% lower first stage effort than random matching and does not induce truthful revelation of types, which undoes any policy that aims to reallocate types based on performance. Pairing the best with the best, however, yields only 5% higher first stage effort than random matching.[1] [2]

[1]This is a joint paper with my PhD supervisors Thomas Gall and Michael Vlassopoulos. For this paper, I initiated the research project, conducted the experiment, performed quantitative analyses, and edited the draft. On the other side, I greatly appreciate the effort Thomas and Michael had invested in this paper to improve its quality throughout the project. They had provided world-class supervision to ensure the success of the project, and the paper cannot be as polished as today without their effort. In particular, I would like to thank Thomas for his contribution in developing the theoretical model and his financial support for this project.

[2]This paper greatly benefited from comments and suggestions received from seminar participants at Southampton, Marburg, the London Experimental Workshop, the 12th Workshop Matching in Practice, the 2017 RES Annual Conference and the 2017 Economic Science Association European Meeting in WU Vienna.

3.1 Introduction

Individuals' payoffs depend on the economic environment they are placed in. For example, peers' attributes can affect an individual's payoff in the workplace or the classroom and spouse's attributes will affect one's payoff in marriage. However, the attributes that determine payoffs are likely to be the consequence of prior choices made, in full anticipation of the later assignment to other people, peers, tasks, or jobs. For instance, expectations about the future assignment into colleges or firms may well provide powerful incentives for accumulating human capital. This raises the interesting question: does the manner of how individuals are assigned to each other or to tasks, jobs, schools etc. affect their prior choices, such as, earlier stage investments and performance?

In the workplace, there is a wide range of methods used in practice to assign workers to tasks and to each other. For instance, some firms assign workers into teams that are heterogeneous in ability, by partnering strong performers with weaker ones, in order to facilitate learning or to provide role models that lead to productivity gains (Hamilton et al., 2003). This practice may limit an individual's desire to exert effort at an earlier stage, i.e., there is an equity-efficiency trade-off. Conversely, if the best performers are assigned to better partners this will provide additional incentives for effort at an earlier stage. Depending on the degree of production complementarity (Franco et al., 2011) and the strength of incentives (Bandiera et al., 2013), this pattern can also be the outcome when workers are allowed to choose their own teammates since workers will tend to match positively assortatively in ability. Of course, team formation may also be left to chance, for instance, if assignment is by sequence of arrival, follows a rotation system or is guided by alphabetic order of names (e.g. Bartel et al., 2014).

Also outside the workplace assignment mechanisms of individuals vary widely, sometimes as the result of an explicit policy, but often as the result of a decentralised market place. For example, in higher education there is a marked difference between the US and the UK where students self-select by academic ability into universities guided by detailed rankings, and continental Europe where students focus more on the city a university is located in. In secondary education, countries differ substantially in the degree to which they sort pupils by academic achievement, i.e. tracking (cf. Betts, 2011; Hanushek and Woessmann, 2011). Evidence on the marriage market suggests that mating is assortative in educational achievements (see e.g. Fernandez et al., 2005).

In all these examples, individuals who anticipate that their later assignments and outcomes will depend on their earlier stage choices will therefore respond to the assignment mechanisms used at later stages. This reasoning has been considered in the theoretical literature, examining e.g. investments taken before marriage or business partnerships (see e.g. Cole et al., 2001; Felli and Roberts, 2016; Peters and Siow, 2002; Bidner, 2010), providing some insights into the incentive effects of different matching mechanisms (e.g. Booth and Coles, 2010; Gall et al., 2006, 2012), and mechanism design (Hatfield et al., 2017). However, there has been no comparable interest in examining empirically the dynamic effects of different assignment mechanisms.[3] The aim of the current paper is to fill this void.

We design a real effort experiment with two work stages: in a first stage, participants perform a task individually and do not receive compensation. In the second stage, they are assigned to teams of size two to perform the task, based on their performance in the previous stage, and receive compensation that depends on the average performance of the team. However, the tasks worked on permit learning by doing, introducing a dynamic complementarity by increasing individual productivity in the later stage. Given the novelty of examining the resulting dynamic incentive effects we opt for a clean design and shut down static complementarities or substitutabilities, i.e. peer effects within teams in the second stage. Thus the design will allow some extrapolation of the results for the presence of positive or negative peer effects.

The experimental variation stems from varying the rule that matches participants in the second stage. We examine three salient forms of team assignment: random matching (RAM), matching participants randomly ignoring their performance in the first stage as a baseline treatment, positive assortative matching (PAM), in which the best performer is matched with the second best and so on, and negative assortative matching (NAM), in which the best performer is matched with the worst and so on. Besides the practical relevance, mechanism design theory would suggest that these assignment policies are interesting for another reason: PAM rewards higher first stage performance with a better match, and thus has a positive dynamic incentive effect on first stage effort. Under RAM, the second stage match is unrelated to first stage performance, thus shutting down the dynamic incentive effect. By contrast, under NAM higher first stage yields a partner with worse first stage performance, yielding a negative dynamic incentive effect. However, while PAM tends to induce investment behavior that is strictly monotone in productivity, thus revealing agents' types through their investment choices, this is not necessarily the case with NAM, as there is a tradeoff between one's own performance and the partner's performance in the second stage. Finally, as we are also interested in comparing the efficacy of team formation policies as an incentive device in relation to explicit monetary incentives, we implement a fourth treatment ($R\&I$), in which participants also receive an individual piece rate for their first stage performance and are randomly matched into teams.

We use a simple model to derive some theoretical pointers as to how outcomes in the individual work stage might differ across the treatments. Individuals who differ in their cost of effort, exert effort and invest through learning by doing, and then are assigned into teams of size two with payoffs increasing in their partner's effort. The results we obtain are largely consistent with the predictions. Intuitively, NAM leads to the lowest performance in the individual work stage, substantially lower than in the other treatments (20% reduction in mean performance

[3] There has been considerable interest in examining experimentally the static properties of assignment mechanisms, especially strategy-proofness, (see e.g. Pais and Pinter, 2008; Calsamigla et al., 2010; Braun et al., 2014).

compared to RAM, 30% compared to PAM). Perhaps surprisingly, RAM yields quantitatively similar performance outcomes as the positively incentivized treatments PAM and $R\&I$. These results point to an asymmetric effect: punishing effort appears to have a greater effect than rewarding effort, from a baseline of RAM. Somewhat more subtly, the evidence is also consistent with a more complex prediction of the model, namely, that NAM will not allow for truthful revelation of individual ability: since NAM precludes monotonicity of payoffs in one's choice, ex ante behavior will be characterised by an extent of bunching, i.e., different cost types will choose the same investment. Observed behavior under NAM is consistent with this prediction and suggests that measured performance before the assignment is not very informative about true productive and later performance. Interestingly, PAM and $R\&I$ are not statistically distinguishable from each other in terms of effort, which suggests that in our experiment the dynamic implicit incentive is as strong as the within period monetary incentive, offering a possible avenue for efficiency gains in production. In line with the theory, the difference between PAM (NAM) and RAM is more (less) pronounced for a task with less scope for learning-by-doing. Finally, we do not find any differences in performance in the team work stage across all treatments.

Our results have an interesting efficiency-equity implication that has not been highlighted before: a matching policy that aims to increase equity by equalizing average productivity and thus performance across groups (NAM) will impose a dynamic cost, by discouraging effort before the assignment into teams. Although such policies may well be motivated by efficiency considerations at the team work stage they will lead to a substantial shortfall of effort in earlier stages. Far more worrying is the fact that NAM will also not induce truthful revelation: observed effort before the assignment will generally not allow reliable conclusions regarding individuals' productivity, which, however, is needed to effectuate the desired assignment. Hence, achieved equality of the final allocation is likely to fall short of aspirations.

The remainder of this paper is organized as follows. The next section describes the design of our laboratory experiment in more detail. Section 3 presents the theoretical predictions to be tested in our study. Section 4 presents and discusses our experimental results. Section 5 concludes, and the Appendix includes proofs, additional tables, and the experimental instructions.

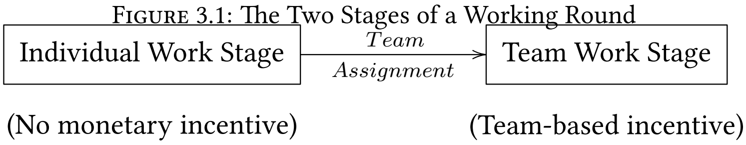
3.2 Experimental Design

To study the dynamic incentive effects of various assignment mechanisms we designed an experiment with the key feature that a participant's initial performance in a real effort task determines the partner that the participant is matched with in a later stage according to a known assignment rule. In this later stage compensation depends on the average performance of the pair. This provides an implicit incentive for a participant to exert effort in the earlier stage, in anticipation that this will lead to a more desirable match. This setup allows us to examine our main question of how early stage performance is affected by the mechanism, which varied across treatments, assigning participants to each other in a later stage.

3.2.1 The Stages of the Experiment

Specifically, participants in our experiment performed a real effort task in two stages. The first stage was an individual work stage, followed by a team work stage, with each stage lasting four minutes (see Figure 3.1). In addition, participants had two minutes to practice the task before the individual work stage. Subjects received live feedback about their score and the remaining time during practice. In addition, to help participants develop a better sense of their relative performance, they received feedback about their rank among all subjects based on the final scores achieved in practice.

Figure 3.1: The Two Stages of a Working Round



In the individual work stage subjects had four minutes to work on the task and their performance was rewarded implicitly or explicitly, depending on the treatment, which will be defined in the next subsection. At the end of the stage, we also elicited their belief about their performance relative to that of the other participants in their session in an incentive compatible way.[4] Then individuals were informed about their true rank as well as the maximum and minimum scores and were assigned to a partner for the team work stage. The assignment rule depends on the treatment and is explained in detail below. Subjects were shown their partner's rank and score before beginning the team work stage.

In the team work stage subjects had another four minutes to work on the task. Performance was rewarded by monetary payment, which was based on the average of own and the partner's performance. Similar feedback was given as in the previous stage. At the end of the stage subjects were informed about their partner's final score and the average of own and partner's score, which determined payment. Note that the team work stage does not entail joint production per se, as individuals perform the same task individually, but the compensation scheme induces a local public good.

This pattern was repeated three times (rounds). In each round participants performed a different task. After the last round subjects answered a brief questionnaire, eliciting subjects' preferences over risk and time, as well as altruism and competitiveness, and their socioeconomic characteristics (including gender, age, nationality, and native speaking language) and educational achievements (major fields of study on university, academic level, and years of study in university).

[4] They received £0.4 for correctly guessing in which quartile of the performance distribution they belong.

The sequence of events, the assignment mechanism in place, and the payment rules were communicated very clearly to the subjects at the very beginning of the experiment in both written and spoken form. In particular, considerable effort was made to ensure that the instructions informed subjects about the payment rules while not emphasising that the individual work stage is not directly incentivised, in the three treatments where this was the case.

3.2.2 Treatments

To test possible effects of different assignment mechanisms on the incentives to exert effort in the individual work stage prior to the assignment, our experiment involved four treatments that were implemented in a between-subject design. We focus on random matching as a benchmark and two polar assignment regimes, in the hope that this will allow extrapolation to a general class of assignment mechanisms. The first three treatments relied exclusively on the implicit incentives for first stage behaviour generated by different assignment mechanisms, while the fourth treatment added explicit monetary incentives in stage 1 to allow a quantitative comparison of the strength of explicit and implicit incentives.

Random matching (RAM)

Our benchmark treatment is random matching: each individual is assigned to any other individual with equal probability. This assignment regime reflects both actual randomised assignments as well as situations where the assignment is based on markers that are orthogonal to prior performance (such as the alphabetical order of surnames or the sequence of arrivals).

Positive assortative matching (PAM)

Positive assortative matching assigns individuals into groups based on their effort before the assignment. Specifically, the individual with the best performance is assigned to the individual with the second best performance, the third best individual to fourth best individual, and so on. This assignment mechanism rewards higher performance, corresponding to a higher effort, with a better partner. This pattern often endogenously arises in situations when individuals are allowed to choose their partners, since absent compensation payments individuals will tend to match positively assortatively (see e.g. Chen et al., 2015).

Negative assortative matching (NAM)

Similarly to PAM, negative assortative matching assigns individuals into groups based on their effort before the assignment, but the higher one's own performance the lower the performance of one's match. Specifically, the individual with the best performance is assigned to the individual with the worst performance, the individual with the second best performance to the one with the second worst performance, etc. This assignment mechanism provides low performers with high performing partners, thus generating balanced teams in terms of average performance of members. Real life examples include the formation of balanced teams, or the assignment of better employees to support weaker colleagues.

Random matching with monetary incentives ($R\&I$)

The three treatments presented above do not use any explicit monetary incentives to encourage effort before individuals are assigned to each other; only effort within teams is rewarded by monetary payments. In contrast, the final treatment ($R\&I$) rewards effort before assignment explicitly by a payment depending on performance. Assignment is by random matching, as under RAM.

3.2.3 Real Effort Tasks

To measure participants' effort and possible effects of differential assignment mechanisms, we used three computerised real effort tasks: the slider task (Gill and Prowse, 2012), the counting zeros task (Abeler et al., 2011) and the word encryption task (Erkal et al., 2011). The use of different real effort tasks was intended to provide subjects with a modicum of variety to maintain motivation through the 90 minutes duration of the experiment, and to account for the possibility that subjects' elasticity of effort provision to monetary incentives might differ between tasks (Araujo et al., 2016). All tasks are simple to understand, do not require preexisting knowledge and offer little gains from guessing. Hence, the performance, or score, achieved in a real effort task is a good measure of individual effort.

In the *Slider Task* (as proposed by Gill and Prowse, 2012) forty-eight sliders appear on screen, each with a range of integer values from 0 to 100, initially positioned at 0, see Figure B3 in the appendix. Subjects were tasked to use their mouse to position the slider at 50, which requires a certain degree of manipulation. Subjects' performance in the task, i.e. their "score", was given by the number of sliders successfully positioned at exactly 50 within the allotted time.

The *Grid Task* consists of counting the number of 0's in a 5×5 grid of randomly distributed 0's and 1's. Subjects were asked to enter the number on the screen, see Figure B4 in the appendix. If the number entered was correct, they continued to the next grid. The score in this task was the number of correctly counted grids within the allotted time. This task is similar to the task by Abeler et al. (2011), although they use 10×15 grids and impose no time limit.

In the *Word Encryption Task* subjects were shown combinations of three letters (words) and tasked to transcribe them into numbers using an encryption table mapping letters uniquely to numbers range from 0 to 100, see Figure B5 in the appendix. Once subjects entered the correct encryption, they were given a new random three letter combination to encode. The score in this task was the number of correctly encoded words. To limit training effects the encryption table was re-randomized before each stage (individual and team work), both changing the position of letters (not in alphabetical order) and the mapping from letters to numbers. Therefore subjects could not profit from memorising the encryption table nor the location of the keys. This task is similar to the task by Erkal et al. (2011), although they do not vary the encryption table. The double-randomisation of letters and numbers in the word encryption task was introduced by Benndorf et al. (2014) who find that it limits learning behavior when repeating the task.

We calibrated the difficulty of the tasks based on the results of pilot sessions, so that the average performance is approximately a score of 9 per two minutes for all tasks. Importantly, the three tasks differ in their scope for learning-by-doing through improving hand-eye coordination allowing us to test our prediction (iii) in Proposition 3.1 which can be found in the next section. In particular, for both the slider and the grid task previous studies have found that subjects improve performance over time (Georganas et al., 2015; Vranceanu et al., 2013), whereas the version of the encryption task we employ has been shown not to allow for significant improvement due to learning (Benndorf et al., 2014).

3.2.4 Payments

In each of the treatments RAM, PAM and NAM subjects were paid only in the team work stage and according to the average of the scores of the two partners. At the end of a session, for each subject one of the three tasks was randomly chosen and the subject's payment was her average team work stage score in that task multiplied by a piece rate of £0.4 per score point.

Treatment $R\&I$ additionally rewards individual performance in the individual work stage, given by the subject's score in that stage. In this treatment for each subject one of the three tasks and one of the individual and the team work stages is randomly chosen with equal probabilities and the subject's payment will be the subject's team, respectively individual score, in the selected task at the selected stage multiplied by a piece rate of £0.4 per score point.

3.2.5 Procedures

The experiment was conducted at the Social Sciences Experimental Lab (SSEL) of the University of Southampton, in spring of 2016. We ran three sessions of each of the four treatments described above (RAM, $R\&I$, PAM and NAM), for a total of 12 sessions. The order of treatments and the sequence of tasks within sessions was randomized, under the condition that each of the three tasks was the first one to be performed in a session exactly once for each treatment. Each session had 16 student subjects from various departments, with 192 participants in total (104 females and 88 males).[5] The subjects were recruited from the SSEL subject pool, using ORSEE (Greiner, 2004). The experimental instructions were provided to each subject in

[5]We invited 20 randomly selected subjects to each session. The first 16 subjects who showed up at the lab participated in the experiment. The other subjects received a show-up fee of £4 and were asked to leave the laboratory.

written form and were also read aloud to the subjects. Seating positions were randomised and seat numbers were given in the order of arrival. To ensure subject-experimenter anonymity actions and payments were linked to seat number only. After reading the instructions and before performing the task, subjects completed a quiz to ensure understanding of the rules and the assignment mechanism in their treatment. Each subject was paid a show-up fee of £4 and earned an average of a further £10 during the experiment. Subjects were paid privately in cash at the end of each session. The experiment was programmed in z-Tree (Fischbacher, 2007).

3.3 Predictions

To organise thoughts, consider the following model of investment and matching. An economy is populated by a continuum of agents, characterized by their types θ, and lasts for two stages. Suppose θ is distributed on an interval $[\underline{\theta}, \overline{\theta}]$ with $0 < \underline{\theta} < \overline{\theta}$. Suppose also that the distribution of productivity θ has full support.[6]

Production

In each stage t each agent can exert effort e_t to generate output y_t. Effort e_t comes at a utility cost c_t, however. To best reflect the experimental setup we impose the following formal assumptions on the cost function.

Assumption 1. *Stage* 1 *and* 2 *effort cost functions are given by:*

$$c_1(e_1) = \frac{e_1^2}{2\theta} \text{ and } c_2(e_1, e_2) = \frac{e_2^2}{2(\theta + \lambda e_1)},$$

where $\lambda > 0$. *They have the following properties:*

(i) $c_t(.)$ *does not depend on other individuals (no peer effects in production),*

(ii) $c_t(.)$ *strictly increases and is strictly convex in* e_t *(real effort),*

(iii) The marginal cost of effort strictly decreases in type θ *(heterogeneity),*

(iv) The marginal cost of stage 2 *effort* e_2 *strictly decreases in stage* 1 *effort* e_1 *(learning-by-doing).*

In the experiment, individual performance in a task only depends on the individual's own actions, in all stages. This will preclude peer effects, at least in performing the task, and is reflected in our first assumption. Since participants engage in a real effort task we impose

[6]Assuming a discrete set of agents or types will not change results for NAM, RAM and $R\&I$, in which one's equilibrium match is random. The assumption will preclude pure strategy equilibria under PAM. Using mixed strategies or introducing some noise in performance or the matching would ensure that an equilibrium exists and has similar properties as the one derived below. The intuition, namely, that dynamic incentives will boost stage 1 effort under PAM, reduce them under NAM, and be absent under random matching, will still be present.

increasing marginal cost, which will imply diminishing returns to effort. Heterogeneity of effort cost captures differences between individuals in their ability to perform the experimental tasks. Since performance in all experimental tasks can be enhanced by practice (even if it is only by familiarising oneself with the user interface), the cost of effort will decrease in past experience, i.e. there is learning-by-doing. The strength of the intertemporal learning spill-over is given by the parameter λ. The remainder of this section will use a fully parameterised cost function, but the results derived below are driven by the four qualitative assumptions given above, not by the functional form.

Finally, suppose that individual output is simply given by effort:[7]

$$y_t = e_t.$$

Assignment

After stage 1 individuals are assigned into teams of size 2, with attributes (e_1, θ) and (e'_1, θ'). The assignment takes one of the four forms described above: RAM, $R\&I$, PAM, or NAM. Team output in stage 2 is the sum of individual stage 2 output y_2 and y'_2:

$$Y_2 = y_2 + y'_2.$$

Payoffs

An individual's monetary payoff is given by a piece rate 1 per unit of average team output in stage 2, $Y_2 = y_2 + y'_2$. Hence, an individual's overall payoff is given by

$$(y_2 + y'_2)/2 - c_1(e_1) - c_2(e_1, e_2).$$

Under $R\&I$, the individual receives a piece rate w both per unit of stage 1 performance y_1 and stage 2 average team output $y_2/2$, yielding payoff:

$$w(y_2 + y'_2)/2 + wy_1 - c_1(e_1) - c_2(e_1, e_2).$$

In the experiment we set $w = 1$, but only one of the two stages was selected to determine the payment, each with probability $1/2$. Hence, if stage 1 and stage 2 output are of similar size, $R\&I$ will induce a similar aggregate wage bill as RAM.

Note that no peer effects in production implies that an individual's productivity does not depend on their match, but their payoff will. This also implies that, holding constant stage 1 effort e_1, aggregate output in stage 2 and surplus are independent of the assignment mechanism used. Hence, no peer effects in production implies that any differences in aggregate

[7] An equivalent formulation could specify output as a strictly increasing and concave function of effort and use a linear cost, for instance, and impose assumptions on the output function analogous to Assumption 1.

output and utility in stage 2 are entirely due to the dynamic incentives effects of the different assignment mechanisms.

Solution Concept

The type of assignment mechanisms may affect participants' stage 1 behaviour. Individual stage 2 payoff increases in the effort of their partner. If individuals anticipate that the quality of their partner in stage 2 depends on their own stage 1 performance (in the non-random assignments), stage 1 effort will be rewarded (or punished) with a better (worse) partner. Of course, stage 1 performance may or may not be informative of stage 2 behaviour, as both are part of equilibrium behaviour. To identify equilibrium behaviour we use therefore a sub-game perfect Nash equilibrium in individual effort choice e_1 and e_2 in the two-stage game played by individuals. We omit general properties, e.g., existence because this type of matching cum investment game with a continuum of players has been explored elsewhere, for instance in the work by Cole et al. (2001) and the premarital investment game by Peters and Siow (2002), both imposing PAM, and by Booth and Coles (2010) and Gall et al. (2012) who also allow for RAM, respectively, NAM.

Stage 2 **behavior**

Since stage 1 equilibrium behaviour will depend on anticipated stage 2 equilibrium outcomes, we use backward induction to derive an equilibrium and start with behavior in stage 2. In stage 2, individuals are assigned into teams and choose individual effort e_2 given their assignment and their stage 1 effort choice e_1. That is, an individual chooses effort e_2 to solve

$$\max_{e_2} \frac{e_2 + e_2'}{2} - \frac{e_2^2}{2(\theta + \lambda e_1)},$$

where e_2' denotes the effort of the individual's partner. Note that the cost of stage 1 effort is sunk. Hence, individual optimal stage 2 effort satisfies:

$$e_2^* = (\theta + \lambda e_1)/2,$$

and $e_2^* = w(\theta + \lambda e_1)/2$ under $R\&I$. Since this must be true for each individual in all teams, an individual's overall payoff from both stages under RAM, PAM and NAM is given by:

$$u(e_1, \theta, e_1', \theta') = \frac{\theta + \lambda e_1 + 2(\theta' + \lambda e_1')}{8} - \frac{e_1^2}{2\theta}.$$

This payoff clearly increases in the attributes e_1' and θ' of one's partner in stage 2. Under $R\&I$ an individual additionally obtains payoff $y_1 = we_1$, so that $u(.) = w^2 \frac{\theta+\lambda e_1+2(\theta'+\lambda e_1')}{8} + we_1 - \frac{e_1^2}{8\theta}$, which also increases in the stage 2 partner's characteristics. This implies the following fact.

Fact 1. An individual's payoff strictly increases in the quality θ' and stage 1 effort e_1' of their stage 2 partner.

This property relies on the stage 2 reward scheme, introducing a local public good, and learning-by-doing for the assertion on stage 1 efforts.

Stage 1 **behavior**

In stage 1 an individual chooses only effort e_1. This choice depends on the continuation payoff in stage 2 through two possible channels: higher stage 1 effort will reduce the cost of stage 2 efforts through learning-by-doing (the intertemporal spill-over channel), and it may affect the attributes e_1' and θ' of one's stage 2 partner through the assignment mechanism in place (dynamic incentive channel). Moreover, under $R\&I$, stage 1 effort choice will additionally depend on the reward for stage 1 performance w directly. That is, an individual chooses e_1 to solve

$$\max_{e_1} u(e_1, \theta, e_1', \theta'),$$

taking into account that own stage 1 effort may change attributes of one's match e_1' and θ'.

We start by examining the two random assignment mechanisms RAM and $R\&I$. Under both mechanisms e_1' and θ' do not depend on an agent's choice of e_1. That is, both RAM and $R\&I$ shut down the dynamic incentive channel, and $R\&I$ introduces direct piece rate incentives for stage 1 effort on top of the intermporal spill-over channel. The first order conditions become:

$$\frac{e_1^{RAM}}{\theta} = \frac{\lambda}{8} \text{ and } \frac{e_1^{R\&I}}{\theta} = w^2\frac{\lambda}{8} + w.$$

Under PAM and NAM both channels are present, but will have opposite signs. Under PAM e_1' increases in e_1, and θ' will also depend on e_1. The individual optimization problem becomes now:

$$\max_{e_1} \frac{\theta + \lambda e_1 + 2(\theta'(e_1) + \lambda e_1'(e_1))}{8} - \frac{e_1^2}{2\theta}. \tag{3.1}$$

If, as under RAM, e_1 increases in type θ, then higher e_1 also implies being matched to a higher type θ'; see the appendix for a proof that this is indeed the case. In this case $e_1^{PAM} > e_1^{RAM}$ because PAM will reward stage 1 effort through the dynamic incentive channel in addition to the positive intertemporal spill-over through learning-by-doing. See the appendix for the full derivation.

Under NAM the stage 1 effort of one's partner e_1' (weakly) decreases in own effort e_1. But this means that the two channels are in conflict: a higher stage 1 effort e_1 will be rewarded by lower effort cost in stage 2 (intertemporal spill-overs), but punished by receiving a partner with lower e_1' (dynamic incentives). Hence, strategies need not increase in type, and a semi-pooling equilibrium will result (see appendix for details): lower productivity participants will choose effort $e_1 = 0$, while higher productivity participants will choose $e^{NAM} = e^{RAM}$. This is because individuals at the top will be matched with a random partner among all individuals who set $e_1 = 0$, and increasing effort even marginally at the bottom would imply a discrete loss in the quality of stage 2 assignment.

To sum up, the Nash equilibrium for each assignment mechanism has the following properties (details are in the appendix):

Fact 2. Individual stage 1 effort e_1 in a Nash equilibrium depends on the assignment mechanism as follows:

- Under RAM effort is $e_1^{RAM} = \frac{\lambda\theta}{8}$.
- Under $R\&I$ effort is $e_1^{R\&I} = w^2 \frac{\lambda\theta}{8} + w\theta$.
- Under PAM effort is $e_1^{PAM} = \frac{3\lambda\theta}{16} + \sqrt{9\lambda^2 + 64}\frac{\theta}{16}$.
- Under NAM there is $\hat{\theta}$ such that $e_1^{NAM} = 0$ for agents with $\theta < \hat{\theta}$ and $e_1^{NAM} = \frac{\lambda\theta}{8}$ for agents with $\theta > \hat{\theta}$.

Using these expressions for equilibrium effort in stage 1 allows us to compare the different regimes in terms of observable outcomes, yielding testable predictions.

Proposition 3.1 (Predictions)**.** *Comparing equilibrium first stage effort levels under the different assignment mechanisms:*

(i) PAM and, if $\lambda < 8w/(1-w^2)$, $R\&I$ induce higher effort for all types than RAM, which in turn induces higher effort than NAM, and strictly so for some types.

(ii) PAM induces higher effort than $R\&I$ if the degree of learning-by-doing λ is sufficiently high and the difference increases in λ, i.e. there is $\bar{\lambda}(w) \geq 0$ such that $e_1^{PAM} > e_1^{R\&I}$ for all $\lambda > \bar{\lambda}(w)$, where $\bar{\lambda}(1/2) = 0$ and $\bar{\lambda}$ increases in w.

(iii) The percentage difference in effort between PAM ($R\&I$) and RAM decreases, but the percentage difference in effort between RAM and NAM increases in the degree of learning-by-doing λ.

Very intuitively, dynamic incentives under PAM boost stage 1 effort, while those under NAM reduce it, compared to RAM where no dynamic incentives are present, as one's stage 2 partner does not depend on stage 1 effort. Consequently our main result, the comparison between RAM, PAM, and NAM, relies on the assumption of learning-by-doing (NAM and RAM coincide if $\lambda = 0$) and no peer effects in production combined with splitting the surplus in a pair. Allowing for positive peer effects, that is, letting efforts be complements, would generate a game with increasing differences and increase equilibrium effort under PAM in stage 2 and thus also in stage 1, decrease both under NAM and not affect expected effort in the random protocols. The opposite will hold for negative peer effects. The comparison to $R\&I$, which uses monetary incentives in stage 1 as well as in stage 2 depends on the power of incentives. For $w > 1/2$ stage 1 effort is higher than RAM for all λ. This is because under RAM stage 1 effort is only rewarded through the intertemporal spill-over, while $R\&I$ rewards stage 1 effort directly through a piece rate. Comparing PAM to $R\&I$ generates some ambiguity: PAM induces higher effort than $R\&I$ for sufficiently high degrees of learning-by-doing; note, however, this is the case for all $\lambda > 0$, if $w = 1/2$. In the experiment, in treatment $R\&I$ only one of the two stages was chosen randomly for payment, using the same piece rate as in the other treatments for both stages.

The degree of learning-by-doing inherent in the task affects the comparison between the different mechanisms: outcomes under RAM become closer to outcomes under PAM and $R\&I$ and less close to those under NAM, and effort under PAM increases faster than under $R\&I$, as the degree of learning-by-doing increases.

Incentive Compatibility and Truthful Revelation

Since stage 2 assignments are based on stage 1 outcomes, a social planner wishing to alter the assignment of productive types into teams relies on the equilibrium choices of stage 1 effort to reveal individual types.[8] Truthful revelation of individual types is implicitly guaranteed under PAM, RAM and $R\&I$ since equilibrium stage 1 effort choices strictly increase in type θ_i.

Fact 2 states that this is not true for NAM, however, since it induces partial bunching (or pooling): all agents with types $\theta_i \leq \hat{\theta}$ will choose the same effort level 0 and higher types will choose the same effort as under RAM. This implies that effort choices are more dispersed under NAM than under RAM. The comparison of equilibrium efforts under PAM and $R\&I$ to those under NAM, respectively, RAM does not allow for a clear-cut characterisation (computations show that it depends on λ and the distribution of θ, and on whether the cost function is separable in θ and λ). Partial pooling under NAM also implies that stage 1 effort choices are not strictly monotone in type, and therefore stage 1 performance does not reveal types. Hence, using NAM conditional on stage 1 performance will not induce NAM conditional on true types θ_i.

Recalling that stage 2 effort choices strictly increase in type under all assignment mechanisms, this reasoning generates another prediction on the correlation of individuals' ranks in stage 1 and stage 2 performance.

Corollary 3.2 (Correlation)**.** *The correlation of individual ranks in equilibrium stage* 1 *and stage* 2 *performance is* 1 *under RAM, PAM and $R\&I$, and strictly less than* 1 *under NAM with random tie-breaking.*

[8] Note, however, that the predictions in Proposition 3.1 carry over to a setting where assignment is in terms of productivity types θ, in contrast to e.g. the ratchet effect (see Cooper et al., 1999; Charness et al., 2011).

3.4 Results

3.4.1 Sample

Summary statistics of participant characteristics are presented in Table B4. Using Chi-square, t-, and Mann-Whitney-U-tests (M-W test) Table B7 in the appendix shows that participants' characteristics are balanced across treatments with the exception of academic level and age (which are plausibly highly related).

3.4.2 Effort in The Individual Work Stage

We use the score achieved in each task as a measure of an individual's effort. Table B1 summarizes performance in the individual work stage for the whole sample and by each treatment and task separately.[9] The mean score across treatments and tasks was 23. While the mean scores in the slider and the grid tasks were very similar, the mean score in the word encryption task was significantly lower.[10]

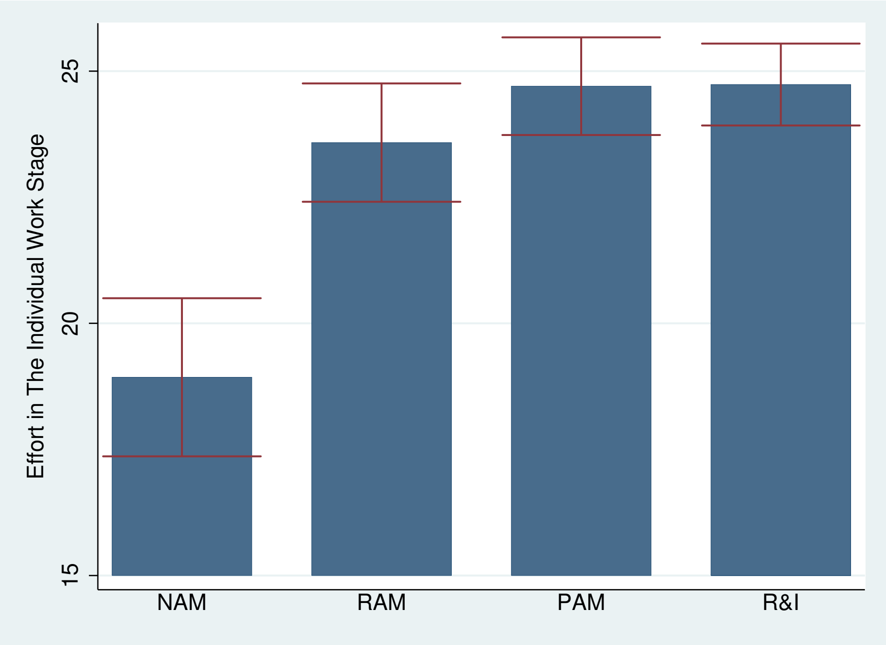
To address our main question of whether individual effort in stage 1 is affected by the assignment mechanism used in stage 2, Figure 3.2 shows the mean individual performance levels by treatment. In particular, the point estimate of the mean score under NAM is about 20% less than under RAM, while PAM and $R\&I$ are virtually indistinguishable and both induce 5% higher scores than RAM. However, mean performance under RAM, PAM and $R\&I$ are not statistically different at conventional levels. Observed patterns in terms of point estimates, though not necessarily in terms of significance of the differences, are indeed in line with our expectations stemming from the theoretical model in Section 3.3, reflecting that the dynamic incentive effect will have a positive impact on first stage effort under PAM, be absent under RAM and $R\&I$ and negative under NAM. That RAM generates higher first stage effort than NAM is consistent with a presence of the intertemporal spill-over. Adding direct monetary rewards under $R\&I$ was expected to yield higher first stage performance than RAM.

As mentioned above, the three different tasks may have differed in terms of learning-by-doing or sensitivity to explicit incentives. To assess possible differences in outcomes, Figure 3.3 shows individual stage performance across treatments for each task and yields a nuanced picture: while the differences of outcomes between treatments are similar across tasks, the magnitudes of the differences vary considerably. While performance under RAM comes close to the one under PAM and $R\&I$ for both the grid and the slider task, this is not the case for the word encryption task in which performance under NAM is about 13% less than under RAM, while

[9]Following the practice by Gill and Prowse (2012), we leave out of the analysis one participant (in the treatment $R\&I$) who scored 0 in all three stages of the slider task. Our qualitative results do not depend on this sample selection and the quantitative results would change only marginally.

[10]Both the paired t-test (p-values < 0.001) and the Wilcoxon signed rank sum test (p-values < 0.001) reject equality of mean score in the word encryption and the other two tasks. Comparing mean scores in the slider and grid tasks the Wilcoxon test indicates a significant difference (p-value = 0.048), but not the t-test (p-value = 0.361).

FIGURE 3.2: Effort in The Individual Work Stage



Notes: The top end of the bars indicates the mean effort in the individual work stage, and the line segments represent the 95% confidence intervals.

PAM and $R\&I$ both induce $8-9\%$ higher performance than RAM (Table B1). Since the word encryption task offers less opportunity for learning-by-doing than the other tasks, this result is consistent with our prediction (iii) in Proposition 3.1.[11]

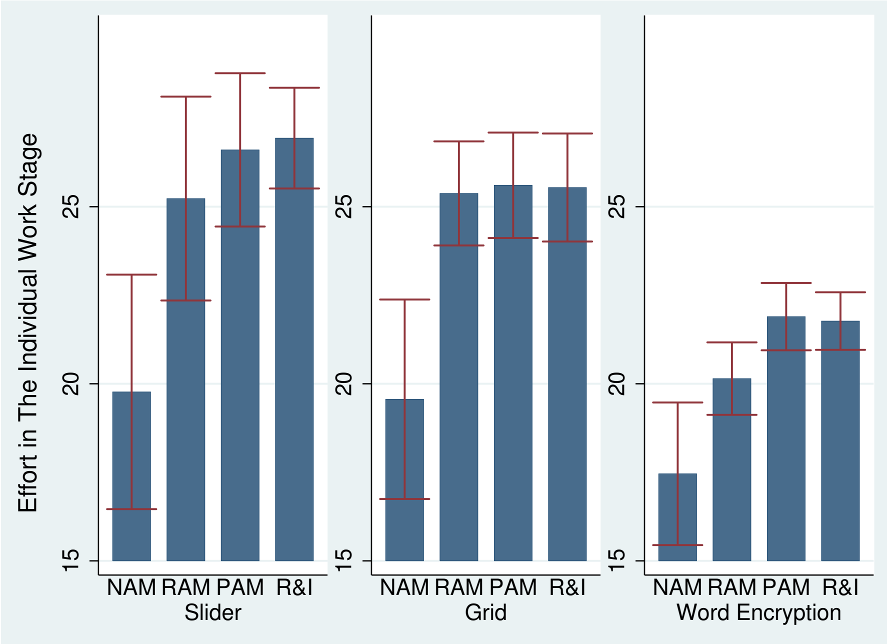
Table B2 reports the results of the tests for possible differences between treatments. Summing up over the scores achieved in all tasks, both nonparametric test (M-W) and t-test yield significant differences between NAM and the other three treatments. On the other hand, we find no statistically significant differences between Treatments RAM, PAM and $R\&I$.[12] Table B2 also contains the results for each task separately, confirming the picture in Figure 3.3: for the word encryption task performance under RAM was significantly smaller than under each of PAM and $R\&I$, whereas this was not the case for the other two tasks.

To be able to control for variation at the task and round level, as well as individual characteristics, we complement the previous analysis with OLS regressions displayed in Table B3. Column (1) presents the results of a regression with only the treatment dummies as independent variables, column (2) adds task and round fixed effects to capture unobservable variation across

[11]One possible way to test for differences in learning-by-doing across tasks is to examine the relative performance improvement between the two work stages in treatment $R\&I$, since effort is incentivized with monetary payments in both stages. Doing this we find that the slider task has the largest improvement (mean = 8.4%), followed by the grid task (mean=3.5%) and the word encryption task (mean = 1.1%). A pairwise test rejects equality between the slider and the word encryption task (a paired t-test has a p-value of 0.024 and a Wilcoxon test 0.063), but fails to reject equality between the grid and the word encryption task (a paired t-test has p-value 0.313 and a Wilcoxon test 0.272).

[12]These results remain the same when using a weighted average of the scores instead of simply adding up. Detailed re-weighting methods and results are available upon request.

FIGURE 3.3: Effort in The Individual Work Stage by Task



Notes: The top end of the bars indicates the mean effort in the individual work stage, and the line segments represent the 95% confidence intervals.

tasks and rounds. Column (3) adds preference indicators, constructed from subjects' answers to the questions asked during the experiment. The preference indicators capture subjects' accuracy of beliefs about relative performance, competitiveness, altruism, time discounting and risk attitudes (see Appendix B for details on the construction of these variables and Table B4 for descriptive statistics.). Columns (4) to (6) add demographic covariates (academic level, gender, and finally controls for nationality and an economics-related degree subject) to account for possible differences in the sample composition, although the selection into treatments was fairly balanced on observables.[13] The coefficients for the different treatments remain relatively stable across the different specifications.

Overall, the regression analysis confirms the results above, indicating that NAM was associated to a decrease in score of $4.6 - 4.7$ (about 20%) relative to RAM, while PAM and $R\&I$ were associated to an increase in score of $1.0 - 1.3$ (about 5%) each. The drop in performance under NAM is statistically significant for all specifications, as is the increase under $R\&I$, although the significance level drops to 10% as we saturate the model with controls. The performance increase under PAM is only significant in some specifications, however, and only at the 10% level. Finally, the coefficients of PAM and $R\&I$ are statistically indistinguishable in all specifications.[14]

[13]Notice that individual age, years of study, and native speaking language are not included in the regressions as they are collinear with academic level and nationality, respectively.

[14]The results are very similar when using an individual random effects estimation approach (which may be warranted as individuals are not independent within each session). The same is true when including observations from the one subject dropped because of a failure to score at all in the slider task. Results are also qualitatively unchanged

This analysis does not include a proxy for individual ability at a task. One could, however, use individual performance at the team work stage to proxy for individual ability in a given task. It is, however, plausible that the team work stage effort choice could be affected by the treatment, i.e., by the team composition, which would generate an endogeneity problem. Nevertheless, adding team work performance into our specifications does not change our conclusions above.

3.4.3 Truthful Revelation

As noticed above in Section 3.3, individual work stage effort choice may be strategic and not reflect individuals' true productivities. The theoretical model predicts that individual work stage effort choices are indeed strictly monotone in productivity type under PAM. This means that individuals with better performance in the individual stage can be expected to perform better in the team work stage, so that inequality of individual performance in teams should be low within teams, but high across teams. For NAM the model predicts that a positive measure of the population will choose the same effort level (zero) in the individual work stage. That is, in mechanism design terms, NAM will induce bunching and is not incentive compatible: individual stage performance is not necessarily informative about true productivity. Strategic behavior in the individual work stage would imply, of course, that performance rankings of individuals will differ between the individual and the team work stage under NAM, but not in the rest of the treatments. The data offer support for this prediction: Table B5 shows the rank correlation of individual and team work stage performance, which is significantly (at conventional levels) lower under NAM than under the remaining treatments.

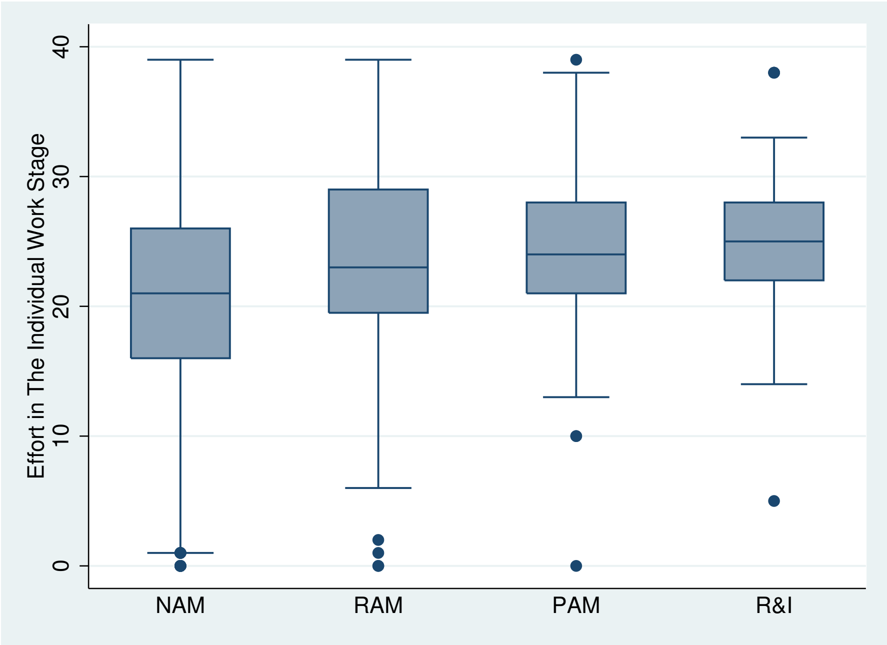
The theoretical prediction of bunching for part of the individuals under NAM would also imply that first stage performance is more dispersed under NAM than under RAM (recall that the comparison between NAM and either PAM or $R\&I$ is ambiguous). Figure 3.4 shows box-plots of the distribution of individual stage efforts across treatments, indicating that moving across treatments from NAM to $R\&I$ the average effort increases while its dispersion decreases, as does its standard deviation (see the fourth column in Table B1).[15]

Note that subjects in both NAM and PAM who perform at the average level are paired with a counterpart who shares a similar performance. For instance, in our case, in NAM subjects who ranked 7 and 8 are teamed with 10 and 9, respectively. In PAM, subjects who ranked 7 is paired with the one who ranked 8, and ranks No.9 and No.10 are paired into teams. Figure 3.5 indicates that the performance of individuals from these comparably matched groups in NAM and PAM are significantly lower in NAM than in PAM. Moreover, in line with the theoretical prediction that individuals with extreme ranks who are paired in NAM, i.e. the best and the worst performing ones, exert even less effort than the average performing ones from the same

when using the logarithm of the dependent variable, although the treatment effect size increases. Regression results are available upon request.

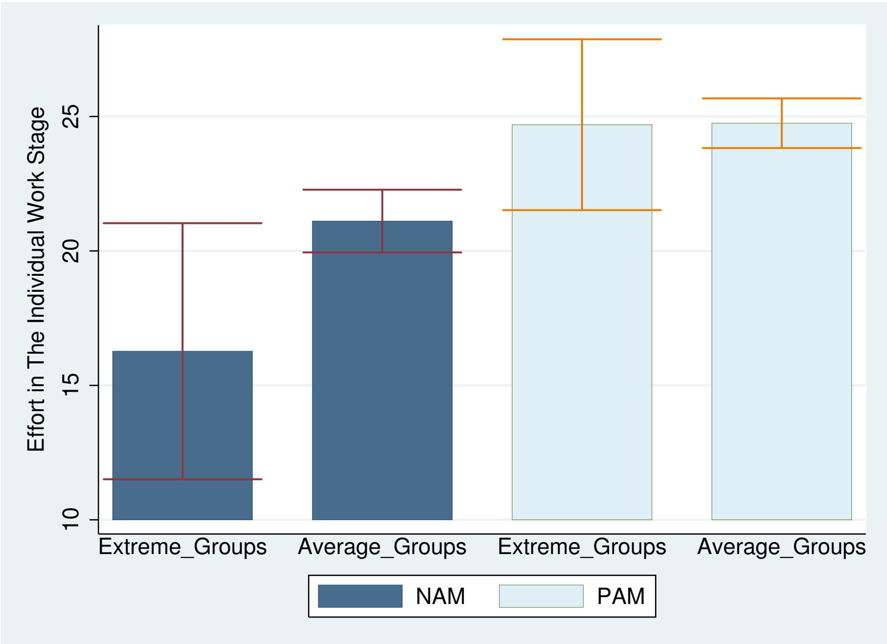
[15]Pairwise F-tests of equality of variance across treatments reject equality in all cases.

FIGURE 3.4: Dispersion of Effort in The Individual Work Stage



treatment as they bunch at zeros. However, this is not the case in PAM, meaning that our sample precisely offers us a symmetric distribution even under an incentive mechanism.

FIGURE 3.5: Effort in The Individual Work Stage by Groups

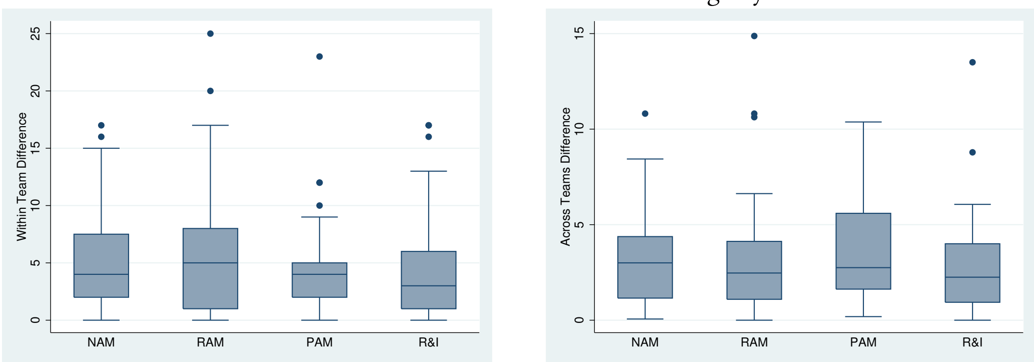


Notes: Extreme groups denote subjects who are ranked No.01, No.02, No.15, and No.16, while average groups denote subjects who are ranked No.07, No.08, No.09, and No.10.

All this strongly suggests that NAM does not truthfully implement a negative assortative matching of true productivity types, but will involve some randomness. This is relevant since one possible motivation for the use of NAM may be a concern for inequality. For instance, PAM will induce very little inequality *within* teams in terms of individual attributes (i.e., past performance) but substantial inequality *across* teams. The converse will be the case for NAM: there will be considerable inequality within teams (matching the best to the worst performers, etc.), but very little inequality across teams. If past performance reflects individual ability this difference in within and across teams inequality of past performance should be mirrored by the performance in the team work stage.

A failure to truthfully implement negative assortative matching in true ability types implies that inequality in *team stage performance* is lower within teams and higher across teams than inequality in individual stage performance. This implication seems consistent with the data from the experiment. Figure 3.6 shows the performance difference both within and across teams in the different treatments. PAM is clearly distinguishable from the other treatments and shows both low within team and high across teams inequality of actual team work stage performance, while NAM does not appear to differ substantially from the two treatments that match randomly.[16]

FIGURE 3.6: Team Difference in The Team Work Stage by Treatment



A. Within Team Difference

B. Across Teams Difference

3.4.4 Subgroup Analysis

There are some plausible reasons why certain subgroups of individuals might be expected to react more to our treatments than others. First, our theoretical model predicts a degree of strategic behavior that is informed by individuals' knowledge about the type distribution. Hence, we would expect that behavior of participants who are well informed about their relative performance among all subjects conforms more closely to the theory. Second, since participants in the experiment received feedback on their relative performance, highly competitive individuals could be expected to be motivated intrinsically and to respond less to extrinsic incentives.

Therefore, we explore possible differences of treatment effects between subgroups by splitting the sample along two dimensions: first with respect to how accurately they were able to predict

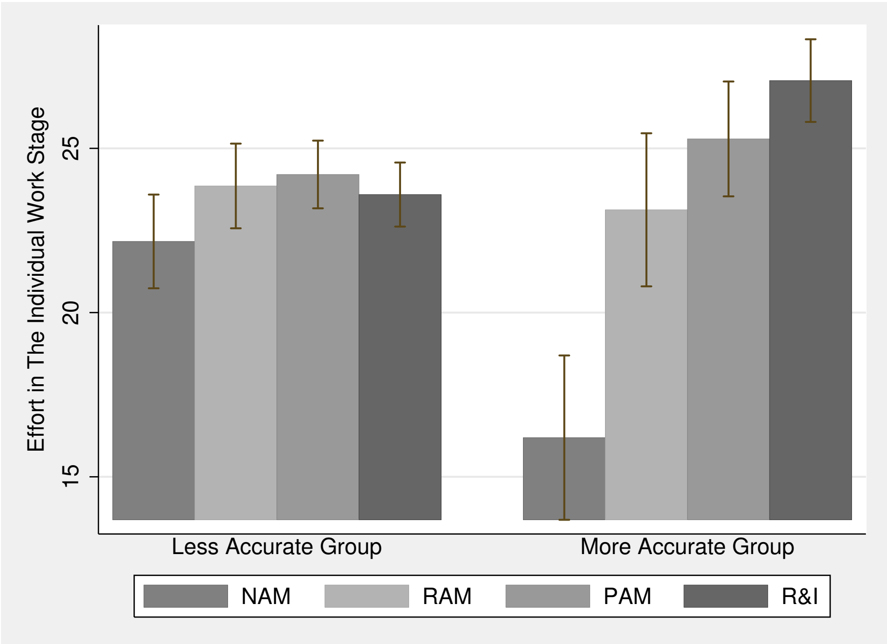
[16]The differences between PAM and the other treatments are statistically significant in some but not all of the comparisons (results are available upon request).

their relative position in the performance distribution and second with respect to intrinsic ability or motivation as measured by actual performance in the team work stage.

3.4.4.1 Splitting The Sample Based on Accuracy of Beliefs about Relative Performance

Recall that at the end of the individual work stage of each round we elicited participants' beliefs about their relative performance. More specifically, participants predicted the quartile, in which they believed their performance to lie, and received a reward if they were correct. Indeed, a sizable fraction of participants (41% to 47%) accurately predicted their quartile, about 20% underestimated it, and 33%-39% overestimated it. Overall, 42.6% of participants were able to accurately predict their relative performance in their session in at least two of the three rounds, with 11.8% correctly predicting their quartile in all three rounds. 20.9% did not predict correctly in any of the rounds.[17]

FIGURE 3.7: Subgroup Analysis by Accuracy of Beliefs about Relative Performance



Notes: The top end of the bars indicates the mean effort in the individual work stage, and the line segments represent the 95% confidence intervals.

[17]See Table B8 in the appendix for the demographic composition of the groups. 56.3% of the participants who had correct beliefs about their relative performance are male, and there seems to be a higher incidence of them under NAM, although the difference is only statistically significant at 10% (regression results are available upon request). Furthermore, more than half of the participants who had correct beliefs about their relative performance for at least two of the three rounds are non-competitive (84.1%) and risk-averse (81.6%). This result makes sense because an overconfident subject is less likely to develop accurate (rational) expectations about their relative standings so as to follow the dominant strategy, and overconfident subjects are more likely to compete, and more risk-averse subjects are less likely to compete since the tournament involves more risk than the piece rates (Bartling et al., 2009). Therefore, competition-lovers and risk-seekers are most unlikely to follow the dominant strategies in a real-effort experiment like ours.

Since strategic behavior requires individual expectations of relative performance to be reasonably accurate (at least under NAM and PAM) one would expect that the treatment differences are more pronounced among the group of participants that well predicted their own relative performance. To explore this possibility, we split the sample into two similarly sized groups: one group ($n = 110$) that predicted their ranking correctly in at most one round, and the other group ($n = 82$) that correctly predicted their ranking in at least two rounds. Figure 3.7 shows mean performance by treatment separately for each of the two groups. For the subject group that predicted their relative performance more accurately the pattern of treatment effects mirrors closely the theoretical predictions: performance was very low under NAM, while PAM and $R\&I$ were significantly higher than RAM. On the other hand, the performance of the subjects who predicted less accurately does not differ much across treatments, except for NAM, which yields slightly lower performance, albeit significantly higher than under NAM for the other group.[18]

These observations carry over to a regression analysis similar to the one for the whole sample reported above. The results, shown in Table B9 in the appendix, indicate that for the more accurate group treatment effects are much greater in magnitude than for the other group (-7% vs. -28% for NAM, 1.5% vs. 9% for PAM and 0% versus 17.5% for $R\&I$), and the increase under $R\&I$ is statistically significant throughout.[19]

3.4.4.2 Splitting The Sample Based on Team Work Stage Performance

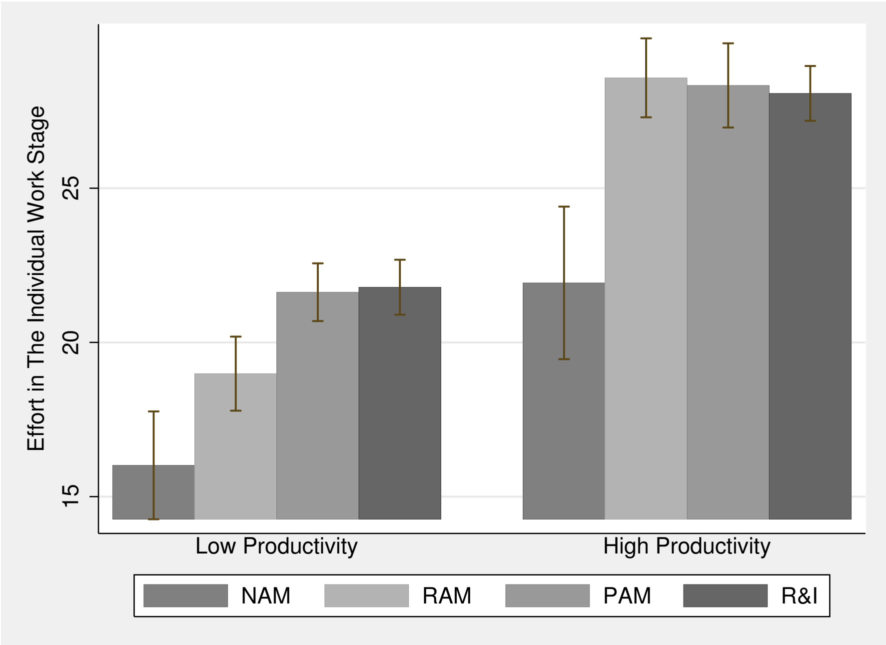
Our second subgroup analysis addresses a possible concern of any real effort experiment: subjects might exert substantial effort regardless of the experimental treatment because of intrinsic motivation. This might be due to a desire to perform well, either because participants enjoy working on the task, or because they feel challenged. Alternatively, subjects may feel a moral obligation to exert effort knowing that they will receive a compensation for participating in the experiment. Finally, the feedback on their relative performance that participants received may already offer substantial non-monetary incentive for status-concerned individuals to exert effort. That is, some participants may already be exerting effort close to their capacity and thus make it difficult to detect variations across treatments. On the other hand, participants that lack such intrinsic motivation will perform well below their full potential, making it more likely to detect treatment effects.

To examine this possibility, we split the sample into two similarly sized groups on the basis of their performance in the team work stage. Subjects with less than the median performance

[18] In particular, both t-test and Mann-Whitney U test indicate statistically significant differences in subjects' performances between the two subgroups within treatment NAM (mean difference 5.974, p-value of t-test $<$ 0.001, and p-value of Mann-Whitney U test < 0.001) and $R\&I$ (mean difference -3.470, p-value of t-test < 0.001, and p-value of Mann-Whitney U test 0.005). On the other hand, the differences are statistically insignificant for RAM (mean difference 0.726, p-value of t-test 0.556, and p-value of Mann-Whitney U test 0.666) and PAM (mean difference -1.083, p-value of t-test 0.272, and p-value of Mann-Whitney U test 0.118).

[19] The results are robust to clustering standard errors at the individual level, with the exception that the coefficients for treatment NAM become statistically insignificant for the less accurate group.

FIGURE 3.8: Subgroup Analysis by Team Stage Performance



Notes: The top end of the bars indicates the mean effort in the individual work stage, and the line segments represent the 95% confidence intervals.

form a low productivity group (n=101) and those with higher than the median performance form a high productivity group (n=91).[20] Recall that the team work stage is the only stage where individual effort is explicitly incentivised (with a team piece rate). Hence, individual performance in the team work stage is arguably a reasonable proxy for individuals' intrinsic motivation, in particular since average team work stage performance did not vary across treatments (see Section 3.4.5).

Figure 3.8 depicts the performance for each group by treatment and shows marked differences between the two groups in all treatments. What we find is that for the low productivity group observed treatment effects closely mirror the theoretical predictions. On the other hand, for the high productivity group, only treatment NAM is distinguishable from other treatments. These observations can be further seen in a regression analysis by productivity group. The regression results in Table B11 in the appendix indicate that for the low productivity group (columns 1 and 2) there are statistically significant differences across treatments except for the difference between PAM and $R\&I$. For the high productivity group (columns 3 and 4), however, only treatment NAM shows the expected drop in performance compared to the other treatments, while performances in PAM and $R\&I$ are not statistically different from those in RAM.

[20]See Table B10 in the appendix for details on the composition of the two groups. The main difference appears to be that the high productivity group has a higher share of UK nationals and, reassuringly, of participants who prefer competitive settings (regression results available upon request).

3.4.5 Effort in The Team Work Stage

While the main focus of this paper lies on individual performance before the assignment into teams, it is of interest to examine whether different assignment mechanisms affected participants' performance once they were assigned a partner. Table B6 presents the average individual performance in the team work stage by treatment. Mean performance is very similar across treatments, and this remains true when examining the three different tasks separately.[21] Thus, the treatment in form of assignment mechanism has no effect on the individual performance after the assignment.

3.4.5.1 Peer Effects

Though not at the focus of our analysis, later stage performance could have been affected by the assignments into teams through the presence of peer effects, which have received considerable attention in the literature (e.g. Eisenkopf, 2010; Falk and Ichino, 2006; Mas and Moretti, 2009, among many others). In our experiment, subjects received information on their own and their partner's absolute and relative performance, but no live feedback was given. This setup allows for a possible peer effect through the knowledge of being paired with a better or worse performing peer. The possible effect is ambiguous: on one hand a better peer may make free-riding more attractive, but on the other hand reciprocity or inequity aversion may induce higher effort anticipating higher effort of one's peer.

To examine possible peer effects we estimate an OLS regression of individual performance in the team work stage on two dummy variables indicating whether an individual's partner had performed better, respectively, worse than that individual in the individual stage. We constrain our sample to treatments RAM and $R\&I$, because both treatments used the same assignment mechanism (random matching), thus excluding selection bias. In this subsample, 45.3% of participants were paired with a better partner, 45.6% were paired with a worse partner, and 9.1% had the same individual stage performance as their partner. The results (see columns (1) and (2) of Table B12 in the appendix) suggest peer effects were modest: being assigned a better (worse) partner (compared to one's own individual work stage performance) is associated negatively (positively) with own performance in the team work stage, but not significantly so. The coefficients for the two dummies have the expected (i.e. different) signs and differ significantly from each other. The negative sign of the coefficient for a better partner is consistent with a free-riding effect. This is corroborated by regressing individual team work performance on the continuous individual work stage performance of one's partner (instead of the dummy indicating a better or worse partner), see columns (3) and (4) of Table B12 in the appendix.[22]

[21] Results of statistical tests are available upon request. In addition, an F-test confirms that there are no statistically significant differences in the standard deviations (column SD in Table B6) across treatments in the team work stage.

[22] When standard errors are clustered at the individual level the coefficients for the two dummies become weakly significant and the implications remain the same for the continuous individual work stage performance of one's partner. Results are available upon request.

3.5 Conclusion

Does the manner of how individuals are assigned to each other affect prior effort choice? Our results from a real effort task experiment strongly suggest that the answer is in the affirmative. Specifically, we find that subjects substantially reduce prior effort under an assignment rule that matches high performers with low performers relative to a scenario where individuals are randomly matched. The evidence is consistent with strategic behavior in the early stage because NAM is not incentive compatible so that early stage performance does not reveal true productivity types – i.e. the usual disclaimer applies: past performance is not indicative of future results.

This finding confirms expectations of an equity-efficiency trade-off: an assignment rule that yields teams similar in average prior performance of their members comes at the cost of reducing effort ex ante, much as Ramsay logic would suggest. While our results give a possible reason for caution when using matching on attributes based on prior choice in experiments, perhaps more important in practice are adverse implications for policies e.g. in school admission or personnel organisation that are designed to implement heterogeneity in terms of markers correlated with prior performance, such as race. Further research on this matter would appear highly desirable in order to inform policy.

Assignment policies that match better performing individuals with better partners or explicitly reward early stage effort with monetary payments tend to outperform random matching in terms of early stage performance, but the effect is relatively small. More interesting is perhaps the finding that effort choices under both explicit (monetary) and implicit (assignment) incentives are statistically indistinguishable. That is, using some form of positive assortative matching can replace costly monetary payment in earlier stages (perhaps reminiscent of the use of low or unpaid internships before workers are promoted to full-paid positions).

The analysis in this paper is a first pass at bringing an investment and matching framework to the lab. There are several directions in which the analysis could be extended. For instance, the assignment could be made endogenous, allowing participants to submit preference rankings over peers and then employing tried and tested matching algorithms. Moreover, while our results suggest the presence of learning-by-doing, effort in our experiment was not explicitly designed as an investment. Explicitly incorporating investment before assignment could be a potentially valuable approach to model educational policies in the lab. Moreover, many effort and investment decisions are taken in a team environment, potentially subject to peer effects. Hence, a repeated team formation and effort choice setup may shed some more light on productive processes.

Despite the effects found for performance before team formation we do not find significant differences in effort across treatments at the team work stage. This is not entirely unexpected as the real effort task performed in teams is independent across members and the payoff additively linear in individual performance. Hence, there are no peer effects by design and one's peer matters only through group incentives. Corresponding to the latter we do find some evidence for mild free-riding at the team work stage. Of course, a potentially fruitful direction for further research could be to incorporate complementarities at the team work stage, for instance by tweaking the payments to reflect increasing or decreasing differences of joint production in individual output. In particular, when weaker individuals profit more from stronger teammates than stronger individuals (decreasing differences) a tension will arise between static optimisation (favoring NAM-like policies) and dynamic considerations in terms of crowding out earlier stage effort (favoring PAM-like policies).

Chapter 4

The Role of Gender and Compensation Scheme in Managers' Mentoring: Evidence from British Workplaces

Abstract. This paper studies both non-monetary and monetary determinants of mentoring relationships between managers and employees in British firms by using data from the Workplace Employment Relations Survey. In particular, I focus on the role of a manager's gender and the use of managerial incentive schemes. Past literature suggests a significant association between a manager's gender and mentoring behaviour. However, using longitudinal data this paper finds that the significant relationship disappears once firm fixed effects are included. The results also show a positive but weak association between managerial incentive schemes and managers' mentoring behaviour. Widespread mentorships are more likely to be found in firms where managers' payments are linked to organisational profits.[1]

[1]I greatly benefited from comments and suggestions received from seminar participants at Southampton.

4.1 Introduction

Mentoring has become increasingly popular as a personnel management device at workplaces around the world. Mentorships not only help mentees to develop skills and progress in their careers,[2] but also benefit firms from mitigating costly mistakes as employees turn to their mentors for advice. More importantly, employees who experience mentoring relationships are more often satisfied with their jobs, which turns the workplace into a more positive work environment and reduces employee turnover.[3] However, although 93 per cent of the 11,000 small and medium-sized enterprises (SMEs) surveyed in 17 countries across the world acknowledge that mentoring helps employees to succeed, only 28 per cent of them make use of recognisable mentors (according to the 2013 Sage Business Index).

There is an extensive body of study on the determinants of mentoring focusing on the function of mentoring and the characteristics of mentors, or mentees, or both such as their similarities.[4] In particular, gender has received tremendous attention. Female managers are expected to provide more mentoring than their male counterpart based on the assumption that females are more caring and nurturing.[5] However, once hired, it is not necessary that female managers take up on the role of mentoring or manage differently from males (Wajcman, 1996). One possible reason is that most of these studies, which use survey data that is collected at one point in time or within a single organisation, fail to account for other important factors that determine mentoring. This exclusion is significant as mentoring can be organised costly or complimentarily that depending on the condition of the employers. The benefits may also vary substantially with the nature of the firm, its employees and its culture. For example, managers might be more demotivated to provide mentoring support and employees might be more disinclined to seek mentoring support if they perceive that the employer does not value such behaviours. Theory suggests that this can give rise to sorting: firms who believe in the value of mentoring attract employees with the same beliefs (Van den Steen, 2005). The resulting alignment of beliefs of different employees could develop into the corporate culture and persist for a long time (Van den Steen, 2010). Thus, this paper contributes by re-examining the association between managers' gender and the provision of mentoring while taking the firm's time-invariant corporate culture into account by controlling for the firm fixed effects.

My analysis is based on 1,733 British firms across 12 industries observed in the Workplace Employee Relations Surveys (WERS). Data from 600 of these firms are collected repeatedly in 2004 and 2011. Mentoring regarding personal career development is captured by a survey question, asking employees to what degree they agree with the statement that managers at

[2] See Allen et al. (2004); Ragins and Cotton (1999) for reviews.

[3] See Chao (1997); Haggard et al. (2011); Karatepe (2013); Payne and Huffman (2005); Ragins et al. (2000) for details.

[4] See Ghosh (2014) for a review.

[5] See Eagly and Crowley (1986), Eisenberg and Fabes (1991), Burleson et al. (1996), Markiewicz et al. (2000), Fletcher and Ragins (2008), and Powell (2010) for a review.

the workplace encouraged people to develop their skills.[6] The specific structure of this dataset allows me to control for numerous sources of heterogeneity that threaten the identification. For instance, time-varying observed firm heterogeneity (including company's preference to fill vacancies with internal employees or external applicants, the quality of product or service, company's market share, and the degree of competition in the market), time-invariant unobserved firm heterogeneity (e.g. firm's culture), industry-specific factors, formal status-specific factors, region-specific factors, and economy-wide trends (as the value of mentoring is recognised through time) that may additionally affect the evolution of managers' mentoring at firms.

I find that in the absence of firm fixed effects, the higher the ratio of women in management positions in a firm the greater share of employees, especially females, affirmed that their managers encouraged them to develop skills. However, the statistical significance of this association vanishes after I include the firm fixed effects. The reasoning has been considered in the theoretical literature, sorting e.g. employees based on the employer's beliefs (see e.g. Van den Steen, 2005). A firm who believes in the value of mentoring attracts employees with the same beliefs and eliminates the employees who have different views. As a result, the alignment of beliefs of different employees could further develop into the corporate culture and persist for a long time (e.g. Van den Steen, 2010). Hence, the time-invariant beliefs of firms are important factors to determine employees' behaviours in the workplace, as argued in Rob and Zemsky (2002).

Alternatively, I propose a potential candidate - managerial incentive schemes. For instance, group performance pay schemes may provide incentives for managers to help workers (e.g. Drago and Turnbull, 1988; Lazear, 1989; Itoh, 1991; Kandel and Lazear, 1992; Siemsen et al., 2007). If group performance is strongly rewarded (such as profit-related pay which ensures that the improvement of subordinates increases the realised incentives of the manager), managers will spend less time on individual tasks and devote more effort to help others. On the other hand, by selecting less powerful group incentives and strong individual incentives, an organisation will induce employees to work harder on individual tasks that do not require teamwork. I find that the occurrence of career mentoring is positively associated with managerial incentives, but only weakly so. In particular, if managers were offered a group performance (organisational profit) related pay, the occurrence of mentoring in this firm is 0.04 (0.06) percentage points higher compared to companies did not offer the performance (profit) related payment schemes. However, these are only preliminary results which cannot be interpreted as causal effects. It encourages future studies to go beyond easily observable manager characteristics and further explore the impact of managerial incentives on the mentoring relationships in the workplace.

[6] In general, mentoring is defined as a relationship between a mentor (a more experienced and qualified senior) and a mentee (a less experienced individual) that is formed to facilitate career and personal development (Kram, 1985). A more detailed explanation of this variable can be found in Section 4.2.1.

The findings for managers' gender is in contrast to a literature that finds female mentors are perceived to provide more support, motivation, and inspirations (e.g. Allen and Eby, 2004; Burxe et al., 1993; Sosik and Godshalk, 2000) with survey data collected at one point in time or in a single organisation. The omission of organisation fixed effects leads to a failure of accounting for important determinants of the employee-manager mentoring relationship in the workplace, such as organisational culture (Ragins and Cotton, 1999; Ragins et al., 2000). This paper examines this relationship with a large sample of firms across various industries and over time. The time-invariant determinants (e.g. organisational context) will be captured by the firm fixed effects. My finding is consistent with recent evidence from a very specialised labour market: Bednar and Gicheva (2014) show that an athletic director's gender appears to be uncorrelated with the decision on hiring and retaining female coaches. However, this earlier finding may not be immediately generalizable to a wider context, whereas my study covers a wide range of industries.

While the manager's gender seems to not be the driver, an alternative determinant may be the remuneration structure of managers. This paper contributes to the growing empirical evidence on the support that a manager gives to workers, and their relation to managerial compensation. For instance, Bandiera et al. (2009) finds that an exogenous change from a fixed wage to a bonus scheme for managers changes their way of assigning workers to different jobs, Devaro and Kurtulus (2010) shows a positive relationship between group performance-related pay and the delegation of worker authority by using the 1998 British WERS, and managerial ownership plays an important role to determine employee compensation (Cronqvist et al., 2009). In contrast, my study provides novel evidence supporting that managers are more likely to encourage workers to develop their skills when their own payment is related to group performance or corporate profit.

The remainder of the paper is structured as follows: Section 2 describes the data sources and presents summary statistics of key variables. Section 3 outlines the econometric approach and Section 4 reports and discusses the results. Section 5 concludes.

4.2 Data

I obtain data from the survey records of the Workplace Employee Relations Surveys, collected in two consecutive waves 2004 and 2011 (WERS 2004-2011), respectively.[7] This is a nationally representative stratified random sample covering British workplaces with at least five to nine

[7]Notice that panel data can only be constructed between two consecutive survey waves. The workplaces participated in the survey are followed up only once, and it is intended that there is no overlap between the 1998-2004 panel sample and the 2004-2011 panel sample by the historic design.

employees across a dozen industries.[8][9] A major feature of the WERS survey is the use of two entirely separate samples. Either the first wave (2004) or the second wave (2011) can be used as a fresh cross-section sample. Alternatively, one can construct a sample with a panel structure by combining these two waves.

This study focuses on the management questionnaire (MQ) and the employee questionnaire (EQ). The respondents of MQ were the senior managers dealing with personnel, staff or employment relations at the workplace according to Chaplin et al. (2005). At the end of each survey, the manager was asked for permission to distribute the employee questionnaires among employees at the workplace. If the manager agreed, 25 employees were randomly selected by the interviewer from a list of all employees at the workplace provided by the manager or all employees were interviewed if the workplace had no more than 25 employees. Since the employee respondents were either randomly selected by the interviewer or all being surveyed, the problem of bias in the selection of employees can be eliminated.

A total of 1,733 workplaces (pooled cross-section sample) returned employee questionnaires from the 2,295 workplaces, and 600 workplaces provided employee questionnaires in both 2004 and 2011 (panel sample). Notice that only the management questionnaires can be used to construct a panel structure, because the employee questionnaires do not necessarily cover the same employees for both years.

4.2.1 Career Mentoring

The EQ provides an information in the form of a 5 item Likert scale which asks the employees "to what extent do you agree or disagree about the managers at the workplace encouraging people to develop their skills". This is a natural definition of mentoring according to Kram (1985), who describes mentoring as "an experienced mentor helps a less experienced person to enhance career by developing specific skills and knowledge". An act of encouragement from the manager can be seen as one of the supports given to the employees. In line with classic mentoring theory and most available evidence (Dreher and Cox, 1996; Fagenson, 1989; Kram, 1985; Scandura, 1992; Wallace, 2001), this measure of mentoring is positively related to job satisfaction in our data (see Appendix B).

Since the skills managers encouraged employees to develop at the workplace are in most cases career/job related, "career mentoring" is used as a shorthand to describe this encouragement throughout the rest of the paper. A respondent can choose among five options in which 1 = strongly disagree, 2 = disagree, 3 = neither agree nor disagree, 4 = agree, and 5 = strongly

[8]The twelve industries covered in the survey are: Manufacturing, Electricity, gas and water, Construction, Wholesale and retail, Hotels and restaurants, Transport and communication, Financial services, Other business services, Public administration, Education, Health and Other community services. Workplaces are geographically dispersed all over the United Kingdom, including North, Yorkshire and Humberside, East Midlands, East Anglia, South East, South West, West Midlands, North West, Wales and Scotland.

[9]Workplaces were selected as a stratified random sample from the issued sample of workplaces to provide representative results for the population of workplaces in existence at the time of the survey.

agree. As shown in Table 4.1, around 58% respondents asserted that their managers at the workplace encouraged people to develop their skills in both 2004 and 2011.[10] Standard errors are computed by using employee sampling weights.

Table 4.1: Summary Statistics of the 5-point Scale Item on Mentoring

	Proportion					
	Strongly Disagree	Disagree	Neither Disagree Nor Agree	Agree	Strongly Agree	Observations
Sample	(1)	(2)	(3)	(4)	(5)	(6)
2004	0.052	0.126	0.237	0.431	0.154	21,769
	(0.002)	(0.004)	(0.004)	(0.005)	(0.004)	
2011	0.048	0.117	0.249	0.430	0.156	21,536
	(0.003)	(0.004)	(0.005)	(0.006)	(0.005)	

Notes: Source: employee questionnaires. Employee sampling weights are used to calculate the standard errors.

Since this study is interested in firms' characteristics, all variables will be used at the establishment level. For career mentoring, firstly, I transformed the ordinal variable into a dummy variable which equals to 1 if the response is either agree or strongly agree and otherwise 0. Then the mean of this dummy variable is computed for each workplace to give the mean percentage of employees who stated that managers at this workplace encouraging people to develop their skills.[11]

4.2.2 Share of Female Managers and Managerial Incentives

The first objective of this paper is to test whether an increase in the presence of female managers is associated with more career mentoring at the workplace. The share of female managers is derived by dividing the number of female managerial employees (both full-time and part-time) by the total number of managerial employees (summation of male and female managerial employees).[12]

Furthermore, I am interested in examining the effect of managerial incentives on managers' career mentoring. The two managerial incentive schemes used in this study are *Performance related pay* (PERFM) and *Profit related pay* (PROFT), respectively. These are the only two incentive schemes from the survey that can be linked to managers at the workplace.[13] For both incentive schemes I create dummy variables which equal 1 if managers in the workplace were offered the corresponding scheme and 0 otherwise. A PERFM scheme links part or all of pay to either individual or group performance, often in terms of the achievement of agreed objectives and targets. In contrast, a PROFT scheme is one where part of the pay is linked to company profitability.

[10] Notably, not all employees answered this question. The response rate is 97% in 2004 and 98% in 2011, respectively.

[11] The employee sampling weights, which designed to remove known biases introduced by the sample selection and response process, are used for this computation. See Appendix C for a full discussion of sampling weights.

[12] Detailed definition of manager is presented in Appendix A.

[13] In addition, there is one complex incentive scheme called the employee share ownership scheme, which includes share incentive plan, save as you earn share options scheme, enterprise management incentives, company share option plan, and other employee share scheme.

As documented in the 2011 WERS *Interview Handbook*, PERFM may include an element of merit rating, and personal qualities and inputs are assessed in conjunction with outputs. Payouts may be given as bonuses or incremental pay awards, and may replace part or all general pay increases. PROFT is a performance measure that reflects an employee's contribution to the company's profits. Most PROFT is paid as a bonus rather than as part of basic pay or as a replacement for annual pay increases, and payments are usually made in cash or shares on an annual basis. Most importantly, it is unlikely that both incentive schemes were used in conjunction. This is because, when the respondent was asked whether managers in the workplace were paid by PERFM, they received a card showing the precise definition of this incentive scheme and stating that this payment does not include PROFT.

The panel sample used consists of all participating workplaces from 2004 that remained in existence in 2011. As shown in Table C1 (Appendix D) establishment characteristics are identical between the pooled cross sectional sample and the panel sample in 2004 and Table C3 further shows that none of the firms' attributes used in this paper are statistically predictive of selection into the panel sample. Therefore, the possibility of selection bias can be ruled out while using the panel sample.

4.3 Empirical Strategy

To test the association between career mentoring and the share of female managers, I estimate the following regression model for gender representation on the management of firm i:

$$Y_{i,t} = \alpha + \beta_1 ShareFemales_{i,t} + \beta_2 PERFM_{i,t} + \beta_3 PROFT_{i,t} + Z'_{i,t}\beta + \epsilon_{i,t}, \tag{4.1}$$

where $Y_{i,t}$ is the dependent variable, measuring the percentage of employees who claimed that with their managers' mentoring activities in firm i at time t. ShareFemales is the share of female managers. PERFM and PROFT are binary variables indicate whether the manager was offered group performance related pay and whether the manager was offered profit related pay, respectively. The control variables which is denoted by Z include year dummies, industry and formal status dummies, region dummies, number of employees, recruitment preferences, competition condition in the market, quality of products/services, and UK market share; ϵ is the error term.

Year fixed effects are included to control for any economy-wide shocks and general trends affecting manager's mentoring behaviour symmetrically across all workplaces. However, there may also be factors which influence the behaviour or supply of managers that vary across industries and formal statuss. For instance, mentoring may be more effective for individuals working in financial services than those working in construction. Therefore industry and formal status-specific factors are added in the specification. Furthermore, the distribution of either industries or formal statuss may vary across regions. Therefore, region fixed effects are also included in specification (4.1).

Even after accounting for year, industry, formal status and region fixed effects, there may still remain differences across firms in factors that vary over time and influence the prevalence of mentoring relationships, such as recruitment preferences, product/service quality, firm size, market share, and market condition.[14] At firms where internal promotions are preferred to external applicants mentoring relationships are probably more prevalent (Laband and Lentz, 1999). Similarly, in a company that produces high-quality products and services, mentoring programs may have been used more frequently than in others to minimise human errors.

There is also evidence showing that formalisation of HRM practices is associated with organisational size (Marsden et al., 1996). Big companies with large market share are attracting talented graduates by offering prolonged and advanced mentoring programs, such as Boeing, Intel, and GE. Market conditions may affect the prevalence for mentoring as well: firms that operate in a competitive market, where the labour supply is perfectly elastic, may find good candidates without providing a reliable mentoring program.

Going beyond specification (4.1), firm fixed effects can be included when using a panel sample:

$$Y_{i,t} = \alpha + \gamma_i + X'_{i,t}\beta + \epsilon_{i,t}, \tag{4.2}$$

where γ_i are firm fixed effects. This allows me to control for time-invariant unobserved workplace attributes that may influence manager's mentoring behaviours, such as company culture. X denotes the share of female managers, PERFM, PROFT, and other controls Z as discussed in specification (4.1).

Notice that although the panel data nature allows to control for the time-invariant unobserved heterogeneity, it cannot account for time-variant unobserved heterogeneity. Under the assumption of strict exogeneity, $\epsilon_{i,t}$ does not correlate with $X'_{i,t}$ for all time periods (i.e. $E(\epsilon_{i,t}|X_{i,1}, X_{i,2}, ..., X_{i,T},) = 0, t = 1, ..., T$), my results can be interpreted as causal effects.[15]

To measure the potential effect of managerial incentive schemes on manager's career mentoring activities I exploit again the panel structure of the sample: the data set consists of two periods and allows me to identify the change of managerial incentive schemes in each workplace i. It thus raises the possibility to use a Difference-in-Differences (DiD) approach. However, the parallel trend assumption cannot be tested, because there are only two periods of data. The specification is given by:

$$Y_{i,t} = \alpha + \delta_1 P_i + \delta_2 T_t + \delta_3 (P * T)_{i,t} + X'_{i,t}\beta + \epsilon_{i,t} \tag{4.3}$$

[14] Detailed definitions of these control variables are provided in Appendix A.

[15] One could argue that the supply of female graduates can be used as an instrumental variable for the share of female managers, but the number of graduates may also suffer an endogeneity problem. A better educated labour force entering the industry implies more managerial potentials, which may lead to more mentorships.

where T_t equal to one, if the observations are from 2011 and zero otherwise. Denote by X the vector of control variables and ϵ is the error term. P_i indicates the change of a payment scheme for the manager in the workplace i. Given the originality of measuring the impact of managerial incentives on managers' career mentoring I only estimate the effect of either introducing or abolishing a single payment scheme. Therefore, the parameter of interest δ_3 gives a clean measure of the impact of PERFM/PROFT. Take PERFM, for example - P_i is constructed as:

Introducing PERFM

$P_i(I) = 1$ if the workplace had used PERFM in 2011 but not in 2004;

$P_i(I) = 0$ if the workplace had used PERFM in neither 2004 nor 2011;

Abolishing PERFM

$P_i(A) = 1$ if the workplace had used PERFM in 2004 but not in 2011;

$P_i(A) = 0$ if the workplace had used PERFM in both 2004 and 2011.

Results for both introducing a new payment scheme and abolishing a previous payment scheme will be shown in Section 4.4.4. With the assumption of strict exogeneity and unviolated parallel trend the results can be interpreted as causal effects.

4.4 Results

4.4.1 Descriptives

Table B4 reports descriptive statistics for our outcome variable of interest career mentoring, the share of female managers, and the two managerial incentive schemes - the performance related pay (PERFM) and the profit related pay (PROFT). The first row of Panel A shows that, on average, 65% of employees from both pooled cross sectional sample and panel sample stated that their managers used career mentoring in 2004.[16] In contrast, in 2011, the figures drop 1% for the pooled cross sectional sample and further fall by 7% for the panel sample as can be seen in Panel B.

The second row of each Panel shows that the share of female managers is almost identical within a single survey wave across samples. However, there is a roughly five percent rise in the mean percentage of female managers from 2004 to 2011 for both samples.

The last two rows of Panel A illustrate that 24% of the establishments from the pooled cross sectional sample offered managers PERFM while 26% offered PROFT in 2004. However, for

[16] The pooled cross sectional data consists of all establishments that have been surveyed in the two periods. Each followed-up establishment that existed in both years gives two distinct observations. However, the panel data is a sub-sample of the pooled cross sectional data. It only includes the 2004 establishments which have been followed up in 2011.

firms from the panel sample, 21% of them offered PERFM and only 17% offered PROFT in 2004. In 2011, as shown in Panel B, the mean percentage of establishments offered managers PERFM increases by 6% while the figure for PROFT remains numerically identical compared to 2004. For the panel sample, more firms offered managers PERFM and PROFT in 2011 than in 2004, about 9% and 6% rise respectively.

Table 4.2: Descriptives on Career Mentoring, the Share of Female Managers, and Managerial Incentive Schemes

	Pooled Cross Sectional Data				Longitudinal Data			
VARIABLES	Mean	Proportion	Std. Err.	Obs	Mean	Proportion	Std. Err.	Obs
Panel A. 2004								
Career Mentoring	0.647		0.012	1,732	0.654		0.019	599
Share of Female Managers	0.392		0.018	1,564	0.409		0.031	538
PERFM		0.241	0.017	1,729		0.208	0.028	598
PROFT		0.262	0.018	1,728		0.171	0.026	598
Panel B. 2011								
Career Mentoring	0.635		0.014	1,921	0.571		0.024	599
Share of Female Managers	0.439		0.017	1,825	0.454		0.027	572
PERFM		0.308	0.020	1,920		0.295	0.033	598
PROFT		0.264	0.019	1,921		0.237	0.032	599

Notes: "Career Mentoring" indicates the percentage of employees who stated that managers at the workplace encouraging people to develop their skills. "PERFM" indicates the performance related pay scheme while "PROFT" for the profit related pay scheme. Establishment sampling weights are used to calculate the standard errors.

Overall, the evidence in Table B4 for the pooled cross sectional sample indicates that the percentage of employees who stated that their managers provided career mentoring and the number of firms offered managers PROFT are consistent over time, while the number of firms that offered managers PERFM and the share of female managers have grown rapidly. However, the firms interviewed in the follow-up survey (panel sample) have a smaller percentage of employees stating that managers used career mentoring in 2011 than in 2004 while the figures for other variables escalate. This may suggest a negative association between career mentoring and the share of female managers/managerial incentives.

4.4.2 Linear Regressions v.s. Fixed Effects

To test the relationship between career mentoring (the percentage of employees who claimed that managers at the workplace encouraging people to develop their skills) and the share of female managers, specification (4.1) which does not include the firm fixed effects is estimated with both pooled cross sectional data and panel data. The fixed effects model - specification (4.2) is only examined by using the panel data.

The estimates are reported in Table 4.3. Columns 1-4 present the estimates for linear regressions (by performing OLS) in the absence of firm fixed effects, while columns 5 and 6 report the fixed effect estimations (FE) where firm fixed effects are included. Other control variables include year dummies, industry and formal status dummies, region dummies, number of employees, recruitment preferences, competition condition in the market, quality of products/services, and UK market share. Robust standard errors clustered at establishment level are reported in brackets below the estimates, which allow errors in different time periods for a given establishment to be correlated, while errors for various establishments are assumed to be uncorrelated.

Table 4.3: Linear Regressions (OLS) v.s. Fixed Effects (FE)

	Pooled Cross Sectional Data		Longitudinal Data			
	OLS	OLS	OLS	OLS	FE	FE
	(1)	(2)	(3)	(4)	(5)	(6)
Share of Female Managers	0.0880***	0.0876***	0.1070***	0.1050***	-0.0032	-0.0090
	(0.0156)	(0.0156)	(0.0290)	(0.0291)	(0.0654)	(0.0668)
PERFM		0.0288***		0.0013		0.0255
		(0.0088)		(0.0153)		(0.0321)
PROFT		0.0082		0.0161		0.0210
		(0.0108)		(0.0197)		(0.0393)
Constant	0.4630***	0.4470***	0.4510***	0.4450***	0.3620**	0.3500**
	(0.0285)	(0.0291)	(0.0447)	(0.0455)	(0.1440)	(0.1440)
Observations	3,386	3,374	1,108	1,101	1,108	1,101
Number of Firms	2,832	2,823	554	551	554	551
R-squared	0.153	0.155	0.252	0.252	0.759	0.761
Firm Fixed Effects	NO	NO	NO	NO	YES	YES
Other Controls	YES	YES	YES	YES	YES	YES

Notes: The dependent variable is Career Mentoring which indicates the percentage of employees who claimed that managers at the workplace encouraged people to develop their skills. The empirical analysis is conducted at the establishment level, and robust standard errors clustered at establishment level are reported in brackets below the estimates. Columns (1) and (2) report estimates for the baseline model (excluding firm fixed effects) with the cross-sectional sample. Columns (3) and (4) report estimates for the baseline model with the panel sample. By using the panel data, columns (5) and (6) include firm fixed effects. PERFM equals 1 if managers were offered performance related pay scheme and otherwise 0. PROFT equals 1 if managers were offered profit related pay scheme and otherwise 0. Other controls include year dummies, industry and formal status dummies, region dummies, firm size, recruitment preferences, competition condition in the market, quality of products/services, and UK market share. *** Significant at 1% level, ** significant at 5% level, * significant at 10% level.

Column 1 indicates that the share of female managers is positively associated with career mentoring in the absence of firm fixed effects. The effect is statistically significant at the 1 % level in the dataset with a cross-sectional structure. A one percentage point increase in the share of female managers is associated with a 0.088 percentage points rise in the occurrence of career mentoring. In column 2, the two different types of manager's payment schemes, i.e., performance related pay (PERFM) and profit related pay (PROFT), are added into the regression specification. The share of female managers retains its positive and significant association with career mentoring in the absence of firm fixed effects, and PERFM is positively associated with career mentoring (p-value < 0.01) while the association between PROFT and career mentoring is positive but statistically insignificant. The coefficient for PERFM, for example, can be interpreted as in firms who offered managers performance related pay the average proportion of employees who affirmed that their managers provided career mentoring is about 0.0288 percentage points higher than those firms who did not offer this payment scheme. Compared to columns 1 and 2, columns 3 and 4 repeat the same estimations with the panel sample. The coefficients for the share of female managers increased by 24% and are statistically significant at 1 % level. However, the association between PERFM and career mentoring diminished dramatically while the coefficients of PROFT remain positive and insignificant.

To address the main question of whether the correlation between the occurrence of career mentoring and the presence of female managers will remain significant after including firm fixed effects, columns 5 and 6 show the results from estimating specification (4.2). The coefficients of the share of female managers become negative and statistically insignificant. For PERFM and PROFT, the estimates remain positive, but insignificant.[17][18]

Overall, these results confirm that the positive and significant association between the share of female managers and career mentoring can only be found in the absence of firm fixed effects. A possible explanation is that the percentage of women in the managerial occupations determines (or be determined by) a firm's culture, which may be time invariant and may further form other unobservables such as informal mentoring relationships between managers and employees. For managerial incentive schemes results remain ambiguous. This may be because a pooled estimation only compares firms that had an incentive scheme with those that did not have one. A subsample analysis will be presented in Section 4.4.4 to look into the differences between companies who changed their managerial incentives over time and those who did not.

These results can be interpreted as causal effects if there is no time varying omitted variable that is correlated with both managers' career mentoring and the share of female managers/-managerial incentives.

4.4.3 Subsample Analysis: Does Employees' Gender Play a Role?

It has been argued that when managers and employees are similar along an easily observable demographic characteristic, such as gender, mentoring is more profound (Athey et al., 2000). Relationship research suggests that, on average, people prefer others who they perceive as similar to them (Hinde, 1997). Studies find that the low number of females in upper management positions, coupled with the fact that the few available female mentors are already overloaded with female mentees, means that women are often less likely to find mentors (Burke and McKeen, 1997; Ragins, 1989). As a consequence, a larger share of female managers may increase the likelihood for female employees to establish mentoring relationships, which will be tested in this section.

To examine whether the number of female employees who confirmed managers' mentoring activities increases when the share of female managers rises, I split the employee questionnaires into two groups based on employees' gender. By estimating specifications (4.1) and (4.2), Panel A of Table 4.4 shows the estimates for female employees, while results for males are presented in Panel B. Robust standard errors reported in brackets below the estimates are clustered at the establishment level.

[17] The results are identical when standard errors are calculated by establishment sampling weights or not adjusted at all. I also performed the same analyses at the employee level by using the original mentoring variable (5-point Scale) and controlling for employees' characteristics such as gender, age, education, etc. The conclusions remain the same. All of the results are available upon request.

[18] One could also include the interactions between the share of female managers and managerial incentives into the specifications. The conclusion remains the same. Results are available upon request.

Table 4.4: Linear Regressions (OLS) v.s. Fixed Effects (FE) by Employee Respondents' Gender

	Pooled Cross Sectional Data	Longitudinal Data	
	OLS	OLS	FE
	(1)	(2)	(3)
Panel A. Female Employee Respondents			
Share of Female Managers	0.1040***	0.1110***	0.0167
	(0.0182)	(0.0356)	(0.0903)
PERFM	0.0218*	-0.0094	0.0033
	(0.0115)	(0.0207)	(0.0442)
PROFT	0.0152	0.0300	0.0338
	(0.0148)	(0.0281)	(0.0651)
Constant	0.4710***	0.3800***	0.2730
	(0.0419)	(0.0794)	(0.3120)
Observations	3,129	1,020	1,020
Number of Firms	2,619	510	510
R-squared	0.094	0.163	0.725
Panel B. Male Employee Respondents			
Share of Female Managers	0.0364	0.0416	-0.1170
	(0.0248)	(0.0479)	(0.1320)
PERFM	0.0229*	-0.0037	0.0192
	(0.0125)	(0.0218)	(0.0488)
PROFT	0.0072	0.0047	0.0377
	(0.0140)	(0.0264)	(0.0543)
Constant	0.4880***	0.5470***	0.9680***
	(0.0380)	(0.0640)	(0.1710)
Observations	2,964	980	980
Number of Firms	2,474	490	490
R-squared	0.095	0.123	0.727
Firm Fixed Effects	NO	NO	YES
Other Controls	YES	YES	YES

Notes: In Panel A (B), the dependent variable indicates the percentage of respondents among female (male) employees who claimed that managers at the workplace encouraged people to develop their skills. The empirical analysis is conducted at the establishment level, and robust standard errors clustered at establishment level are reported in brackets below the estimates. Columns (1) and (2) report estimates for the baseline model (excluding firm fixed effects) with the cross-sectional sample and the panel sample, respectively. By using the panel data, columns (3) includes firm fixed effects. PERFM equals 1 if managers were offered performance related pay scheme and otherwise 0. PROFT equals 1 if managers were offered profit related pay scheme and otherwise 0. Other controls include year dummies, industry and formal status dummies, region dummies, firm size, recruitment preferences, competition condition in the market, quality of products/services, and UK market share. *** Significant at 1% level, ** significant at 5% level, * significant at 10% level.

Consistent with the gender similarity theory, columns 1 (pooled cross sectional sample) and 2 (longitudinal sample) indicate that a larger share of female managers is significantly associated with more female employees among their own gender cohort stating that managers encouraged people to develop their skills at the workplace in the absence of firm fixed effects. Consistent with the results found in Table 4.3, the coefficient of the share of female managers becomes much smaller and insignificant after including the firm fixed effects as shown in column 3. In contrast, coefficients for male respondents are relatively small in size and statistically insignificant across all specifications. These results suggest that gender does not play a role to mediate the effect of female managers on the occurrence of mentoring when the strict exogeneity assumption for fixed effects models is not violated.

4.4.4 Subsample Analysis: Managerial Incentives

By using specification (4.3) I further compare firms who shared the same managerial incentives in 2004 but made some changes in 2011 to the firms who did not change their incentive systems. In particular, column 1 in Table 4.5 reports the estimated coefficients of introducing the performance related payment scheme, which compares 96 firms that have used PERFM in 2011, but not 2004, with 316 firms that have used this scheme in neither 2004 nor 2011. Column 2 reports the estimated coefficients of abolishing this payment scheme, which compares 85 firms that have used PERFM in 2004, but not 2011, with 99 firms that have used this scheme in both 2004 and 2011. Similarly, column 3 compares 70 firms that have introduced PROFT in 2011 with 412 firms that have never used PROFT and column 4 compares 55 firms that have abolished PROFT in 2011 with 60 firms that have used PROFT in both years.[19]

Table 4.5: The Effect of Changing Managerial Incentives on Career Mentoring

	Introduce PERFM (1)	Abolish PERFM (2)	Introduce PROFT (3)	Abolish PROFT (4)
PERFM	-0.0135	0.0594*		
	(0.0290)	(0.0347)		
PERFM*2011	0.0433	-0.0704*		
	(0.0339)	(0.0425)		
PROFT			-0.0158	0.0130
			(0.0346)	(0.0514)
PROFT*2011			0.0606*	-0.0572
			(0.0345)	(0.0622)
2011	-0.0507***	0.0278	-0.0453***	0.0068
	(0.0159)	(0.0267)	(0.0134)	(0.0385)
Constant	0.458***	0.504***	0.431***	0.619***
	(0.0589)	(0.0801)	(0.0532)	(0.119)
Number of firms	412	184	482	115
R-squared	0.245	0.303	0.250	0.277
Other Controls	YES	YES	YES	YES

Notes: OLS estimations based on specification 4.3. The dependent variable is Career Mentoring which indicates the percentage of employees who claimed that managers at the workplace encouraged people to develop their skills. The empirical analysis is conducted at the establishment level, and robust standard errors clustered at establishment level are reported in brackets below the estimates. PERFM equals 1 if managers were offered performance related pay scheme in the firm and otherwise 0. PROFT equals 1 if managers were offered profit related pay scheme in the firm and otherwise 0. Other controls include year dummies, industry and formal status dummies, region dummies, firm size, recruitment preferences, competition condition in the market, quality of products/services, and UK market share. *** Significant at 1% level, ** significant at 5% level, * significant at 10% level.

The coefficient of interest δ_3 is given by the coefficients of PERFM $*$ 2011 which is positive if PERFM was introduced into the workplace, but negative if PERFM was abolished in 2011. This suggests that PERFM has a positive effect on career mentoring. The occurrence of career mentoring increased by 0.0433 percentage points (statistically insignificant), if PERFM was introduced in 2011 and decreased by 0.0704 percentage points (statistically significant at the 10% level), if PERFM was abolished.[20] This weak significant relationship disappears if the share of female managers is included in the specification as shown in Table C5.

With respect to the profit related payment scheme (PROFT), introducing PROFT into the workplace increased the occurrence of career mentoring by 0.0606 percentage points, which is statistically significant at the 10% level. Adding the share of female managers into the specification

[19]Detailed descriptives are reported in Table C4 in appendix.

[20]Similar results can be produced when standard errors are calculated by establishment sampling weights or not adjusted at all. However, the coefficients for PERFM $*$ 2011 when PERFM was abolished are no longer significant, and the coefficient for PERFM $*$ 2011 when PERFM was introduced becomes statistically significant at 10% level if standard errors are calculated by establishment sampling weights. Results are available upon request.

does not change this result (See Table C5). Abolishing PROFT gives a negative and insignificant coefficient (-0.0572).[21]

Therefore, the results provide weak evidence supporting a positive association between managerial incentive schemes and career mentoring in the workplace. Since the (group) performance related pay is quantifiable, the positive association can be explained as managers encouraged their employees to develop skills to improve the overall rating of their performance. For managers whose payments are related to corporate profits, they may use mentoring as a tool to retain and promote talented employees and manifest it as their contributions to the company. Nonetheless, the insignificant results may be due to the small sample size and a short time span.[22]

The validity of the identification strategy and the causal interpretation given to the results relies on two assumptions. The first is the parallel trend assumption. The second underlying identifying assumption is that the effect of changing managerial incentives on managers' career mentoring does not change over time for reasons other than economy-wide trends, industry-specific factors, formal status-specific factors, region-specific factors, firm size, recruitment preferences, competition condition in the market, quality of products/services, and UK market share. This ensures that there are no time-varying unobservables that (i) are correlated with managers' career mentoring and (ii) determine the change of managerial incentives.

4.5 Conclusion

Does an increase in the number of female managers help employees to receive more encouragement and mentoring from managers? Using a large sample of British workplaces this study suggests that this is not the case. Although by using a cross-sectional sample this study shows that the share of female managers is positively and significantly associated with the occurrence of mentoring relationships in the workplace, especially for female employees, adding firm fixed effects shows that this significant association disappears completely. However, on the other hand, the empirical evidence points to a positive but weak effect of group performance pay and profit related pay schemes on the occurrence of mentoring. These results suggest that it is beneficial for firms to link their managers' payment to corporate profits rather than focus on the gender of managers, if they value mentoring.

The data do not permit to infer causation, and a word of caution appears appropriate. Further research on this matter is required for informing policy, ideally in the form of a field experiment. A potential direction could be to implement different managerial incentive schemes in a real-life workplace and see how managers react to these incentives in terms of mentoring.

[21]The implications remain the same for abolishing PROFT if standard errors are calculated by establishment sampling weights or not adjusted at all, but the coefficients for PROFT $*$ 2011 when PROFT was introduced become statistically insignificant. Results are available upon request.

[22]However, specification 4.3 can not be tested with the original mentoring variable (5-point Scale) due to the estimates of δ_3 are biassed and inefficient in a nonlinear model (Ai and Norton, 2003).

Indeed, there are several possible reasons for engaging in mentoring. One, it could be thought of as a "warm glow", by receiving personal pride and gratification from observing an assigned mentee grow and develop. It is a sense of satisfaction one feels from the success of an assigned mentee that one has encouraged and helped to achieve. Nevertheless, researchers are encouraged to go beyond easily observable manager characteristics so as to further explore the impact of HRM practices on mentoring relationships in the workplace, such as managerial incentive schemes.

Second, and perhaps the more important consideration is that conventional stereotype about gender differences in terms of females being more caring and therefore providing more mentoring might be misleading. The gender based stereotyping may result in differences in structure of compensation by gender and further lead to the gender pay gap (Albanesi and Olivetti, 2009; Albanesi et al., 2015; Johnson and Scandura, 1994). An employer who appoints a female manager to a position where mentoring is desired may not compensate her for mentoring as gender stereotype would enforce mentoring as an intrinsic part of female characteristic. A male manager on the other hand would be compensated for providing mentoring as it would be considered an additional responsibility. The evidence of a negative relationship between the share of female managers and the use of group based incentives is also found in my data, see Figure C3 in the appendix.

Chapter 5

Conclusions

In this Thesis, I study three aspects of organisational economics focusing on teams, especially organisational capital and incentives in teams and before team assignment. In particular, I have touched practices covering peer effects (mentoring), dynamic effects of team composition (training investments), and remuneration (multi-tasking in teams).

By using a large sample of British workplaces, I do not find evidence that gender is strongly predictive of mentoring supports. But, I show that there is a positive but weak effect of group-based incentives on the occurrence of mentoring. These results suggest that researchers study manager-employee supporting relationships should go beyond individual characteristics and to explore the role of management practices such as the remuneration system.

Taking it further, I conduct a field experiment in two Chinese factories to shed light on the impact of higher-powered incentives on multitasking leaders' organisational behaviours. Group leaders are usually responsible for organising teams and contributing to the goal as a member. As in the context of our experiment, a group leader - called foreman in the manufacturing factories - who makes products but also organises the group production process. However, the foreman's organisational inputs that contribute to a public good are hard to measure. By implementing a monitoring system that subjectively evaluates the foremen, the organisational inputs become quantifiable. I then introduce an incentive scheme that provides each foreman with a bonus depending on the foreman's relative position among other foremen in the factory. By using the difference-in-difference technique, I find an overall 6% increase in workers' productivity in the treated factory relative to the controlled one. As for the multitasking foremen, I show that the new bonus scheme increases their effort provisions in both tasks as they invest more time on the job, given that there is a positive spillover effect in the workplace.

Lastly, team composition has also been identified to play a vital role to influence team performance through peers' attributes. However, the attributes that determine payoffs are likely to be the consequence of prior choices made. We designed a laboratory experiment to test whether subjects anticipate the assignment which is based on prior performance and how this may affect their prior performance. We find that pairing the worst performing individuals with the best yields 20% lower pre-assignment effort than random matching and does not induce truthful revelation of types, which undoes any policy that aims to reallocate types based on performance. Pairing the best with the best, however, yields only 5% higher pre-assignment effort than random matching.

Together these studies provide a wide research agenda to pursue to better understand incentives in teams. One potential direction is to combine our technology used in the Chinese firms and the algorithms implemented in the lab to reassign group leaders to different groups based on their rankings. For instance, the highest ranked foreman will be teamed up with the lowest productive production line if the negative assortative matching is used. This would allow us to test the dynamic effect of team paring mechanisms in a multi-tasking environment.

Appendix A

Appendix for Chapter 2

A.1 Model Details

A.1.1 Assumptions on w

We impose the agent's limited liability constraint on the piece rate w in the firm's profit function $E\pi(.)$, therefore, it must be positive. We also impose that individual efforts cannot be negative.

First, we assume that w is large enough. By looking at equations 2.1, 2.2, and 2.3 we know that all effort levels are negative if $\lambda w > 2$. This means that no one exerts any effort and the total production output equals zero, so the firm's profit is negative ($-b$) which would not occur in practice. Similarly, if $\lambda w = 2$ output is infinite and profit is infinite. Hence, we assume $\lambda w < 2$ to rule out these cases.

Now, the inspection of λw in equation 2.2 reveals that g_f equals to zero if $\lambda w \leq 1$ because individual efforts cannot be negative. Both the foreman and the worker solve their maximisation problems considering $\overline{g}_f = 0$, and we have $\overline{e}_f = \overline{e}_w = w$.

Furthermore, if $1 < \lambda w < 2$, g_f^*, e_f^*, and e_w^* as expressed in equations 2.1, 2.2, and 2.3 are all positive. Thus, all three first order conditions hold, and we would predict that the foreman and the worker who are offered a piece rate such that $1 < \lambda w < 2$ would choose the interior levels of effort g_f^*, e_f^*, and e_w^*.

A.1.2 Proof of Lemma 2.1

We first analyse the case when $1 < \lambda w < 2$, substituting the optimal effort levels e_w^*, e_f^*, and g_f^* into the output functions the firm's optimisation problem can be written as below:

$$\max_{\{w\}} E\pi = (p-w)(y_f(e_f^*, g_f^*) + y_w(e_w^*)) - b$$
$$= (p-w)\frac{\lambda w^2(1+\lambda w)}{(1-(\lambda w-1)^2)^2} - b.$$

The first order condition with respect to w is therefore given by:

$$\lambda(4+\lambda w)(p-w) - (2+\lambda w - \lambda^2 w^2) = 0 \tag{A.1}$$

solving this equation we get:

$$w^* = \frac{4p\lambda - 2}{5\lambda - p\lambda^2}. \tag{A.2}$$

Substituting firm's offer w^* into the foreman's and worker's optimal effort levels, we have:

$$g_f^* = \frac{5p\lambda - 7}{12\lambda - 6p\lambda^2},$$
$$e_f^* = \frac{5 - p\lambda}{12\lambda - 6p\lambda^2},$$
$$e_w^* = \frac{4p\lambda - 2}{12\lambda - 6p\lambda^2}.$$

Taking together, the firm's expected profit when $1 < \lambda w < 2$ can be written as:

$$E\pi^* = (p-w^*)\frac{\lambda(w^*)^2(1+\lambda w^*)}{(1-(\lambda w^*-1)^2)^2} - b = (p-w^*)\frac{1+\lambda w^*}{\lambda(2-\lambda w^*)^2} - b$$
$$= (p - \frac{4p\lambda-2}{5\lambda-p\lambda^2})\frac{1+\lambda\frac{4p\lambda-2}{5\lambda-p\lambda^2}}{\lambda(2-\lambda\frac{4p\lambda-2}{5\lambda-p\lambda^2})^2} - b$$
$$= (p - \frac{4p\lambda-2}{5\lambda-p\lambda^2})\frac{(p\lambda+1)(5-p\lambda)}{12\lambda(2-p\lambda)^2} - b$$
$$= \frac{p(p\lambda+1)(5-p\lambda)}{12\lambda(2-p\lambda)^2} - \frac{(4p\lambda-2)(p\lambda+1)}{12\lambda^2(2-p\lambda)^2} - b$$
$$= \frac{[p\lambda(5-p\lambda)-(4p\lambda-2)](p\lambda+1)}{12\lambda^2(2-p\lambda)^2} - b$$
$$= \frac{(p\lambda+1)(2-p\lambda)(p\lambda+1)}{12\lambda^2(2-p\lambda)^2} - b = \frac{(p\lambda+1)^2}{12\lambda^2(2-p\lambda)} - b,$$

note that from equation A.2 we know that the underlying assumption for π^* ($1 < \lambda w < 2$) only holds if $1.4 < p\lambda < 2$.

If $\lambda w \leq 1$, substituting $\overline{g}_f = 0, \overline{e}_f = \overline{e}_w = w$ into the firm's expected profit function, we have:

$$E\overline{\pi} = (p - w)2w - b.$$

Solving this maximisation problem for the firm, the firm would prefer to set its piece rate at $\frac{p}{2}$. The optimal effort levels of the foreman and the worker are therefore given by: $\overline{g}_f = 0, \overline{e}_f = \frac{p}{2}$, and $\overline{e}_w = \frac{p}{2}$. These results only exist if $\lambda\overline{w} = \lambda * \frac{p}{2} < 1 \Rightarrow p\lambda < 2$, and we can rewrite the firm's expected profit as:

$$E\overline{\pi} = (p - \overline{w})2\overline{w} - b = \frac{p^2}{2} - b.$$

To see the firm's choice between these two potential outcomes, given the fact that π^* only exists if $1.4 < p\lambda < 2$, taking the difference between $E\pi^*$ and $E\overline{\pi}$ gives us:

$$\begin{aligned} E\pi^* - E\overline{\pi} &= \frac{(p\lambda + 1)^2}{12\lambda^2(2 - p\lambda)} - b - (\frac{p^2}{2} - b) \\ &= \frac{(p\lambda + 1)^2}{12\lambda^2(2 - p\lambda)} - \frac{p^2}{2} \\ &= \frac{(p\lambda + 1)^2 - 6p^2\lambda^2(2 - p\lambda)}{12\lambda^2(2 - p\lambda)} \\ &= \frac{p^2\lambda^2 + 2p\lambda + 1 - 12p^2\lambda^2 + 6p^3\lambda^3}{12\lambda^2(2 - p\lambda)} \\ &= \frac{(2p\lambda - 1)(3(p\lambda)^2 - 4p\lambda - 1)}{12\lambda^2(2 - p\lambda)}, \end{aligned}$$

the solutions of this function when it is equal to zero are $p\lambda = \frac{2-\sqrt{7}}{3} \approx -0.22, 0.5, and \frac{2+\sqrt{7}}{3} \approx 1.55$. $E\pi^* - E\overline{\pi} > 0$ if $\frac{2-\sqrt{7}}{3} < p\lambda < 0.5$ or $p\lambda > \frac{2+\sqrt{7}}{3}$. $E\pi^* - E\overline{\pi} < 0$ for $0.5 < p\lambda < \frac{2+\sqrt{7}}{3}$. Because both the market price and the team efficiency spillover are greater than zero, we have $E\pi^* > E\overline{\pi}$ if $0 < p\lambda < 0.5$ or $p\lambda > \frac{2+\sqrt{7}}{3}$, $E\pi^* < E\overline{\pi}$ if $0.5 < p\lambda < \frac{2+\sqrt{7}}{3}$, and $E\pi^* = E\overline{\pi}$ if $p\lambda = 0.5$ or $\frac{2+\sqrt{7}}{3}$.

As we already know that $\overline{\pi}$ exists if $p\lambda < 2$, and π^* is attainable if and only if $1.4 < p\lambda < 2$. Taking together, the profit maximising firm is indifferent between $\overline{w}$ and w^* if $p = p^* = \frac{2+\sqrt{7}}{3\lambda}$. It chooses $w = w^*$ if $p^* < p < \frac{2}{\lambda}$. Under w^*, the foreman and the worker both invest positive effort levels and the expected profit of the firm equals $\frac{(p\lambda+1)^2}{12\lambda^2(2-p\lambda)} - b$. On the other hand, if $0 < p < p^*$ the firm sets the piece rate at $\overline{w}$. The foreman's organisational effort is zero while both the foreman and the worker choose production effort equals to the piece rate $\overline{w}$. The firm receives $\frac{p^2}{2} - b$ as a return.

A.2 Other Figures

Figure B1: The Disposable Infusion Sets

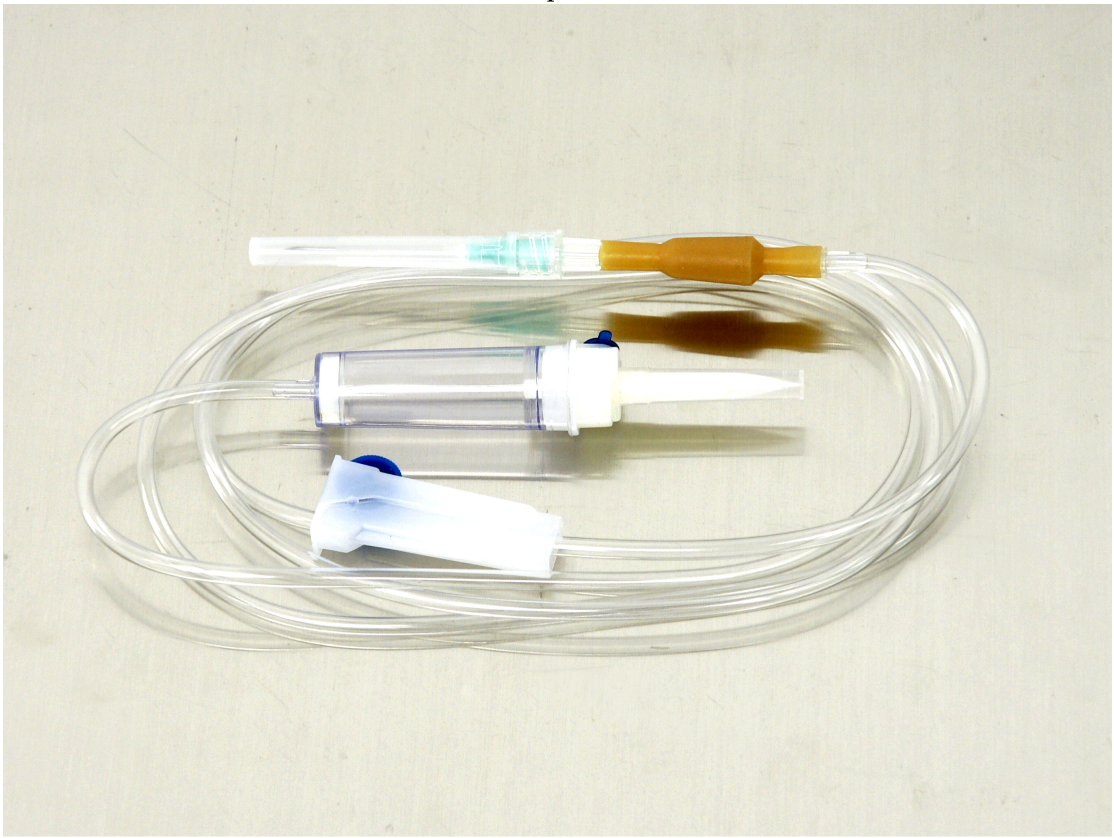
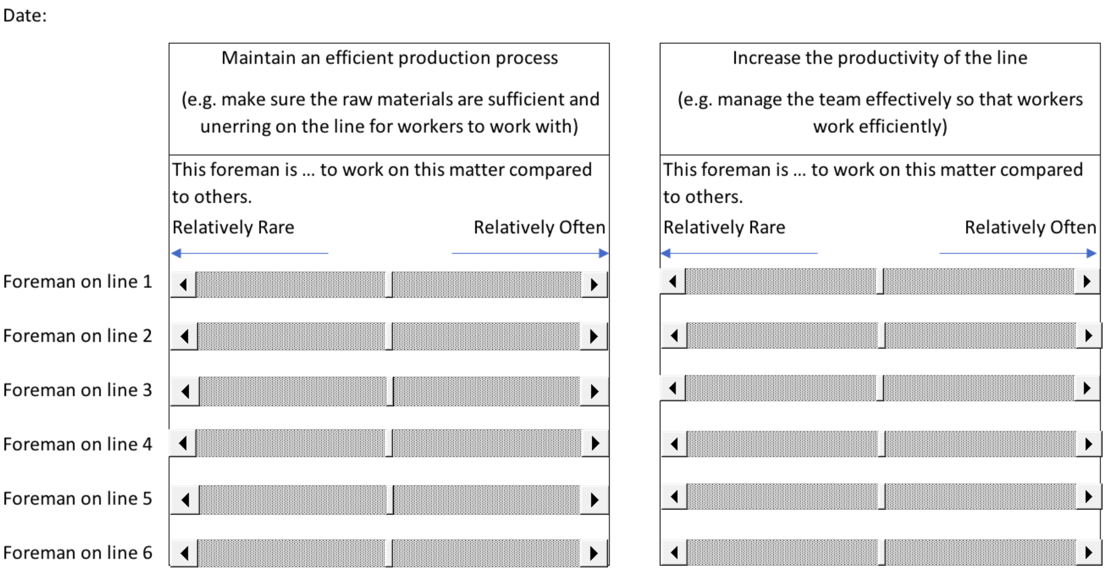


FIGURE B2: The Production Lines

FIGURE B3: Sliders for Ranking the Foremen



Reduce line defect rates (e.g. constantly remind workers to use standardised operating procedure in order to reduce the number of faulty products)	Team building (e.g. provide support and communication to foster a friendly and positive work environment)	**Overall Ranking**
This foreman is ... to work on this matter compared to others. Relatively Rare ⟵⟶ Relatively Often	This foreman is ... to work on this matter compared to others. Relatively Rare ⟵⟶ Relatively Often	
		1
		1
		1
		1
		1
		1

Figure B4: The Poster Showing the Evaluation Results of Foremen

A.3 Other Tables

Table C1: Summary Statistics for Other Individual Characteristics (Fengcheng)

	N	mean	sd	min	max
Female	43	1	0	1	1
Married	43	0.977	0.152	0	1
Live in the factory	43	0.279	0.454	0	1
Commute by factory bus	43	0.698	0.465	0	1
Commute by bike	43	0.047	0.213	0	1
Commute by motorbike	43	0.140	0.351	0	1
Number of years worked in the factory	43	2.930	2.005	0	7
Number of different types of products worked per day	43	1.919	0.288	1.630	2.439
Number of different products worked per day	43	2.284	0.374	1.917	3.030
Number of temporary coworkers from other lines	43	1.416	1.094	0	3.041
Education level:					
Illiterate	42	0.214	0.415	0	1
Primary school	42	0.405	0.497	0	1
Secondary school	42	0.333	0.477	0	1
High school	42	0.048	0.216	0	1

Table C2: Summary Statistics for Other Individual Characteristics (Fuzhou)

	N	mean	sd	min	max
Female	27	1	0	1	1
Married	27	1	0	1	1
Live in the factory	27	0	0	0	0
Commute by factory bus	24	0.375	0.495	0	1
Commute by bike	24	0.125	0.338	0	1
Commute by motorbike	24	0.500	0.511	0	1
Number of years worked in the factory	27	8.111	3.105	1	13
Number of different types of products worked per day	27	1.024	0.016	1.01	1.049
Number of different products worked per day	27	1.047	0.028	1.01	1.086
Number of temporary coworkers from other lines	27	0	0	0	0
Education level:					
Illiterate	27	0.037	0.192	0	1
Primary school	27	0.333	0.480	0	1
Secondary school	27	0.593	0.501	0	1
High school	27	0.037	0.192	0	1

Table C3: The Parallel Trend Assumption Test for Worker's Productivity

	DiD
Fengcheng × Weeks 1-2	-0.001
	(0.014)
Fengcheng × Weeks 3	0
	(0)
Fengcheng × Weeks 4-5	0.042**
	(0.011)
Fengcheng × Weeks 6-7	0.052**
	(0.014)
Fengcheng × Weeks 8-9	0.062**
	(0.014)
Fengcheng × Weeks 10-11	0.066**
	(0.016)
Fengcheng × Weeks 12-13	0.104**
	(0.018)
Fengcheng × Weeks 14-15	0.058**
	(0.018)
Observations	5,655
Clusters	57
R^2	0.782
Controls	YES

Notes: The unit of observation is worker i. The dependent variable is the log of worker's productivity. Productivity is a measure of the output per hour. Worker fixed effects, factory-line fixed effects, week fixed effects, the day of the week fixed effects (e.g. Monday), and an indicator variable for whether the worker is recorded sick or if there is an organisational error are included in all regressions. Robust standard errors clustered at the worker level are reported in brackets below the estimates. *** Significant at 1% level, ** significant at 5% level, * significant at 10% level.

Table C4: The Parallel Trend Assumption Test for Worker's Production Output

	DiD
Fengcheng × Weeks 1-2	-0.015
	(0.015)
Fengcheng × Weeks 3	0
	(0)
Fengcheng × Weeks 4-5	0.066**
	(0.015)
Fengcheng × Weeks 6-7	0.123**
	(0.020)
Fengcheng × Weeks 8-9	-0.006
	(0.018)
Fengcheng × Weeks 10-11	0.216**
	(0.020)
Fengcheng × Weeks 12-13	0.050*
	(0.023)
Fengcheng × Weeks 14-15	0.077**
	(0.021)
Observations	5,655
Clusters	57
R^2	0.538
Controls	YES

Notes: The unit of observation is worker i. The dependent variable is the log of worker's production output. Worker fixed effects, factory-line fixed effects, week fixed effects, the day of the week fixed effects (e.g. Monday), and an indicator variable for whether the worker is recorded sick or if there is an organisational error are included in all regressions. Robust standard errors clustered at the worker level are reported in brackets below the estimates. *** Significant at 1% level, ** significant at 5% level, * significant at 10% level.

Table C5: The Parallel Trend Assumption Test for Forewoman's Productivity

	DiD
Fengcheng × Weeks 1-2	0.001
	(0.024)
Fengcheng × Weeks 3	0
	(0)
Fengcheng × Weeks 4-5	0.023
	(0.018)
Fengcheng × Weeks 6-7	0.022
	(0.022)
Fengcheng × Weeks 8-9	0.041**
	(0.013)
Fengcheng × Weeks 10-11	0.035*
	(0.019)
Fengcheng × Weeks 12-13	0.072***
	(0.019)
Fengcheng × Weeks 14-15	0.021
	(0.020)
Observations	1,312
Clusters	13
R^2	0.846
Controls	YES

Notes: The unit of observation is forewoman i. The dependent variable is the log of forewoman's productivity. Productivity is a measure of the output per hour. Forewoman fixed effects, factory-line fixed effects, week fixed effects, the day of the week fixed effects (e.g. Monday), and an indicator variable for whether the forewoman is recorded sick or if there is an organisational error are included in all regressions. Robust standard errors clustered at the forewoman level are reported in brackets below the estimates. *** Significant at 1% level, ** significant at 5% level, * significant at 10% level.

Table C6: The Parallel Trend Assumption Test for Forewoman's Production Output

	DiD
Fengcheng × Weeks 1-2	-0.045
	(0.029)
Fengcheng × Weeks 3	0
	(0)
Fengcheng × Weeks 4-5	0.044
	(0.038)
Fengcheng × Weeks 6-7	0.121***
	(0.035)
Fengcheng × Weeks 8-9	-0.033
	(0.040)
Fengcheng × Weeks 10-11	0.195***
	(0.048)
Fengcheng × Weeks 12-13	-0.016
	(0.061)
Fengcheng × Weeks 14-15	0.051
	(0.037)
Observations	1,312
Clusters	13
R^2	0.361
Controls	YES

Notes: The unit of observation is forewoman i. The dependent variable is the log of forewoman's production output. Forewoman fixed effects, factory-line fixed effects, week fixed effects, the day of the week fixed effects (e.g. Monday), and an indicator variable for whether the forewoman is recorded sick or if there is an organisational error are included in all regressions. Robust standard errors clustered at the forewoman level are reported in brackets below the estimates. *** Significant at 1% level, ** significant at 5% level, * significant at 10% level.

Table C7: The Treatment Effect on Worker's Working Time

	Number of Minutes Worked in a Day			
	Jul-Sep	Jul	Aug	Sep
	(1)	(2)	(3)	(4)
Fengcheng	-71.597	57.206***	99.724***	109.510***
	(46.606)	(5.448)	(28.232)	(20.904)
Post	-36.565***	-35.215***	-47.421***	-24.140**
	(10.198)	(9.938)	(10.760)	(10.634)
Fengcheng*Post	11.333	35.951***	12.837	-2.043
	(9.391)	(10.299)	(10.060)	(11.370)
Observations	5,655	2,770	2,647	2,712
Clusters	57	57	57	57
R^2	0.422	0.577	0.411	0.507
Controls	YES	YES	YES	YES

Notes: The unit of observation is worker i. The dependent variables in Columns 1-4 are the working time (number of minutes) a worker worked in a day. Columns 1 shows the results for the full sample includes observations from June 7th until September 30th while Columns 2-4 compare the observations from the pre-treatment period (June) to each post-treatment month separately. Worker fixed effects, factory-line fixed effects, week fixed effects, the day of the week fixed effects (e.g. Monday), and an indicator variable for whether the worker is recorded sick or if there is an organisational error are included in all regressions. Robust standard errors clustered at the worker level are reported in brackets below the estimates. *** Significant at 1% level, ** significant at 5% level, * significant at 10% level.

Appendix B

Appendix for Chapter 3

Proof of Fact 2

Start with the benchmark case RAM. Then e_1' and θ' does not depend on an agent's choice of e_1 and the optimal stage 1 effort (and Nash equilibrium effort) is given by:

$$e_1^{RAM} = \frac{\lambda\theta}{8}.$$

If an individual is additionally paid a piece rate of 1 for output in the first stage, $y_1 = e_1$, the Nash equilibrium effort increases to:

$$e_1^{R\&I} = \frac{\lambda\theta}{8} + \theta.$$

Under PAM e_1' increases in e_1. Suppose that strategies are strictly monotone increasing and differentiable in type.[1] Since θ has full support by assumption, so does e_1 and the positive assortative assignment satisfies $e_1' = e_1$. Moreover, since θ' is a function of e_1', anticipating the matching outcome θ' is a function of e_1. The individual optimization problem becomes thus:

$$\max_{e_1} \frac{\theta + \lambda e_1 + 2(\theta'(e_1) + \lambda e_1'(e_1))}{8} - \frac{e_1^2}{2\theta}. \tag{B.1}$$

Since the optimisation problems are the same for any two individuals of the same type θ, equilibrium strategies $e_1^*(\theta)$ will be the same and thus $\theta'(e_1') = \theta(e_1) = (e_1^*)^{-1}(\theta)$. Hence an optimal choice of e_1 satisfies

$$e_1^* = \frac{3}{8}\lambda\theta + \frac{\theta}{4}\frac{\partial\theta(e_1^*)}{\partial e_1}.$$

If $\lambda = 0$, $e_1^{PAM}(\theta) = \frac{\theta}{2}$ will solve this equation. Solving the differential equation for $\lambda > 0$ yields $e_1^{PAM}(\theta) = \frac{3\lambda+\sqrt{9\lambda^2+64}}{16}\theta$, however.

[1] While strict monotonicity will be guaranteed when stage 1 effort decreases effort cost in stage 2, there may be a "pooling" equilibrium when there is no learning (i.e., $\lambda = 0$).

Under negative assortative matching the stage 1 effort of one's partner (weakly) decreases in own effort. Hence, strategies need not increase in type. The individual optimization problem becomes:

$$\max_{e_1} \frac{\theta + \lambda e_1 + 2(\theta'(e_1) + \lambda e_1'(e_1))}{8} - \frac{e_1^2}{2\theta}.$$

Hence, an optimal choice of e_1 satisfies

$$\frac{e_1}{\theta} = \frac{\lambda}{8} + \frac{\lambda}{4}\frac{\partial e_1'}{\partial e_1} + \frac{1}{4}\frac{\partial \theta}{\partial e_1}.$$

Note first that $e_1 < 0$ if $\frac{\partial e_1'}{\partial e_1} < 0$ and $e_1 = \lambda\theta/8$ if $\frac{\partial e_1'}{\partial e_1} = 0$. That is, a positive measure of agents will choose $e_1^{NAM} = 0$, i.e., there is bunching. On the other hand, agents matched to $e_1 = 0$ agents will choose $e_1 = \lambda\theta/8$, since increasing e_1 will still yield a match with $e_1 = 0$ and the same expected type θ' (supposing uniform rationing of $e_1 = 0$ agents). Hence, under NAM an equilibrium is

$$e_1^{NAM} = 0 \text{ if } \theta < \theta^* \text{ and } e_1^{NAM} = \lambda\theta/8 \text{ if } \theta > \theta^*,$$

where θ^* is a cutoff type who is just indifferent between investing $e_1 = \lambda\theta/8$ and investing $e_1 = 0$. The intuition is that investing in the first stage, although profitable in isolation, is made unprofitable, as investment is punished by obtaining a worse match in expectation (both in terms of e_1 and θ).

Notes for Proposition 3.1

For Proposition 3.1 note that while $e_1^{PAM} > e_1^{RAM}$, $e_1^{PAM} > e_1^{R\&I}$ only for λ sufficiently high, and $e_1^{PAM} < e_1^{R\&I}$ otherwise. Moreover, the ratios e_1^{PAM}/e_1^{RAM} and $e_1^{R\&I}/e_1^{RAM}$ are both strictly decreasing in λ. Finally, the ratio e_1^{NAM}/e_1^{RAM} is either 0 or 1 depending on the type θ, so that the ratio of aggregate effort investment must be less than unity.

B.1 Variable Definitions

Accuracy Of Beliefs About Relative Performance: qualitative response to the question "How do you think your individual score ranks among the other participants?" Participants could choose between the following options: "Bottom 25%", "Between 25% and 50%", "Between 50% and 75%", and "Top 25%." In our analysis, this variable is redefined into a dummy variable which equals to 0 if individuals did not manage to predict their relative standings more than once and equals to 1 if successfully predicted their relative standings at least twice.

Time Discounting: we elicited subjects' time discounting preferences using simple hypothetical choices, similar to Falk et al. (2016). Subjects in our experiment were shown a table with 11 rows. In each row they had to decide whether they preferred an early payment "today" (100 pounds) or paying a varying delayed payment "in 12 months" (100 / 103 / 106 / 109 / 112 / 115 / 118 / 121 / 124 / 127 / 130 pounds). In our analysis, subjects who accepted to receive more than 115 pounds in 12 months (the mean of overall amounts offered) are regrouped as "impatient", and for the subjects who accepted to receive 115 pounds and lower are regrouped as "patient". However, for those who misunderstood the question (either switched preferences more than once or chose to receive payment today against high payments in 12 months while chose low payments in 12 months against receiving payment today) are recategorised into the third group - "misunderstand".

Risk Attitude: we elicited subjects' risk preferences using simple lottery choices as used in Falk et al. (2016). Subjects in our experiment were shown a table with 9 rows. In each row, they had to decide whether they preferred a safe option or playing a lottery. In the lottery, they could receive either 10 pounds or 6 pounds with 50 percent probability. The lottery was the same in each row, but the safe option decreased from row to row. In the first row, the safe option was 10 pounds; in the second it was 9.5 pounds, and so on down to 6 pounds in row 9. Similar to the changes in time discounting, the cutting (re-grouping) point is set at the mean of all certain pay offers (which is paying 8 pounds for certain against the lottery). Therefore, 0 indicates the subjects are risk lovers while 1 means risk averse and 2 identifies those who misunderstood the question.

Competitiveness: we used a simple hypothetical choice question to elicit subjects' competitive preferences. Subjects were asked the choices between a tournament payment (16 pounds per score if the score is the highest, otherwise 0) and a piece-rate payment (1 pound per score).

Altruism: To elicit information about subjects' altruistic preferences, we first asked them how much of a prize (100 pounds) he/she would like to share with the other participants if he/she was the lucky winner. Subjects could choose any amount between 0 and 100. In an alternative way, namely by asking individuals to indicate their willingness to share with others without expecting anything in return when it comes to charity on an 11-point scale, with zero indicating completely unwilling to share, and ten indicating complete willingness to share. We use the same wording of the question as in Falk et al. (2016). For altruism, we introduce the product of the two indicators and categorise it into three groups. The first group has the value of 0 implies that the subject is completely unwilling to share. The second group shares the values between 0 and 250 including 250 (where 250 is given by the product of the medians of the two indicators). This group indicates subject's willingness to share is either equal or below the median. Finally, the last group includes all subjects valuing more than 250 which implies these subjects are strongly willing to share.

B.2 Other Tables

Table B1: Summary of Individual Work Stage Effort

Effort in Individual Work Stage	Observations	Mean	SD	Minimum	Maximum
Panel 0. All Treatments					
All Tasks	575	22.98	7.449	0	39
Slider Task	191	24.62	9.178	0	39
Grid Task	192	24.02	7.027	0	37
Word Encryption Task	192	20.32	4.761	0	31
Panel 1. RAM					
All Tasks	144	23.58	7.123	0	39
Slider Task	48	25.23	9.911	0	39
Grid Task	48	25.38	5.060	16	37
Word Encryption Task	48	20.15	3.525	13	27
Panel 2. NAM					
All Tasks	144	18.93	9.515	0	39
Slider Task	48	19.77	11.40	0	39
Grid Task	48	19.56	9.700	0	35
Word Encryption Task	48	17.46	6.934	0	30
Panel 3. PAM					
All Tasks	144	24.70	5.878	0	39
Slider Task	48	26.60	7.454	0	39
Grid Task	48	25.60	5.127	15	37
Word Encryption Task	48	21.90	3.270	13	31
Panel 4. R&I					
All Tasks	143	24.73	4.910	5	38
Slider Task	47	26.94	4.843	16	38
Grid Task	48	25.54	5.251	5	33
Word Encryption Task	48	21.77	2.800	14	27

Table B2: Statistical Differences Across Treatments

Effort in	NAM vs RAM		NAM vs PAM		NAM vs R&I	
Individual	t test	M-W test	t test	M-W test	t test	M-W test
Work Stage	(p-value)	(p-value)	(p-value)	(p-value)	(p-value)	(p-value)
All tasks	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
Slider	0.014	0.003	< 0.001	0.002	< 0.001	0.001
Grid	< 0.001	0.005	< 0.001	0.001	< 0.001	< 0.001
Encryption	0.019	0.092	< 0.001	< 0.001	< 0.001	< 0.001

Effort in	RAM vs PAM		RAM vs R&I		PAM vs R&I	
Individual	t test	M-W test	t test	M-W test	t test	M-W test
Work Stage	(p-value)	(p-value)	(p-value)	(p-value)	(p-value)	(p-value)
All tasks	0.147	0.207	0.112	0.179	0.959	0.841
Slider	0.444	0.956	0.291	0.528	0.798	0.687
Grid	0.826	0.797	0.874	0.488	0.953	0.564
Encryption	0.013	0.018	0.014	0.016	0.841	0.947

Notes: The null hypothesis for t-test/Mann-Whitney U (M-W) test is that the difference between the means/distributions of the two independent samples is zero.

Table B3: OLS Regression

	Dep. Var.: Effort in the Individual Work Stage					
	(1)	(2)	(3)	(4)	(5)	(6)
NAM	-4.653***	-4.653***	-4.612***	-4.636***	-4.678***	-4.708***
	(0.827)	(0.830)	(0.579)	(0.579)	(0.601)	(0.592)
	[1.313]	[1.317]	[1.260]	[1.249]	[1.278]	[1.262]
PAM	1.118	1.118	1.213*	0.974	1.185*	0.997*
	(0.699)	(0.701)	(0.577)	(0.583)	(0.589)	(0.508)
	[0.885]	[0.889]	[0.821]	[0.793]	[0.831]	[0.795]
R&I	1.151***	1.168***	1.265**	1.127**	1.291**	1.003*
	(0.353)	(0.364)	(0.481)	(0.467)	(0.468)	(0.519)
	[0.849]	[0.851]	[0.843]	[0.822]	[0.841]	[0.840]
Constant	23.58***	26.04***	27.54***	27.75***	27.28***	27.57***
	(0.242)	(0.698)	(1.027)	(1.053)	(0.981)	(0.749)
	[0.683]	[0.910]	[1.177]	[1.167]	[1.191]	[1.362]
Observations	575	575	575	575	575	575
Participants	192	192	192	192	192	192
R-squared	0.103	0.177	0.226	0.229	0.227	0.242
Task and Round Fixed Effects:	NO	YES	YES	YES	YES	YES
Attitudes	NO	NO	YES	YES	YES	YES
Academic Level	NO	NO	NO	YES	NO	YES
Gender	NO	NO	NO	NO	YES	YES
Other	NO	NO	NO	NO	NO	YES

Notes: OLS Estimations. Dependent variable is the effort in the individual work stage. The omitted treatment is RAM. Robust standard errors clustered at session level and individual level are reported in brackets and square brackets below the estimates, respectively. (1) reports estimates for the baseline model without control variables. (2) adds task and round fixed effects. (3) adds elicited preferences (accuracy of beliefs about relative performance, competitiveness, time discounting, risk averse, and altruism). (4) adds academic level dummies. (5) adds gender dummy. (6) controls for all individual demographics (gender, academic level, nationality, and degree). *** Significant at 1% level, ** significant at 5% level, * significant at 10% level.

Table B4: Descriptive Statistics of Other Variables

	Participants (1)	Mean (2)	SD (3)	Minimum (4)	Maximum (5)	Fractions (%) (6)
Played Slider Task Before	192	0	0	0	0	
Played Grid Task Before	192	0.016	0.124	0	1	
Played Word Encryption Task Before	192	0.037	0.188	0	1	
Accurately Predicted Relative Standings At Least Twice	192	0.426	0.495	0	1	
Competitive	192	0.130	0.337	0	1	
Patient	182	0.450	0.498	0	1	
Risk Averse	186	0.838	0.368	0	1	
Female	192	0.541	0.499	0	1	
Degree is Econ-related	192	0.405	0.491	0	1	
Final Earning	192	14.75	2.301	6.40	21.60	
Accuracy of Beliefs about Relative Performance	192			1	3	
1 = Accurate						44.52
2 = Underestimate						19.65
3 = Overestimate						35.83
Altruism	192			0	2	
0 = Completely Unwilling to Share						49.39
1 = Willing to Share (Below Average)						31.30
2= Willing to Share (Above Average)						19.30
Nationality:						
1 = UK						44.79
2 = EEA						13.02
3 = Others						40.62
4 = Prefer Not to Say						1.56
Native Speaking Language is English:						
1 = Yes						48.96
0 = No						49.48
2= Prefer Not to Say						1.56
Academic Level:						
1 = Undergraduate						79.69
2 = Postgraduate						19.27
3 = Prefer Not to Say						1.04
Years of Study:						
0 = Less Than 1 Year						60.42
1 = 1 Year						9.38
2 = 2 Years						12.50
3 = 3 Years						13.02
4 = More Than 3 Years						4.17
5 = Prefer Not to Say						0.52

Table B5: Spearman's Rank Correlation Coefficients of Stage 1 and Stage 2 performances across Treatments

	NAM	RAM	PAM	R&I
All tasks	0.368	0.865	0.774	0.797
Slider	0.303	0.731	0.634	0.548
Grid	0.284	0.860	0.799	0.832
Word Encryption	0.292	0.883	0.770	0.708

Table B6: Summary of Team Work Stage Effort

Effort in the Team Work Stage	Observations	Mean	SD	Minimum	Maximum
Panel 0. All Treatments					
All Tasks	575	25.48	5.981	5	45
Slider Task	191	28.70	6.648	5	45
Grid Task	192	26.00	5.185	5	38
Word Encryption Task	192	21.75	3.468	13	30
Panel 1. RAM					
All Tasks	144	25.31	6.338	10	45
Slider Task	48	28.79	7.377	10	45
Grid Task	48	25.88	4.858	15	36
Word Encryption Task	48	21.25	3.829	13	28
Panel 2. NAM					
All Tasks	144	25.12	5.844	14	43
Slider Task	48	28.04	6.633	14	43
Grid Task	48	25.48	5.165	17	38
Word Encryption Task	48	21.85	3.673	14	30
Panel 3. PAM					
All Tasks	144	25.82	6.186	5	43
Slider Task	48	29.10	7.051	5	43
Grid Task	48	26.38	5.354	16	38
Word Encryption Task	48	21.98	3.411	14	30
Panel 4. R&I					
All Tasks	143	25.65	5.560	5	43
Slider Task	47	28.85	5.525	20	43
Grid Task	48	26.27	5.457	5	38
Word Encryption Task	48	21.90	2.955	15	28

Table B7: Tests of Sample Balance on Demographies

	RAM (%)	NAM (%)	PAM (%)	R&I (%)	Chi-square test (p-value)	t-test (p-value)	M-W test (p-value)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Gender:					0.536		
Male	26.14	20.45	25.00	28.41			
	{47.92}	{37.50}	{45.83}	{52.08}			
Female	24.04	28.85	25.00	22.12			
	{52.08}	{62.50}	{54.17}	{47.92}			
Degree:					0.968		
Econ-related	23.68	25.44	25.44	25.44			
	{43.75}	{39.58}	{39.58}	{39.58}			
Not Econ-related	26.92	24.36	24.36	24.36			
	{56.25}	{60.42}	{60.42}	{60.42}			
Nationality:					0.239		
UK	23.26	20.93	23.26	32.56			
	{41.67}	{37.50}	{41.67}	{58.33}			
EEA	16.00	32.00	28.00	24.00			
	{8.33}	{16.67}	{14.58}	{12.50}			
Others	30.77	28.21	24.36	16.67			
	{50.00}	{45.83}	{39.58}	{27.08}			
Prefer Not to Say	0.00	0.00	66.67	33.33			
Native Speaking Language is English:					0.114		
Yes	26.32	21.05	21.05	31.58			
	{52.08}	{41.67}	{41.67}	{62.50}			
No	24.47	29.79	27.66	18.09			
	{47.92}	{58.33}	{54.17}	{35.42}			
Prefer Not to Say	0.00	0.00	66.67	33.33			
Academic Level:					0.031		
Undergraduate	22.88	22.88	27.45	26.80			
	{72.92}	{72.92}	{87.50}	{85.42}			
Postgraduate	35.14	35.14	10.81	18.92			
	{27.08}	{27.08}	{8.33}	{14.58}			
Prefer Not to Say	0.00	0.00	100.00	0.00			
Years of Study:					0.233		
Less Than 1 Year	25.86	28.45	22.41	23.28			
	{62.50}	{68.75}	{54.17}	{56.25 }			
1 Year	22.22	33.33	33.33	11.11			
	{8.33}	{12.50}	{12.50}	{4.17}			
2 Years	16.67	8.33	41.67	33.33			
	{8.33}	{4.17}	{20.83}	{16.67}			
3 Years	24.00	24.00	16.00	36.00			
	{12.50}	{12.50}	{8.33}	{18.75}			
More Than 3 Years	50.00	12.50	12.50	25.00			
	{8.33}	{2.08}	{2.08}	{4.17}			
Prefer Not to Say	0.00	0.00	100.00	0.00			
Age:							
RAM vs NAM						0.013	0.313
RAM vs PAM						0.885	0.000
RAM vs R&I						0.097	0.070
NAM vs PAM						0.097	0.007
NAM vs R&I						0.885	0.474
PAM vs R&I						0.013	0.051

Notes: The null hypothesis for t-test/Mann-Whitney U (M-W) test is that the difference between the means/distributions of the two independent samples is zero. The Chi-square test is used to check if there is a relationship between the demographical variables and treatments. Notice that curly bracket indicates the fraction of the corresponding group within that treatment.

Table B8: Individuals Who Predicted Their Relative Standings At Least Twice

	Observations	Fraction (%)
Treatment:	245	42.61
RAM		22.0
NAM		31.8
PAM		26.9
R&I		19.2
Female	245	43.7
Studied More Than 1 Year	242	37.2
Speak English Natively	239	45.2
From UK	240	41.2
Postgraduate	239	17.6
Degree is Econ-related	245	35.1
Competitive	245	15.9
Patient	245	38.8
Risk Averse	245	81.6

Table B9: Subgroup Analysis by Accuracy of Beliefs about Relative Performance: OLS Regression

	Less Accurate Group[2]		More Accurate Group[3]	
	(1)	(2)	(3)	(4)
NAM	-1.689***	-1.552**	-6.937***	-6.244***
	(0.445)	(0.559)	(1.997)	(1.417)
PAM	0.350	0.344	2.158	2.018
	(0.297)	(0.337)	(1.816)	(1.754)
R&I	-0.262	0.0478	3.934***	4.259***
	(0.922)	(0.939)	(1.223)	(1.307)
Constant	23.86***	26.55***	23.13***	26.78***
	(0.284)	(0.998)	(1.051)	(1.887)
Observations	330	330	245	245
Participants	110	110	82	82
R-squared	0.018	0.248	0.208	0.352
Task and Round Fixed Effects	NO	YES	NO	YES
Other Controls	NO	YES	NO	YES

Notes: OLS Estimations. Dependent variable is the effort in the individual work stage. The omitted treatment is RAM. Robust standard errors clustered at session level are reported in brackets below the estimates. Columns (1) and (3) report estimates for the baseline model without control variables. Columns (2) and (4) add task and round fixed effects, elicited preferences, and individual demographics. *** Significant at 1% level, ** significant at 5% level, * significant at 10% level.

Table B10: Individuals Who Belong to High Productivity Group

	Observations	Fraction (%)
Treatment:	273	47.48
RAM		25.27
NAM		26.01
PAM		24.18
R&I		24.54
Female	273	54.9
Studied More Than 1 Year	272	39.0
Speak English Natively	268	51.9
From UK	271	48.0
Postgraduate	270	17.4
Degree is Econ-related	273	38.1
Competitive	273	16.5
Patient	273	42.9
Risk Averse	273	84.2

Table B11: Subgroup Analysis by Team Stage Performance: OLS Regression

	Low Productivity Group		High Productivity Group	
	(1)	(2)	(3)	(4)
NAM	-2.973***	-3.132***	-6.650***	-5.934***
	(0.490)	(0.524)	(1.382)	(1.445)
PAM	2.642***	3.020***	-0.246	-0.370
	(0.678)	(0.534)	(1.072)	(1.011)
R&I	2.803***	2.347***	-0.505*	0.386
	(0.690)	(0.734)	(0.281)	(0.531)
Constant	18.99***	19.41***	28.58***	31.53***
	(0.409)	(1.207)	(0.224)	(0.895)
Observations	302	302	273	273
Participants	101	101	91	91
R-squared	0.162	0.237	0.148	0.263
Task and Round Fixed Effects	NO	YES	NO	YES
Other Controls	NO	YES	NO	YES

Notes: OLS Estimations. Dependent variable is effort in the individual work stage. The omitted treatment is RAM. Robust standard errors clustered at session level are reported in brackets below the estimates. Columns (1) and (3) report estimates for the baseline model without control variables. Columns (2) and (4) add task and round fixed effects, elicited preferences, and individual demographics. *** Significant at 1% level, ** significant at 5% level, * significant at 10% level.

Table B12: Peer Effects: OLS Regression

	Dep. Var.: Effort in The Team Work Stage			
	(1)	(2)	(3)	(4)
Matched With a More Productive Partner	-2.250	-1.943		
	(1.327)	(1.501)		
Matched With a Less Productive Partner	1.811	1.946		
	(1.410)	(1.562)		
Own Effort (Individual Work Stage)			0.693***	0.690***
			(0.101)	(0.0995)
Partner's Effort (Individual Work Stage)			-0.0333	-0.0297
			(0.0190)	(0.0180)
R&I	0.333	0.278	-0.411	-0.420
	(0.573)	(0.433)	(0.348)	(0.264)
Constant	28.88***	28.51***	11.21**	10.12**
	(2.075)	(2.462)	(3.400)	(2.933)
Observations	287	287	287	287
R-squared	0.360	0.407	0.677	0.698
Task and Round Fixed Effects	YES	YES	YES	YES
Other controls	NO	YES	NO	YES

Notes: OLS Estimations. Dependent variable is the effort in the team work stage. The omitted treatment is RAM. Robust standard errors clustered at session level are reported in brackets below the estimates. Notice that using robust standard errors clustered at individual level will slightly improve the statistical significance of the coefficients for the two dummies (matched with a more productive partner and matched with a less productive partner) in columns (1) and (2), but not change our implications in columns (3) and (4). Columns (1) and (3) report estimates with task and round fixed effects. Columns (2) and (4) further add for all other individual characteristics. *** Significant at 1% level, ** significant at 5% level, * significant at 10% level.

B.3 Experimental Instructions

Instructions [All Treatments]

Thank you for participating in this session. Please raise your hand if you want to ask a question. Apart from asking questions in this way, you must not communicate with anybody in this room. Please now turn off mobile phones and any other electronic devices. These must remain turned off for the duration of this session.

You have been allocated to a computer booth according to the number on the card we gave you as you came in. You must not look into any of the other computer booths at any time during this session. To ensure anonymity, your actions in this session are also linked to this number. From now on, please keep it safe as this card will be required for payment at the end.

You will be paid a show up fee of £4, plus any earnings you accumulate during this session. The amount of money you accumulate will depend partly on your actions, partly on the actions of others and partly on chance. All payments will be made in cash. None of the other participants will see how much you have been paid.

The Setup [All Treatments]

This session consists of three rounds in which you will work on three different tasks. You will perform only one of the tasks in each round and for each task you will get a score based on your performance. The order in which you will perform each task is random.

Each round is divided into three stages: a practice stage, an individual work stage, and a team work stage. The practice stage lasts for 2 minutes and allows you to familiarise yourself with the tasks. Both work stages, individual and team work, last for 4 minutes. Your performance in the individual work stage will be ranked against all other participants. The computer will assign to you another participant as a partner for the team work stage according to a rule explained below **[RAM and R&I]** (Based on this ranking the computer will assign to you another participant as a partner for the team work stage according to a rule explained below **[NAM and PAM]**).

Further details of the payment, the pairing rule and the tasks will be explained below.

Payment [RAM, NAM, and PAM]

In each round your team performance at the team work stage will affect your earnings. In particular, for your team work you earn CREDITS. Your CREDITS are given by the average score of your team.

For example, if player A's score is 38 and player B's score is 28 in the team work stage, each of them earns $\frac{38+28}{2} = 33$ CREDITS.

At the end of the experiment the computer will randomly choose one out of the three rounds to determine your earnings. In other words, all rounds (or tasks) are equally important to you regarding the payment. The CREDITS that you earned from the selected round will determine your payment from performing the tasks: the CREDITS will be exchanged into pounds and the exchange rate will be: 1 CREDIT = £0.40.

As an example, suppose that in the round that is randomly chosen for payment at the end you earned 38 CREDITS. Then your total earnings from performing the tasks will be as follows:

$$\text{Total Earnings} = 38 * \pounds 0.40 = \pounds 15.20$$

Payment [R&I]

In each round your performance will influence your earnings. In particular, for your work you earn CREDITS. In the individual work stage your CREDITS are equal to your score. In the team work stage your CREDITS are given by the average score of your team.

For example, if player A's score is 30 in the individual work stage, player A earns 30 CREDITS. If player A is working in a team with player B in the team work stage, player A's score is 38 and player B's score is 28, each of them earns $\frac{38+28}{2} = 33$ CREDITS.

At the end of the experiment the computer will randomly choose one round (out of the three rounds) and one stage (out of individual work stage and team work stage) to determine your earnings. In other words, both work stages in all rounds (or tasks) are equally important to you regarding the payment. The CREDITS that you earned from the selected round and the selected stage will determine your payment from performing the tasks: the CREDITS will be exchanged into pounds and the exchange rate will be: 1 CREDIT = £0.40.

As an example, suppose that in the round that is randomly chosen for payment at the end you earned 38 CREDITS at the selected stage. Then your total earnings from performing the tasks will be as follows:

$$\text{Total Earnings} = 38 * \pounds 0.40 = \pounds 15.20$$

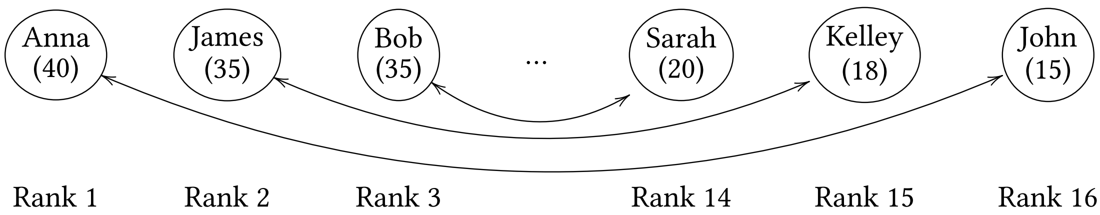
Pairing Rule [RAM and R&I]

The computer will randomly assign to you another participant as a partner for the team work stage. Each team consists of 2 partners.

Pairing Rule [NAM]

The computer will rank all participants according to their scores in the individual work stage. Each team consists of 2 partners. Teams are formed by pairing participants based on their scores in the individual work stage: the best performing participant will be working in a team with the worst performing one, the second best will be working in a team with the second worst, and so on and so forth (see the example in the figure below). If some participants share the same score their rank will be drawn randomly to avoid ties. For instance, Bob and James who have a score of 35 each, have each a chance of 50% to be assigned rank 2, respectively rank 3.

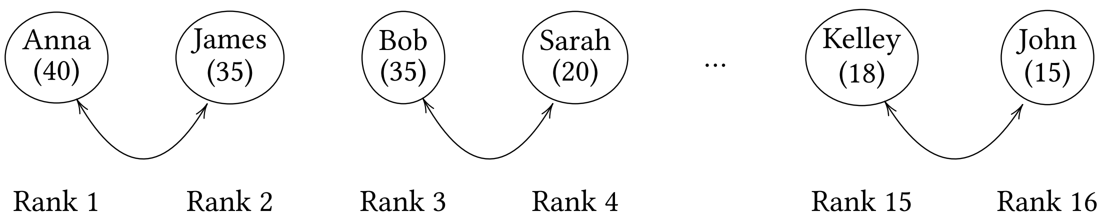
FIGURE B1: Team assignment with 16 participants (individual scores are shown in brackets).



Pairing Rule [PAM]

The computer will rank all participants according to their scores in the individual work stage. Each team consists of 2 partners. Teams are formed by pairing participants based on their scores in the individual work stage: the best performing participant will be working in a team with the second best performing one, the third will be working in a team with the fourth, and so on and so forth (see the example in the figure below). If some participants share the same score their rank will be drawn randomly to avoid ties. For instance, Bob and James who have a score of 35 each, have each a chance of 50% to be assigned rank 2, respectively rank 3.

FIGURE B2: Team assignment with 16 participants (individual scores are shown in brackets).

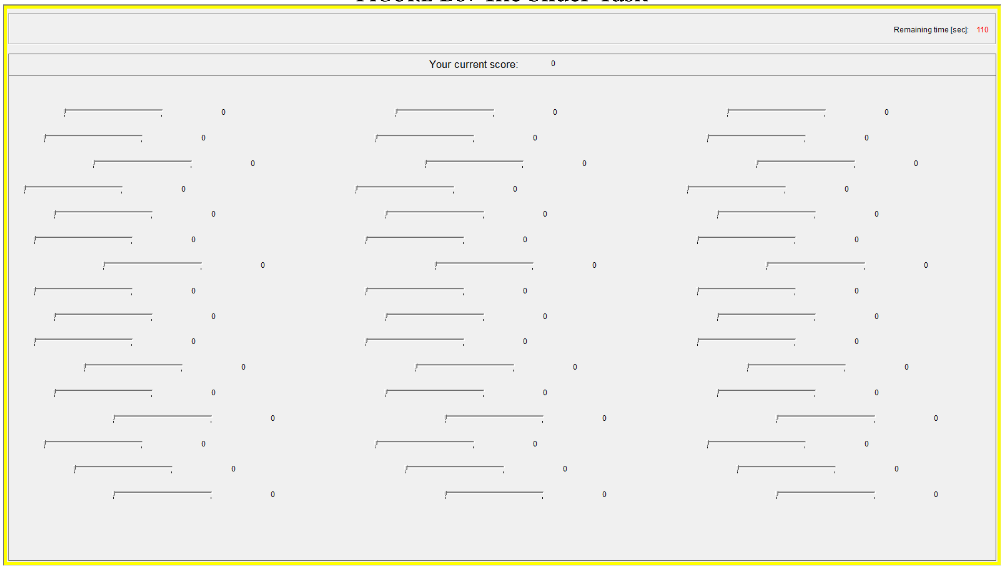


The Tasks [All Treatments]

Slider

The task will consist of a screen with 48 sliders. Each slider is initially positioned at 0 and can be moved as far as 100. Each slider has a number to its right showing its current position. You can use the mouse in any way you like to move each slider. You can re-adjust the position of each slider as many times as you wish. Your task is to position each slider at 50. Your score in the task will be the number of sliders positioned at exactly 50 within 4 minutes. The decision screen is seen in the figure below.

FIGURE B3: The Slider Task



Grid

5 by 5 grids with randomly distributed 0's and 1's will appear on the screen. Your task is to count the number of 0's. Once you count a table correctly, the computer will prompt you with another table which you will be asked to count 0's. Once you count that table, you will be given another table and so on. Your score in the task will be the number of grids with a correct count of 0's entered within 4 minutes. The decision screen is seen in the figure below.

Word Encryption

This task consists of encoding words into numbers. Each word is a combination of three letters. You have to allocate a number (0-100) to each letter. The encryption code can be found in a table below the corresponding word. Once you encode a word correctly, the computer will prompt you with another word which you will be asked to encode. Once you encode that word, you will be given another word and so on. Your score in the task will be the number of words encoded correctly within 4 minutes. As an example, the decision screen can be seen in the figure below.

Note that the encryption table during the experiment will be different from the given example. Before each stage of this task, the computer first selects in the table a new set of random numbers (0-100) to be used for the encoding of the capital letters. Then, the computer program

FIGURE B4: The Grid Task

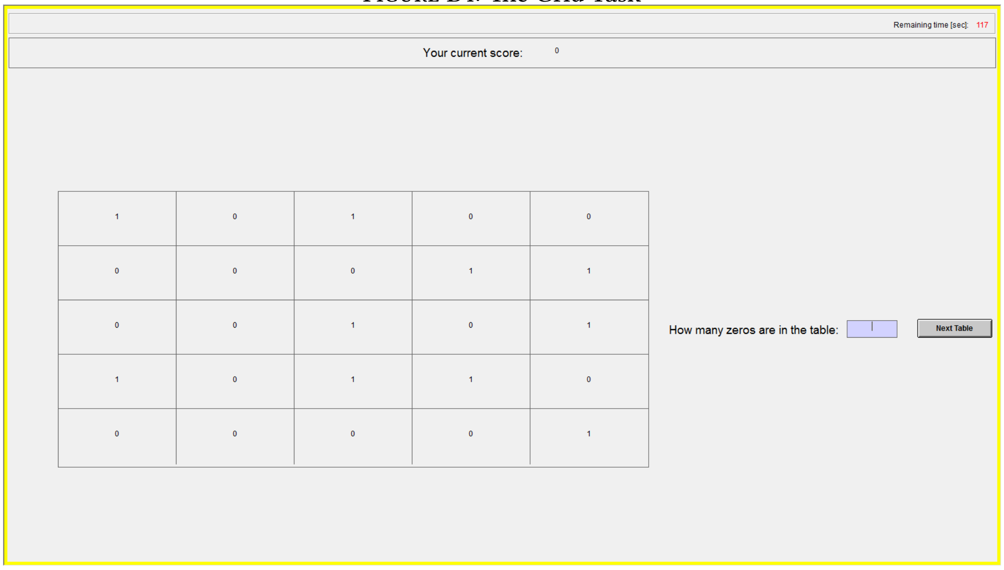


FIGURE B5: The Word Encryption Task

shuffles the position of the capital letters in the table. Note that the encryption table will differ between practice, individual, and team work stages.

Other Information [All Treatments]

During each task, some information will appear at the top of your screen, including the time remaining and your score in the task. After successfully generating all possible teams, the computer will first show you your score, your rank, the highest score and the lowest score among all participants in the individual work stage and then your partner's rank and score. At the end of the team work stage, you will see a summary screen showing your score, your partner's score, and your team's score.

At the end of the session your total cash payment, including the £4 show up fee, will be displayed on your screen. Please leave the computer booth one by one when asked to do so to receive your payment. Please leave all other material on your desk. Thank you for participating. Are there any questions?

Appendix C

Appendix for Chapter 4

Variable Definitions

Managers: number of workers belonging to the following occupational group, as defined in the WERS codebook: "Managers and senior officials": "Managers and senior officials head government, industrial, commercial and other establishments, organisations, or departments within such organisations. They determine policy, direct and coordinate functions, often through a hierarchy of subordinate managers and supervisors. Occupations included are: general managers, works managers, production managers, marketing or sales managers, directors of nursing, catering managers and bank managers. This group also includes police inspectors and senior officers in the fire, ambulance and prison services. This group does not include supervisors or foremen. These employees should be grouped within their skill base e.g. a clerical worker supervising other clerical workers would be grouped with them. A fitter and turner acting as a supervisor or foreman would be classified as a craft or skilled worker."

Recruitment preferences: qualitative response to the question "Which of these statements best describes your approach to filling vacancies at this workplace?" Answers include: 1 = Internal applicants are only source, no external recruitment; 2 = Internal applicants are given preference, other things being equal, over external applicants; 3 = Applications from internal and external applicants are treated equally; 4 = External applicants are given preference, other things being equal, over internal applicants; 5 = External applicants are only source, no internal recruitment.

Competition in the market: employer's rating of the the degree of competition in this market (1 = Very low, 2 = Low, 3 = Neither high nor low, 4 = Hight, 5 = Very high).

Quality of product/service: employer's rating of the establishment's quality of product or service relative to that of other establishments in the same industry (1 = A lot below average, 2 = Below average, 3 = Average, 4 = Above average, 5 = A lot above average).

Market share: qualitative response to the question "What is your company's UK market share for your (main) product or service? Market share is the total value of your company's goods or services as a proportion of all UK sales." Answers include: 1 = Less than 5%, 2 = 5-10%, 3 = 11-25%, 4 = 26-50%, 5 = More than 50%.

Industry categories: Manufacturing; Electricity, Gas, and Water; Construction; Wholesale and Retail; Hotels and Restaurants; Transport and Communication; Financial Services; Other Business Services; Public Administration; Education; Health; Other Community Services.

Formal status categories: Public Limited Company (PLC); Private limited company; Company limited by guarantee; Partnership (inc. Limited Liability Partnership) / Self-proprietorship; Trust / Charity; Body established by Royal Charter; Co-operative / Mutual / Friendly society; Government-owned limited company / Nationalised industry / Trading Public Corporation; Public service agency; Other non-trading public corporation; Quasi Autonomous National Government Organisation (QUANGO); Local/Central Government (inc. NHS and Local Education Authorities).

Region categories: North, Yorkshire and Humberside, East Midlands, East Anglia, South East, South West, West Midlands, North West, Wales, and Scotland.

Notice that, to remain enough degree of freedom in our analysis we reconstruct recruitment preferences and quality of products/services into dummy variables. For recruitment preferences, it equals to 0 if internal applicants are preferred (which originally equals 1 or 2), and equals 1 as long as external applicants are considered. For quality of products/services, value 1 implies above average (which originally equals 4 or 5), otherwise 0.

C.1 Is the Career Mentoring Used a Good Measure?

My analyses about mentoring are based on one single proxy which asked employees to what extent do they agree or disagree about the managers at the workplace encouraging people to develop their skills. But is it a good proxy for mentoring? In this section, I explore potential underlying mechanisms.

Previous research has demonstrated that individuals, who receive more mentoring from managers, earn higher salaries and are more likely to be satisfied with their jobs (Dreher and Cox, 1996; Fagenson, 1989; Kram, 1985; Scandura, 1992; Wallace, 2001). In the data, employees answered qualitatively questions including "the sense of achievement you get from your work," "the opportunity to develop your skills in your job," "the amount of pay you receive" and "your job security" with answers: 1 = Very satisfied; 2 = Satisfied; 3 = Neither satisfied nor dissatisfied; 4 = Dissatisfied; 5 = Very dissatisfied. Using these variables, I find consistent evidence in support of the hypothesis that Career Mentoring defined in this paper is positively associated with employees' satisfaction on achievement from work, opportunity to develop skills, the amount of pay received and job security, as shown in Figure C1.

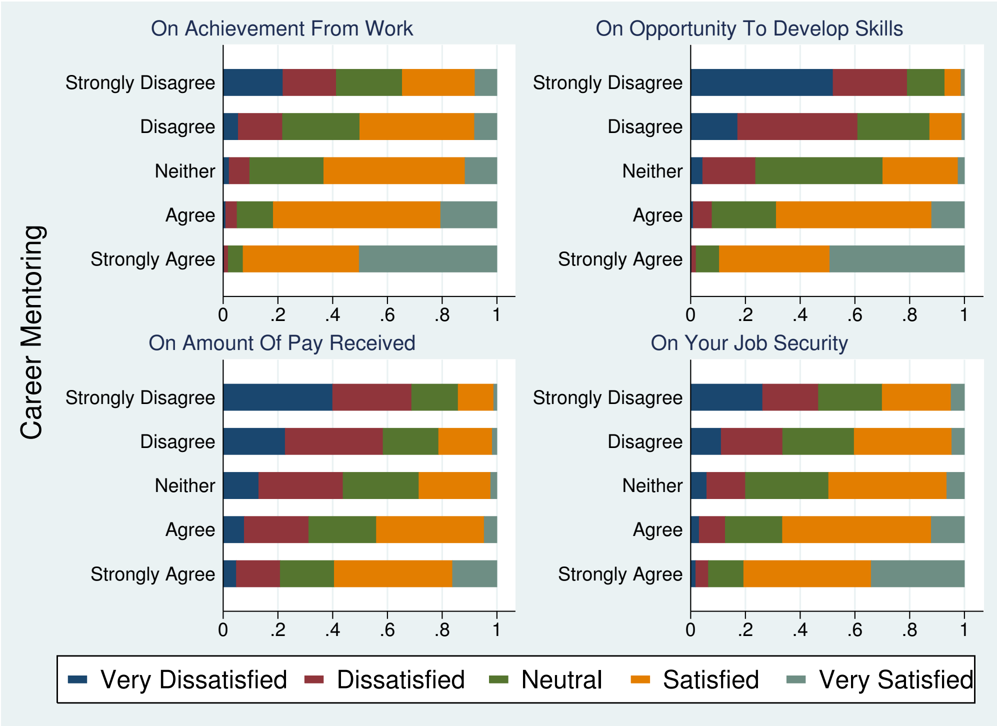
One could argue that the positive association between Career Mentoring and an employee's job satisfaction is due to the fact that individuals who agreed that managers encouraged people to develop their skills, were better paid and therefore had a higher satisfaction on their job. However, this is rejected by the data in Figure C2. Each group of employees with a given opinion about Career Mentoring was a representative in terms of salaries. Therefore, individual attitudes seem not to be driven by individual salaries, which alleviating some concerns of endogeneity.

On the other hand, mentoring also builds confidence (e.g. Eskreis-Winkler et al., 2018) and mental resilience (e.g. Gill et al., 2018) in the mentors themselves. Therefore, there are a lot of benefits for firms to induce mentoring relationships between managers and employees. An interesting question could be: does more mentoring associate with better organisational performance?

Figure C4 shows that the percentage of employees who asserted that their managers at the workplace encouraged people to develop their skills is significantly higher in better-performing firms, where better-performing firms are those whose financial performance is above average among their competitors in the industry. Similarly, average employees' salary in high-performing firms is higher than in the low-performing ones (though the difference is statistically insignificant). Hence, mentoring is indeed positively correlated with firms' performance. Similar patterns can also be found in Figures C5 and C6, where I look at firms' labour productivity and the quality of products/services. Interestingly, the mean of employees' salary is higher in low-productive firms than in high-productive ones (nevertheless, the difference is not statistically significant at a conventional level). Taken together, we find more mentoring relationships in firms that do well. This positive association is consistent with the literature (see Haggard et al., 2011, for a review).

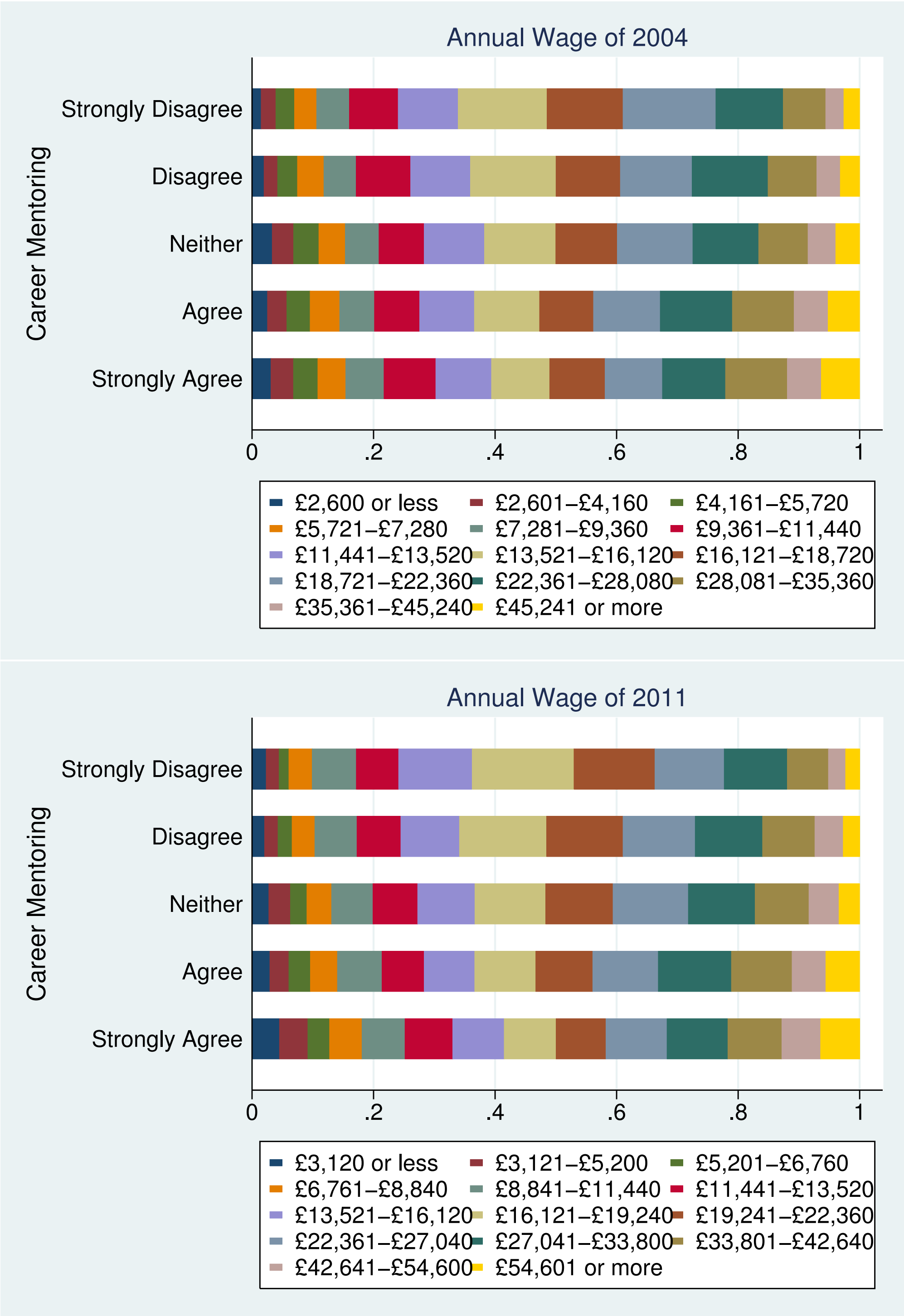
In the literature, studies measured the career mentoring support with the following scales or with somewhat modified versions of these scales: Ragins and McFarlin (1990) 33-item mentor role instrument, Dreher and Ash (1990) 18-item mentoring support scale, and Noe (1988) 29-item mentor functions questionnaire. Out of Ragins and McFarlin (1990) 33 items, three instruments are related to ours, including (1) "... assigns me tasks that push me into developing new skills"; (2) "... gives me tasks that require me to learn new skills"; (3) "... provides support and encouragement". In Dreher and Ash (1990), one of the items that is close to ours is "... given or recommended you for challenging assignments that presented opportunities to learn new skills". Finally, Noe (1988) used, for example "... provided you with support and feedback regarding your performance as an educator"; "... suggested specific strategies for achieving your career goals or accomplishing work objectives"; "... gave you assignments that present opportunities to learn new skills". Although, these survey items are not framed as exactly the same as the one I used in WERS. The idea is very similar, managers' career mentoring includes the activity to provide supports and encouragements to the mentees and help them to develop skills. Therefore, the instrument I used in this study captures the general definition of career mentoring.

FIGURE C1: Career Mentoring and Employee's Satisfactions on the Job



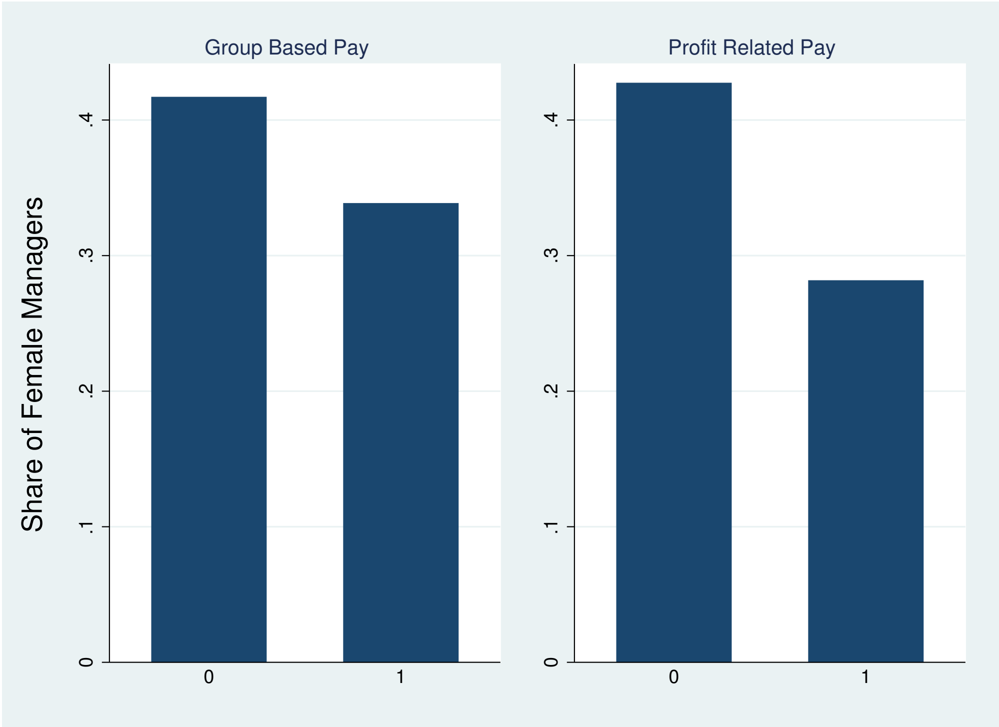
Notes: Y-axises indicate employee's qualitative response to Career Mentoring which is defined as manager at the workplace encouraging people to develop their skills. X-axises indicate the distribution of population regarding their degree of satisfaction on the selected topic (i.e., achievement from work, opportunity to develop skills, amount of pay received or job security) for each item of Career Mentoring (i.e., strongly disagree, disagree, neither, agree and strongly agree). As an example, the top left sub-graph (i.e., titled On Achievement From Work) shows the relationship between Career Mentoring and employee's satisfaction on achievement from the work. For employees who strongly disagreed with Career Mentoring, about 60% of them were not satisfied with their achievements from the work. In contrast, more than 90% of respondents who strongly agreed with Career Mentoring were satisfied with their achievement from the work. Similar interpretations can be applied on other sub-graphs.

FIGURE C2: Career Mentoring and the Employee's Annual Wage



Notes: Y-axises indicate employee's qualitative response to Career Mentoring which is defined as manager at the workplace encouraging people to develop their skills. X-axises indicate the distribution of population regarding their annual wage for each item of Career Mentoring (i.e., strongly disagree, disagree, neither, agree and strongly agree). For instance, about 60% employees who were paid higher than £16,121 in a year strongly agreed with Career Mentoring at the workplace in 2011.

Figure C3: Share of Female Managers and Managerial Incentives

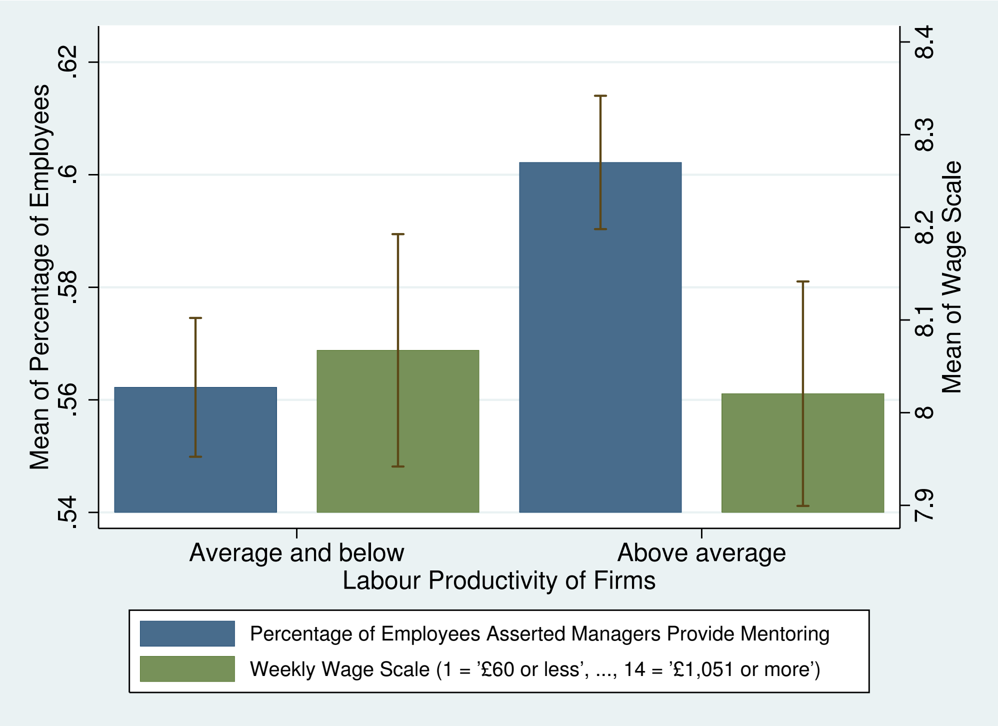


Notes: The differences in the share of female managers between the unpaid (=0) and paid (=1) subgroups are statistically significant for both managerial pay schemes (p-values of either t-test or Mann-Whitney U test are < 0.001).

Figure C4: Mentoring Occurrence and Firms' Financial Performance

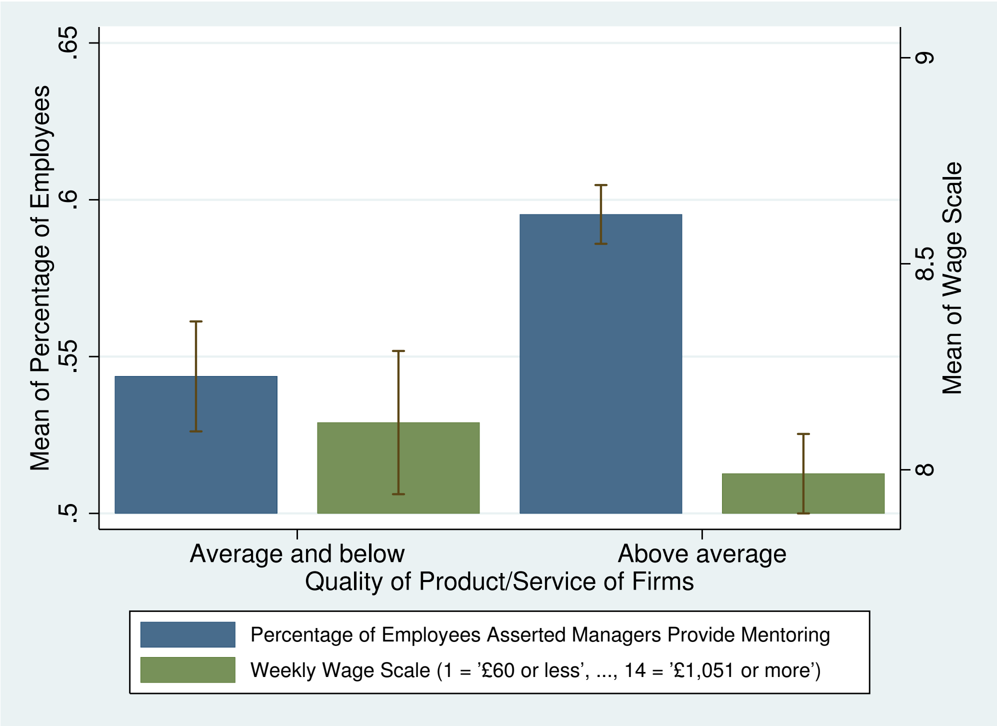
Notes: Above average indicates firms whose financial performances are above average among their competitors in the industry. The line segments represent the 95% confidence intervals.

FIGURE C5: Mentoring Occurrence and Firms' Labour Productivity



Notes: Above average indicates firms whose labour productivities are above average among their competitors in the industry. The line segments represent the 95% confidence intervals.

Figure C6: Mentoring Occurrence and Firms' Quality of Product/Service



Notes: Above average indicates firms whose product/service qualities are above average among their competitors in the industry. The line segments represent the 95% confidence intervals.

C.2 Sampling Weights

In WERS 2004-2011, sampling weights were derived for management questionnaires (MQ) and employee questionnaires (EQ) respectively. For MQ, separate weights are also available for each of the samples (i.e. cross-section and panel). The establishment sampling weight is computed as the inverse of the workplace's probability of selection from the sampling frame, adjusted to account for any observable non-response biases and then post-stratified so that the profile of the weighted sample of workplaces matches the profile of the population by workplace size and industry formal status. The panel sample weights further account for any observable non-response biases that became apparent when attempting to follow up these workplaces in 2011. These weights can be used to bring the profiles of the achieved samples of workplaces and employees into line with the profiles of the respective populations, thereby removing known biases introduced by the sample selection and response process.

For MQ, 53% of the initial sample of 4,293 cases yielded productive interviews, 16% were out-of-scope, and 30% were eligible but unproductive. This gives a response rate of 64% in 2004 compared with a response rate of 43% in 2011. For EQ, the response rate is 50% among all sampled employees in 2011 and 54% in 2004 respectively.

Notice that the EQ received a weight based on the MQ weight but adjusted for:

(i) the probability that the workplace distributed the EQ (estimated using a non-response model with predictors based data from the MQ);

(ii) employee selection probabilities;

(iii) post-stratification to the population of all employees (as reported by managers) by gender.

C.3 Other Tables

Table C1: Descriptive Statistics for Control Variables (2004)

	Cross-sectional				Longitudinal Data			
VARIABLES	Mean	Proportion	Std. Err.	Obs	Mean	Proportion	Std. Err.	Obs
Number of employees	31.423		1.112	1,733	41.416		2.657	600
More than 25 employees		0.235	0.012			0.293	0.025	
Recruitment preference				1,730				598
Internal only		0.003	0.002			0.006	0.006	
Prefer internal		0.211	0.017			0.188	0.030	
Neutral		0.646	0.021			0.697	0.034	
Prefer external		0.016	0.006			0.010	0.006	
External only		0.124	0.016			0.098	0.023	
Market competition				1,203				373
Very low		0.044	0.009			0.056	0.021	
Low		0.071	0.012			0.100	0.027	
Neither high nor low		0.160	0.018			0.189	0.033	
High		0.354	0.022			0.335	0.039	
Very high		0.370	0.023			0.321	0.039	
Quality of product/service				1,621				553
A lot below average		0.001	0.001			0.002	0.001	
Below average		0.014	0.004			0.024	0.015	
About average		0.187	0.016			0.194	0.027	
Better than average		0.546	0.021			0.559	0.036	
A lot better than average		0.253	0.019			0.221	0.030	
UK market share				996				306
Less than 5%		0.539	0.026			0.593	0.046	
5-10%		0.105	0.015			0.098	0.029	
11-25%		0.129	0.016			0.099	0.027	
26-50%		0.108	0.015			0.071	0.024	
More than 50%		0.118	0.016			0.138	0.032	
Industry				1,733				600
Manufacturing		0.102	0.013			0.098	0.021	
Electricity, gas and water		0.001	0.000			0.001	0.001	
Construction		0.047	0.010			0.042	0.015	
Wholesale and retail		0.247	0.019			0.243	0.035	
Hotels and restaurants		0.076	0.011			0.062	0.018	
Transport and communication		0.040	0.008			0.061	0.018	
Financial services		0.040	0.007			0.006	0.003	
Other business services		0.156	0.016			0.112	0.021	
Public administration		0.026	0.005			0.037	0.010	
Education		0.063	0.008			0.084	0.017	
Health		0.124	0.012			0.175	0.023	
Other community services		0.078	0.011			0.080	0.016	

Continued...

	Cross-sectional			Longitudinal Data		
VARIABLES	Proportion	Std. Err.	Obs	Proportion	Std. Err.	Obs
Region			1,733			600
North	0.048	0.008		0.045	0.012	
Yorkshire and Humberside	0.085	0.011		0.087	0.019	
East Midlands	0.070	0.011		0.048	0.013	
East Anglia	0.045	0.008		0.042	0.013	
South East	0.316	0.020		0.291	0.032	
South West	0.089	0.012		0.110	0.024	
West Midlands	0.105	0.014		0.132	0.028	
North West	0.113	0.013		0.117	0.022	
Wales	0.037	0.007		0.033	0.009	
Scotland	0.093	0.012		0.095	0.018	
Formal status			1,733			600
8 Public limited company	0.197	0.016		0.162	0.030	
Private limited company	0.449	0.021		0.367	0.035	
Company limited by guarantee	0.029	0.007		0.046	0.017	
Partnership/Self-proprietorship	0.114	0.013		0.104	0.019	
Trust/Charity	0.042	0.008		0.075	0.016	
Body established by Royal Charter	0.004	0.002		0.002	0.001	
Society	0.017	0.005		0.023	0.011	
GNT	0.013	0.004		0.052	0.018	
Public service agency	0.010	0.003		0.014	0.006	
Other non-trading public corporation	0.002	0.001		0.001	0.001	
QUANGO	0.001	0.001		0.003	0.003	
Local/Central government	0.121	0.011		0.150	0.019	

Notes: "Partnership" indicates Partnership (inc. Limited Liability Partnership). "Society" indicates Co-operative / Mutual / Friendly society. "GNT" indicates Government-owned limited company/Nationalised industry/Trading Public Corporation. "QUANGO" indicates Quasi Autonomous National Government Organisation. "Local/Central Government" indicates Local/Central Government (inc. NHS and Local Education Authorities). Establishment sampling weights are used to calculate the standard errors.

Table C2: Descriptive Statistics for Control Variables (2011)

	Cross-sectional				Longitudinal Data			
VARIABLES	Mean	Proportion	Std. Err.	Obs	Mean	Proportion	Std. Err.	Obs
Number of Employees	31.971		1.172	1,923	56.729		5.750	600
More than 25 employees		0.239	0.013			0.351	0.029	
Recruitment preference				1,915				596
Internal only		0.018	0.005			0.028	0.011	
Prefer internal		0.193	0.016			0.166	0.024	
Neutral		0.665	0.020			0.652	0.035	
Prefer external		0.011	0.004			0.013	0.008	
External only		0.113	0.015			0.141	0.030	
Market competition				1,256				384
Very low		0.030	0.008			0.034	0.011	
Low	0.067	0.012			0.079	0.024		
Neither high nor low		0.168	0.018			0.149	0.030	
High		0.398	0.023			0.386	0.040	
Very high		0.337	0.022			0.353	0.041	
Quality of Product/Service				1,841				572
A lot below average		0.000	0.000			0.000	0.000	
Below average		0.011	0.003			0.026	0.010	
About average		0.173	0.016			0.256	0.035	
Better than average		0.544	0.021			0.475	0.036	
A lot better than average		0.271	0.019			0.243	0.029	
UK market share				1,071				331
Less than 5%		0.567	0.026			0.566	0.044	
5-10%		0.104	0.018			0.058	0.017	
11-25%		0.115	0.017			0.125	0.029	
26-50%		0.099	0.016			0.073	0.017	
More than 50%		0.115	0.015			0.177	0.035	
Industry				1,921				599
Manufacturing		0.096	0.014			0.088	0.019	
Electricity, gas and water		0.001	0.000			0.001	0.001	
Construction		0.042	0.008			0.047	0.015	
Wholesale and retail		0.235	0.020			0.245	0.035	
Hotels and restaurants		0.080	0.010			0.060	0.018	
Transport and communication		0.045	0.008			0.074	0.019	
Financial services		0.007	0.003			0.006	0.003	
Other business services		0.174	0.016			0.114	0.023	
Public administration		0.028	0.005			0.040	0.010	
Education		0.078	0.008			0.082	0.015	
Health		0.137	0.013			0.159	0.022	
Other community services		0.077	0.009			0.083	0.017	

Continued...

	Cross-sectional			Longitudinal Data		
VARIABLES	Proportion	Std. Err.	Obs	Proportion	Std. Err.	Obs
Region			1,923			600
North	0.050	0.009		0.046	0.012	
Yorkshire and Humberside	0.073	0.010		0.087	0.019	
East Midlands	0.078	0.014		0.045	0.013	
East Anglia	0.042	0.011		0.040	0.013	
South East	0.324	0.020		0.295	0.032	
South West	0.111	0.014		0.110	0.024	
West Midlands	0.089	0.012		0.132	0.028	
North West	0.090	0.010		0.116	0.022	
Wales	0.043	0.007		0.033	0.009	
Scotland	0.099	0.011		0.095	0.018	
Formal status			1,923			600
Public Limited Company	0.142	0.015		0.143	0.026	
Private limited company	0.536	0.021		0.425	0.036	
Company limited by guarantee	0.018	0.005		0.027	0.014	
Partnership/Self-proprietorship	0.079	0.010		0.090	0.019	
Trust / Charity	0.068	0.009		0.069	0.015	
Body established by Royal Charter	0.001	0.000		0.002	0.001	
Society	0.012	0.005		0.025	0.011	
GNT	0.016	0.003		0.051	0.018	
Public service agency	0.006	0.002		0.010	0.005	
Other non-trading public corporation	0.001	0.001		0.005	0.003	
QUANGO	0.003	0.002		0.003	0.003	
Local/Central Government	0.117	0.011		0.150	0.019	

Notes: "Partnership" indicates Partnership (inc. Limited Liability Partnership). "Society" indicates Co-operative / Mutual / Friendly society. "GNT" indicates Government-owned limited company/Nationalised industry/Trading Public Corporation. "QUANGO" indicates Quasi Autonomous National Government Organisation. "Local/Central Government" indicates Local/Central Government (inc. NHS and Local Education Authorities). Establishment sampling weights are used to calculate the standard errors.

Table C3: Testing Selection Bias for Panel Sample

	(1)	(2)
Career Mentoring	-0.071	-0.132
	(0.075)	(0.091)
Share of Female Managers	-0.025	-0.002
	(0.040)	(0.053)
More Than 25 Employees	0.011	0.046
	(0.028)	(0.035)
PERFM		0.030
		(0.033)
PROFT		-0.052
		(0.033)
Recruitment Preference		0.022
		(0.023)
Market Competition		0.010
		(0.016)
Quality of Product/Service		-0.002
		(0.022)
UK Market Share		-0.005
		(0.012)
Region FE	YES	YES
Industry FE	YES	YES
Formal status FE	YES	YES
Observations	1563	872
R^2	0.08	0.10

Notes: OLS regression with the dependent variable equals 1 if workplaces in 2004 are observed again in 2011 and otherwise 0. Robust standard errors are clustered at establishment level. "Career Mentoring" indicates the percentage of employees who claimed that managers at the workplace encouraged people to develop their skills. "More than 25 employees" equals to 1 if the establishment had more than 25 employees and otherwise 0. "PERFM" indicates the performance related pay scheme while "PROFT" for the profit related pay scheme. Establishment sampling weights are used to calculate the standard errors.

Table C4: Descriptive Statistics for Managerial Incentives

	PERFM	PROFT
Number of firms had used the scheme in neither 2004 nor 2011	316	412
Number of firms had introduced the scheme in 2011	96	70
Number of firms had used the scheme in both 2004 and 2011	99	60
Number of firms had abolished the scheme in 2011	85	55

Notes: "PERFM" indicates the performance related pay scheme while "PROFT" for the profit related pay scheme.

Table C5: The Effect of Changing Managerial Incentives on Career Mentoring (The Share of Female Mangers is Controlled in all specifications)

	Introduce PERFM (1)	Abolish PERFM (2)	Introduce PROFT (3)	Abolish PROFT (4)
PERFM	-0.0212	0.0587*		
	(0.0301)	(0.0349)		
PERFM*2011	0.0512	-0.0633		
	(0.0348)	(0.0438)		
PROFT			-0.0169	0.00992
			(0.0365)	(0.0513)
PROFT*2011			0.0635*	-0.0483
			(0.0373)	(0.0621)
2011	-0.0626***	0.0168	-0.0562***	0.00232
	(0.0170)	(0.0274)	(0.0141)	(0.0392)
Constant	0.384***	0.490***	0.373***	0.598***
	(0.0610)	(0.0796)	(0.0544)	(0.112)
Number of firms	369	181	437	114
R-squared	0.278	0.313	0.279	0.278
Other Controls	YES	YES	YES	YES

Notes: OLS estimations based on specification 4.3. The dependent variable is Career Mentoring which indicates the percentage of employees who claimed that managers at the workplace encouraged people to develop their skills. The empirical analysis is conducted at the establishment level, and robust standard errors clustered at establishment level are reported in brackets below the estimates. PERFM equals 1 if managers were offered performance related pay scheme in the firm and otherwise 0. PROFT equals 1 if managers were offered profit related pay scheme in the firm and otherwise 0. Other controls include the Share of Female Mangers, year dummies, industry and formal status dummies, region dummies, firm size, recruitment preferences, competition condition in the market, quality of products/services, and UK market share. *** Significant at 1% level, ** significant at 5% level, * significant at 10% level.

Bibliography

Abeler, J., Falk, A., Goette, L., and Huffman, D. (2011). Reference points and effort provision. *American Economic Review*, 101(2):470–92.

Ai, C. and Norton, E. C. (2003). Interaction terms in logit and probit models. *Economics Letters*, 80(1).

Al-Ubaydli, O., Andersen, S., Gneezy, U., and List, J. A. (2015). Carrots that look like sticks: Toward an understanding of multitasking incentive schemes. *Southern Economic Journal*, 81(3):538–561.

Albanesi, S. and Olivetti, C. (2009). Home production, market production and the gender wage gap: Incentives and expectations. *Review of Economic Dynamics*, 12(1):80 – 107.

Albanesi, S., Olivetti, C., and Prados, M. J. (2015). *Gender and Dynamic Agency: Theory and Evidence on the Compensation of Top Executives*, chapter 1, pages 1–59. Emerald Group Publishing Limited.

Allen, T., Poteet, M., Eby, L., Lentz, E., and Lima, L. (2004). Career benefits associated with mentoring for protégés: A meta-analysis. *Journal of Applied Psychology*, 89(1):127–136. cited By 443.

Allen, T. D. and Eby, L. T. (2004). Factors related to mentor reports of mentoring functions provided: Gender and relational characteristics. *Sex Roles*, 50(1-2):129–139.

Apesteguia, J., Azmat, G., and Iriberri, N. (2012). The impact of gender composition on team performance and decision making: Evidence from the field. *Management Science*, 58(1):78–93.

Araujo, F. A., Carbone, E., Conell-Price, L., Dunietz, M. W., Jaroszewicz, A., Landsman, R., Lamé, D., Vesterlund, L., Wang, S. W., and Wilson, A. J. (2016). The slider task: an example of restricted inference on incentive effects. *Journal of the Economic Science Association*, 2(1):1–12.

Armstrong, S. J., Allinson, C. W., and Hayes, J. (2002). Formal mentoring systems: An examination of the effects of mentor/protégé cognitive styles on the mentoring process. *Journal of Management Studies*, 39(8):1111–1137.

Athey, S., Avery, C., and Zemsky, P. (2000). Mentoring and diversity. *American Economic Review*, 90(4):765–786.

Babcock, P., Bedard, K., Charness, G., Hartman, J., and Royer, H. (2015). Letting down the team? social effects of team incentives. *Journal of the European Economic Association*, 13(5):841–870.

Bandiera, O., Barankay, I., and Rasul, I. (2005). Social preferences and the response to incentives: Evidence from personnel data. *The Quarterly Journal of Economics*, 120(3):917–962.

Bandiera, O., Barankay, I., and Rasul, I. (2009). Social connections and incentives in the workplace: Evidence from personnel data. *Econometrica*, 77(4):1047–1094.

Bandiera, O., Barankay, I., and Rasul, I. (2013). Team incentives: Evidence from a firm level experiment. *Journal of the European Economic Association*, 11(5):1079–1114.

Barankay, I. (2012). Rank incentives: Evidence from a randomized workplace experiment. *mimeo, Wharton School, University of Pennsylvania*.

Bartel, A. P., Beaulieu, N. D., Phibbs, C. S., and Stone, P. W. (2014). Human capital and productivity in a team environment: Evidence from the healthcare sector. *American Economic Journal: Applied Economics*, 6(2):231–59.

Bartel, A. P., Cardiff-Hicks, B., and Shaw, K. (2017). Incentives for lawyers: Moving away from "eat what you kill". *ILR Review*, 70(2):336–358.

Bartling, B., Fehr, E., Marechal, M. A., and Schunk, D. (2009). Egalitarianism and competitiveness. *American Economic Review*, 99(2):93–98.

Bednar, S. and Gicheva, D. (2014). Are female supervisors more female-friendly? *American Economic Review*, 104(5):370–75.

Bell, L., Smith, N., Smith, V., and Verner, M. (2008). Gender differences in promotion into top-management jobs. Workingpaper, Aarhus School of Business, Aarhus University, Department of Economics.

Benndorf, V., Rau, H. A., and Sölch, C. (2014). *Minimizing Learning Behavior in Experiments with Repeated Real-Effort Tasks*.

Besley, T. and Ghatak, M. (2008). Status incentives. *The American Economic Review*, 98(2):206–211.

Bettinger, E. P. and Long, B. T. (2005). Do faculty serve as role models? the impact of instructor gender on female students. *American Economic Review*, 95(2):152–157.

Betts, J. R. (2011). The economics of tracking in education. *Handbook of the Economics of Education*, 3:341 – 381.

Bidner, C. (2010). Pre-match investment with frictions. *Games and Economic Behavior*, 68(1):23 – 34.

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

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

Bloom, N., Sadun, R., and Van Reenen, J. (2016). Management as a technology? Working Paper 16-133, Harvard Business School Strategy Unit.

Bodoh-Creed, A. and Hickman, B. R. (2015a). Human capital investment and affirmative action: A structural policy analysis of us college admissions. Technical report, Typescript, University of Chicago Department of Economics.

Bodoh-Creed, A. and Hickman, B. R. (2015b). Using auction theory to study human capital investment in assortative matching markets: a look at affirmative action in college admissions. Technical report, Working paper, University of Chicago Department of Economics.

Booth, A. and Coles, M. (2010). Education, matching, and the allocative value of romance. *Journal of the European Economic Association*, 8(4):744–775.

Bradler, C., Dur, R., Neckermann, S., and Non, A. (2016). Employee recognition and performance: A field experiment. *Management Science*, 62(11):3085–3099.

Braun, S., Dwenger, N., Kübler, D., and Westkamp, A. (2014). Implementing quotas in university admissions: An experimental analysis. *Games and Economic Behavior*, 85:232—251.

Burgess, S., Propper, C., Ratto, M., Kessler Scholder, S. v. H., and Tominey, E. (2010). Smarter task assignment or greater effort: The impact of incentives on team performance. *The Economic Journal*, 120(547):968–989.

Burke, R. J. and McKeen, C. A. (1997). Not every managerial woman who makes it has a mentor. *Women in Management Review*, 12(4):136–139.

Burleson, B. R., Kunkel, A. W., Samter, W., and Working, K. J. (1996). Men's and women's evaluations of communication skills in personal relationships: When sex differences make a difference and when they don't. *Journal of Social and Personal Relationships*, 13(2):201–224.

Burxe, R. J., McKeen, C. A., and McKenna, C. (1993). Correlates of mentoring in organizations: The mentor's perspective. *Psychological Reports*, 72(3):883–896.

Calsamigla, C., Haeringer, G., and Klijn, F. (2010). Constrained school choice: An experimental study. *American Economic Review*, 100:1860—1874.

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

Casas-Arce, P. and Martínez-Jerez, F. A. (2009). Relative performance compensation, contests, and dynamic incentives. *Management Science*, 55(8):1306–1320.

Chan, T. Y., Li, J., and Pierce, L. (2014). Compensation and peer effects in competing sales teams. *Management Science*, 60(8):1965–1984.

Chao, G. T. (1997). Mentoring phases and outcomes. *Journal of Vocational Behavior*, 51(1):15 – 28.

Chaplin, J., Mangla, J., Purdon, S., and Airey, C. (2005). *The Workplace Employment Relations Survey (WERS) 2004 Technical Report (Cross-Section and Panel Surveys).* prepared for Department of Trade and Industry, London.

Charness, G., Kuhn, P., and Villeval, M. C. (2011). Competition and the ratchet effect. *Journal of Labor Economics*, 29(3):513–547.

Charness, G. and Villeval, M.-C. (2009). Cooperation and competition in intergenerational experiments in the field and the laboratory. *American Economic Review*, 99(3):956–78.

Chen, Y., Fehr, E., Fischbacher, U., and Morgan, P. (2015). Decentralized matching and social segregation. *Games and Economic Behavior*, 90:17 – 43.

Cole, H. L., Mailath, G. J., and Postlewaite, A. (2001). Efficient non-contractible investments in large economies. *Journal of Economic Theory*, 101(2):333 – 373.

Cooper, D. J., Kagel, J. H., Lo, W., and Gu, Q. L. (1999). Gaming against managers in incentive systems: Experimental results with chinese students and chinese managers. *American Economic Review*, pages 781–804.

Cronqvist, H., Heyman, F., Nilsson, M., Svaleryd, H., and Vlachos, J. (2009). Do entrenched managers pay their workers more? *The Journal of Finance*, 64(1):309–339.

Delfgaauw, J., Dur, R., Non, A., and Verbeke, W. (2015). The effects of prize spread and noise in elimination tournaments: A natural field experiment. *Journal of Labor Economics*, 33(3):521–569.

Delfgaauw, J., Dur, R., Sol, J., and Verbeke, W. (2013). Tournament incentives in the field: Gender differences in the workplace. *Journal of Labor Economics*, 31(2):305–326.

Devaro, J. and Kurtulus, F. A. (2010). An empirical analysis of risk, incentives and the delegation of worker authority. *Industrial and Labor Relations Review*, 63(4):641–661.

Drago, R. and Turnbull, G. K. (1988). The incentive effects of tournaments with positive externalities among workers. *Southern Economic Journal*, 55(1):100–106.

Dreher, G. F. and Ash, R. A. (1990). A comparative study of mentoring among men and women in managerial, professional, and technical positions. *Journal of applied psychology*, 75(5):539.

Dreher, G. F. and Cox, T. J. (1996). Race, gender, and opportunity: A study of compensation attainment and the establishment of mentoring relationships. *Journal of Applied Psychology*, 81(3):297–308.

Drews, F. A., Pasupathi, M., and Strayer, D. L. (2008). Passenger and cell phone conversations in simulated driving. *Journal of Experimental Psychology: Applied*, 14(4):392.

Eagly, A. H. and Crowley, M. (1986). Gender and helping behavior: A meta-analytic review of the social psychological literature. *Psychological Bulletin*, 100(3):283–308.

Eagly, A. H., Johannesen-Schmidt, M. C., and van Engen, M. L. (2003). Transformational, transactional, and laissez-faire leadership styles: A meta-analysis comparing women and men. *Psychological Bulletin*, 129(4):569–591.

Eby, L. T. d. T., Allen, T. D., Hoffman, B. J., Baranik, L. E., Sauer, J. B., Baldwin, S., Morrison, M. A., Kinkade, K. M., Maher, C. P., Curtis, S., et al. (2013). An interdisciplinary meta-analysis of the potential antecedents, correlates, and consequences of protégé perceptions of mentoring. *Psychological bulletin*, 139(2):441.

Ederer, F. (2010). Feedback and motivation in dynamic tournaments. *Journal of Economics and Management Strategy*, 19(3):733–769.

Eisenberg, N. and Fabes, R. A. (1991). *Prosocial Behavior and Empathy: A Multimethod Developmental Perspective*, volume 12. Sage: Thousand Oaks, CA, in margaret clark (ed.) edition.

Eisenkopf, G. (2010). Peer effects, motivation, and learning. *Economics of Education Review*, 29(3):364–374.

Englmaier, F., Roider, A., and Sunde, U. (2017). The role of communication of performance schemes: Evidence from a field experiment. *Management Science*, 63(12):4061–4080.

Erev, I., Bornstein, G., and Galili, R. (1993). Constructive intergroup competition as a solution to the free rider problem: A field experiment. *Journal of Experimental Social Psychology*, 29(6):463 – 478.

Eriksson, T., Poulsen, A., and Villeval, M. C. (2009). Feedback and incentives: Experimental evidence. *Labour Economics*, 16(6):679 – 688. European Association of Labour Economists 20th annual conference University of Amsterdam, Amsterdam, The Netherlands 18–20 September 2008.

Erkal, N., Gangadharan, L., and Nikiforakis, N. (2011). Relative earnings and giving in a real-effort experiment. *American Economic Review*, 101(7):3330–48.

Ertac, S. (2005). Social comparisons and optimal information revelation: Theory and experiments. *mimeo, University of California, Los Angeles.*

Eskreis-Winkler, L., Fishbach, A., and Duckworth, A. L. (2018). Dear abby: Should i give advice or receive it? *Psychological Science*, 29(11):1797–1806. PMID: 30281402.

Fagenson, E. A. (1989). The mentor advantage: Perceived career/job experiences of proteges versus non-proteges. *Journal of Organizational Behavior*, 10(4):309–320.

Falk, A., Becker, A., Dohmen, T., Huffman, D., and Sunde, U. (2016). The preference survey module: A validated instrument for measuring risk, time, and social preferences. Technical report, IZA Discussion Paper No. 9674.

Falk, A. and Ichino, A. (2006). Clean evidence on peer effects. *Journal of Labor Economics*, 24(1):39–57.

Felli, L. and Roberts, K. (2016). Does competition solve the hold-up problem? *Economica*, 83(329):172–200.

Fernandez, R., Guner, N., and Knowles, J. (2005). Love and money: A theoretical and empirical analysis of household sorting and inequality. *Quarterly Journal of Economics*, 120(1):273—344.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178.

Fletcher, J. K. and Ragins, B. (2008). *Stone Center Relational Cultural Theory: A Window on Relational Mentoring*. SAGE Publications, Inc., Thousand Oaks.

Fliessbach, K., Weber, B., Trautner, P., Dohmen, T., Sunde, U., Elger, C. E., and Falk, A. (2007). Social comparison affects reward-related brain activity in the human ventral striatum. *Science*, 318(5854):1305–1308.

Francis, A. M. and Tannuri-Pianto, M. (2012). Using brazil's racial continuum to examine the short-term effects of affirmative action in higher education. *Journal of Human Resources*, 47(3):754–784.

Franco, A. M., Mitchell, M., and Vereshchagina, G. (2011). Incentives and the structure of teams. *Journal of Economic Theory*, 146(6):2307 – 2332.

Fryer, R. G., Holden, R., et al. (2012). Multitasking, learning, and incentives: A cautionary tale. *NBER Working Paper*, 17752.

Gall, T., Legros, P., and Newman, A. (2012). Mismatch, rematch, and investment. Discussion paper series in economics and econometrics, University of Southampton, Economics Division, School of Social Sciences.

Gall, T., Legros, P., and Newman, A. F. (2006). The timing of education. *Journal of the European Economic Association*, 4(2-3):427–435.

Georganas, S., Tonin, M., and Vlassopoulos, M. (2015). Peer pressure and productivity: The role of observing and being observed. *Journal of Economic Behavior and Organization*, 117:223 – 232.

Ghosh, R. (2014). Antecedents of mentoring support: a meta-analysis of individual, relational, and structural or organizational factors. *Journal of Vocational Behavior*, 84(3):367 – 384.

Gill, D. and Prowse, V. (2012). A structural analysis of disappointment aversion in a real effort competition. *American Economic Review*, 102(1):469–503.

Gill, M. J., Roulet, T. J., and Kerridge, S. P. (2018). Mentoring for mental health: A mixed-method study of the benefits of formal mentoring programmes in the english police force. *Journal of Vocational Behavior*, 108:201 – 213.

Godshalk, V. M. and Sosik, J. J. (2003). Aiming for career success: The role of learning goal orientation in mentoring relationships. *Journal of Vocational Behavior*, 63(3):417 – 437.

Greiner, B. (2004). *An Online Recruitment System for Economic Experiments.*

Griffith, R. and Neely, A. (2009). Performance pay and managerial experience in multitask teams: Evidence from within a firm. *Journal of Labor Economics*, 27(1):49–82.

Haggard, D. L., Dougherty, T. W., Turban, D. B., and Wilbanks, J. E. (2011). Who is a mentor? a review of evolving definitions and implications for research. *Journal of Management*, 37(1):280–304.

Hamilton, B. H., Nickerson, J. A., and Owan, H. (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.

Hanushek, E. A. and Woessmann, L. (2011). The economics of international differences in educational achievement. *Handbook of the Economics of Education*, 3:89 – 200.

Harrison, G. W. and List, J. A. (2004). Field experiments. *Journal of Economic Literature*, 42(4):1009–1055.

Hatfield, J. W., Kojima, F., and Kominers, S. D. (2017). Strategy proofness, investment efficiency, and marginal returns: An equivalence. Technical report, Becker Friedman Institute for Research in Economics Working Paper.

Hinde, R. A. (1997). *Relationships: A Dialectical Perspective.* Psychology Press.

Hjort, J. (2014). Ethnic divisions and production in firms. *The Quarterly Journal of Economics*, 129(4):1899–1946.

Holmstrom, B. and Milgrom, P. (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership and job design. *Journal of Law, Economics and Organization*, 7:24–52.

Hong, F., Hossain, T., List, J. A., and Tanaka, M. (2018). Testing the theory of multitasking: Evidence from a natural field experiment in chinese factories. *International Economic Review*, 59(2):511–536.

Hopkins, E. (2012). Job market signaling of relative position, or becker married to spence. *Journal of the European Economic Association*, 10(2):290–322.

Hoppe, H. C., Moldovanu, B., and Sela, A. (2009). The theory of assortative matching based on costly signals. *The Review of Economic Studies*, 76(1):253–281.

Hossain, T. and List, J. A. (2012). The behavioralist visits the factory: Increasing productivity using simple framing manipulations. *Management Science*, 58(12):2151–2167.

Ichniowski, C. and Shaw, K. (2003). Beyond incentive pay: Insiders' estimates of the value of complementary human resource management practices. *Journal of Economic Perspectives*, 17(1):155–180.

Ichniowski, C., Shaw, K., and Prennushi, G. (1997). The effects of human resource management practices on productivity: A study of steel finishing lines. *The American Economic Review*, 87(3):291–313.

Itoh, H. (1991). Incentives to help in multi-agent situations. *Econometrica*, 59(3):611–636.

Johnson, N. B. and Scandura, T. A. (1994). The effect of mentorship and sex-role style on male-female earnings. *Industrial Relations: A Journal of Economy and Society*, 33(2):263–274.

Kahneman, D., Krueger, A. B., Schkade, D., Schwarz, N., and Stone, A. A. (2006). Would you be happier if you were richer? a focusing illusion. *science*, 312(5782):1908–1910.

Kammeyer-Mueller, J. D. and Judge, T. A. (2008). A quantitative review of mentoring research: Test of a model. *Journal of Vocational Behavior*, 72(3):269 – 283.

Kandel, E. and Lazear, E. P. (1992). Peer pressure and partnerships. *Journal of Political Economy*, 100(4):801–817.

Karatepe, O. M. (2013). The effects of work overload and work-family conflict on job embeddedness and job performance: The mediation of emotional exhaustion. *International Journal of Contemporary Hospitality Management*, 25(4):614–634.

Kishore, S., Rao, R. S., Narasimhan, O., and John, G. (2013). Bonuses versus commissions: A field study. *Journal of Marketing Research*, 50(3):317–333.

Kluger, A. N. and Denisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, pages 254–284.

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

Kram, K. E. (1985). *Mentoring at Work: Developmental Relationships in Organizational Life*. Organizational behavior and psychology series. Scott, Foresman, Glenview, Ill.

Kremer, M., Glewwe, P., and Ilias, N. (2010). Teacher incentives. *American Economic Journal: Applied Economics*, 2(3):205–227.

Laband, D. N. and Lentz, B. F. (1999). The impact of having a mentor on earnings and promotion: Evidence from a panel study of lawyers. *Applied Economics Letters*, 6(12):785–787.

Layard, R. (2011). *Happiness: Lessons from a New Science*. Penguin UK.

Lazear, E. P. (1989). Pay equality and industrial politics. *Journal of Political Economy*, 97(3):561–580.

Lazear, E. P. (2000). Performance pay and productivity. *American Economic Review*, 90(5):1346–1361.

Lazear, E. P. and Rosen, S. (1981). Rank-order tournaments as optimum labor contracts. *Journal of Political Economy*, 89(5):841–864.

Lazear, E. P., Shaw, K. L., and Stanton, C. T. (2015). The value of bosses. *Journal of Labor Economics*, 33(4):823–861.

Liang, J. and Gong, Y. (2013). Capitalizing on proactivity for informal mentoring received during early career: The moderating role of core self-evaluations. *Journal of Organizational Behavior*, 34(8):1182–1201.

Lim, N., Ahearne, M. J., and Ham, S. H. (2009). Designing sales contests: Does the prize structure matter? *Journal of Marketing Research*, 46(3):356–371.

Lizzeri, A., Meyer, M. A., and Persico, N. (2002). *The Incentive Effects of Interim Performance Evaluations*. University of Pennsylvania, Center for Analytic Research in Economics and the Social Sciences.

Manthei, K., Sliwka, D., et al. (2018). Multitasking and subjective performance evaluations: Theory and evidence from a field experiment in a bank. Technical report, Institute for the Study of Labor (IZA).

Mark, G., Voida, S., and Cardello, A. (2012). A pace not dictated by electrons: an empirical study of work without email. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 555–564. ACM.

Markiewicz, D., Devine, I., and Kausilas, D. (2000). Friendships of women and men at work: Job satisfaction and resource implications. *Journal of Managerial Psychology*, 15(2):161–184.

Marsden, P. V., Cook, C. R., and Kalleberg, A. L. (1996). *Bureaucratic Structures for Coordination and Control*. Thousand Oaks, CA: Sage, in organizations in america: analyzing their structures and human resource practices, eds. edition.

Mas, A. and Moretti, E. (2009). Peers at work. *American Economic Review*, 99(1):112–45.

Matsa, D. A. and Miller, A. R. (2011). Chipping away at the glass ceiling: Gender spillovers in corporate leadership. *American Economic Review*, 101(3):635–39.

McManus, S. E. and Russell, J. E. (1997). New directions for mentoring research: An examination of related constructs. *Journal of Vocational Behavior*, 51(1):145 – 161.

Mohnen, A., Pokorny, K., and Sliwka, D. (2008). Transparency, inequity aversion, and the dynamics of peer pressure in teams: Theory and evidence. *Journal of Labor Economics*, 26(4):693–720.

Moldovanu, B., Sela, A., and Shi, X. (2007). Contests for status. *Journal of Political Economy*, 115(2):338–363.

Muralidharan, K. and Sundararaman, V. (2013). Contract teachers: Experimental evidence from india. Working Paper 19440, National Bureau of Economic Research.

Neumark, D. and Gardecki, R. (1998). Women helping women? role model and mentoring effects on female ph.d. students in economics. *The Journal of Human Resources*, 33(1):220–246.

Noe, R. A. (1988). An investigation of the determinants of successful assigned mentoring relationships. *Personnel Psychology*, 41(3):457–479.

O'Brien, K. E., Biga, A., Kessler, S. R., and Allen, T. D. (2010). A meta-analytic investigation of gender differences in mentoring. *Journal of Management*, 36(2):537–554.

Oyer, P. and Schaefer, S. (2010). Personnel economics: Hiring and incentives. Working Paper 15977, National Bureau of Economic Research.

Pais, J. and Pinter, A. (2008). School choice and information: An experimental study on matching mechanisms. *Games and Economic Behavior*, 64(1):303 – 328.

Payne, S. C. and Huffman, A. H. (2005). A longitudinal examination of the influence of mentoring on organizational commitment and turnover. *The Academy of Management Journal*, 48(1):158–168.

Peters, M. and Siow, A. (2002). Competing premarital investments. *Journal of Political Economy*, 110(3):592–608.

Powell, G. N. (2010). *Women and Men in Management*. Thousand Oaks, CA: Sage, fourth edition edition.

Ragins, B. R. (1989). Barriers to mentoring: The female manager's dilemma. *Human Relations*, 42(1):1–22.

Ragins, B. R. and Cotton, J. L. (1999). Mentor functions and outcomes: A comparison of men and women in formal and informal mentoring relationships. *Journal of Applied Psychology*, 84(4):529.

Ragins, B. R., Cotton, J. L., and Miller, J. S. (2000). Marginal mentoring: The effects of type of mentor, quality of relationship, and program design on work and career attitudes. *Academy of Management Journal*, 43(6):1177–1194.

Ragins, B. R. and McFarlin, D. B. (1990). Perceptions of mentor roles in cross-gender mentoring relationships. *Journal of Vocational Behavior*, 37(3):321 – 339.

Rob, R. and Zemsky, P. (2002). Social capital, corporate culture, and incentive intensity. *The RAND Journal of Economics*, 33(2):243–257.

Scandura, T. A. (1992). Mentorship and career mobility: An empirical investigation. *Journal of Organizational Behavior*, 13(2):169–174.

Scandura, T. A. and Schriesheim, C. A. (1994). Leader-member exchange and supervisor career mentoring as complementary constructs in leadership research. *The Academy of Management Journal*, 37(6):1588–1602.

Shearer, B. (2004). Piece rates, fixed wages and incentives: Evidence from a field experiment. *The Review of Economic Studies*, 71(2):513–534.

Siemsen, E., Balasubramanian, S., and Roth, A. V. (2007). Incentives that induce task-related effort, helping, and knowledge sharing in workgroups. *Management Science*, 53(10):1533–1550.

Slade, M. E. (1996). Multitask agency and contract choice: An empirical exploration. *International Economic Review*, 37(2):465–486.

Sosik, J. J. and Godshalk, V. M. (2000). Leadership styles, mentoring functions received, and job-related stress: A conceptual model and preliminary study. *Journal of Organizational Behavior*, 21(4):365–390.

Van den Steen, E. (2005). Organizational beliefs and managerial vision. *The Journal of Law, Economics, and Organization*, 21(1):256–283.

Van den Steen, E. (2010). On the origin of shared beliefs (and corporate culture). *The RAND Journal of Economics*, 41(4):617–648.

Vranceanu, R., El Ouardighi, F., and Dubart, D. (2013). Coordination in teams: A real effort-task experiment with informal punishment. ESSEC Working Papers WP1310, ESSEC Research Center, ESSEC Business School.

Wajcman, J. (1996). Desperately seeking differences: Is management style gendered? *British Journal of Industrial Relations*, 34(3):333–349.

Wallace, J. E. (2001). The benefits of mentoring for female lawyers. *Journal of Vocational Behavior*, 58(3):366–391.

Watson, B. (2003). Chuang-tzu: Basic writings (trans. 1964). *New York*.