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

DISCUSSION PAPER SERIES

IZA DP No. 10454

Working Hours, Promotion and the Gender Gap in the Workplace

Takao Kato
Hiromasa Ogawa
Hideo Owan

DECEMBER 2016

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

DISCUSSION PAPER SERIES

IZA DP No. 10454

Working Hours, Promotion and the Gender Gap in the Workplace

Takao Kato
Colgate University, IZA, TCER, CJEB, CCP, ETLA and Rutgers University

Hiromasa Ogawa
National Graduate Institute for Policy Studies

Hideo Owan
University of Tokyo and RIETI

DECEMBER 2016

IZA – Institute of Labor Economics

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

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

www.iza.org

ABSTRACT

Working Hours, Promotion and the Gender Gap in the Workplace*

This paper presents a novel model of promotion within the firm which sheds new light on the interplay between working hours and the odds of subsequent promotion. The model's key feature is the coexistence of two different sources of asymmetric information: (i) the worker's cost of long working hours: and (ii) the worker's OJT ability (the worker's ability to accumulate valuable human capital on the job through learning by doing). The worker's cost of working long hours is known only to the worker, while the worker's OJT ability is accurately assessed only by the firm observing him/her on the job. Long working hours signal the worker's commitment to the firm, which determines the surplus produced when the worker is promoted. Thus, the firm provides the worker with managerial training only after observing the employee's hours worked, a signal of his/her commitment to the firm or lack thereof. The firm's decision to provide training also depends on its private information about the worker's OJT ability, which affects his/her future productivity if and when the worker gets promoted. Upon completion of training, the firm then promotes the worker. The model illuminates under what conditions, it is efficient for the firm to adopt the information revelation strategy – reveal its private information on the worker's OJT ability to him/her before the worker decides on whether to work long hours and signal his/her commitment. Using the model, we show that under a reasonable set of conditions, the firm may find it optimal to adopt the information revelation strategy for women but not for men, and derive an empirical testable hypothesis that the correlation between working hours and subsequent promotion will be stronger for women than for men. We analyze longitudinal personnel data from a large Japanese manufacturing firm and provide rigorous econometric evidence in support of the hypothesis.

JEL Classification: M51, J16

Keywords: working hours, promotion, rat race, adverse selection, and the gender gap

Corresponding author:
Takao Kato
Department of Economics
Colgate University
13 Oak Drive
Hamilton, NY 13346
USA

E-mail: tkato@colgate.edu

* This study is conducted as a part of the Project "Economic Analysis of Human Resource Allocation Mechanisms within the Firm: Insider econometrics using HR data" undertaken at Research Institute of Economy, Trade, and Industry (RIETI), and this paper is based on the RIETI Discussion Paper Series 16-E-060. This work was also supported by JSPS KAKENHI Grant Number JP25245041.

Working Hours, Promotion, and the Gender Gap in the Workplace

1. Introduction

We develop a new model of promotion within the firm which provides fresh insight on the interplay between the employee's working hours and his/her probability of earning subsequent promotion. Specifically the worker in our model differs in his/her disutility of long working hours or his/her cost of working long hours as well as in his/her ability to accumulate human capital on the job though learning by doing or On-the-Job Training (OJT) ability. The worker's cost of working long hours is known to the worker but not to the firm at the beginning of his/her tenure at the firm---the worker's cost of working long hours is his/her private information. The firm prefers to train employees who are willing to work long hours because a higher rent is generated by promoting such workers.

The productivity of the worker with a higher OJT ability will grow more rapidly over time. In addition, he/she will become a more productive manager if and when gets promoted, for he/she can acquire a variety of skills and knowledge by experiencing a wide range of jobs more effectively thanks to his/her higher OJT ability (see, for instance, Gibbons and Waldman, 2004; Lazear, 2012, Frederiksen and Kato, 2016). The worker's OJT ability is initially unknown to both parties when he/she starts working at the firm, yet it is revealed to the employer earlier than the worker himself/herself at a relatively early stage of his/her tenure.

It is highly unlikely that either the firm or the worker knows accurately how well the worker accumulates human capital (in particular firm-specific human capital) on the job before he/she starts working at the firm. It is plausible that as he/she continues to work and engage in learning-by-doing, the firm will continue to observe the worker's progress in his/her accumulation of human capital on the job and collect/keep relevant information (not only formal but also informal) on the worker's OJT ability. At some point the firm will have a sufficient amount of information to discern the worker's OJT ability with precision. We argue that such information is the firm's private information which the worker cannot access unless the firm decides to share it with him/her. It is possible that the worker himself/herself learns his/her own OJT ability as he/she continues to work. However, having observed many workers and having discerned their OJT abilities in the past as well as understanding deeply what type of human capital is more valuable to the firm, the firm has a clear advantage over the worker in figuring out the worker's OJT ability quickly. Prendergast (1993) have shown that the model with this feature can illustrate how late promotion policy, one of several key characteristics of the Japanese human resource management system, arises in equilibrium.

After learning each worker's OJT ability and having observed his/her hours worked in the first period, the employer will then decide which worker will receive managerial training. Only those workers who have completed such training can potentially perform managerial tasks effectively and thus get promoted to managers.

Using this model, we illustrate how a "rat race" equilibrium in which a substantial portion of workers work inefficiently long hours could arise. We also examine under what conditions it is optimal for the firm to share with the worker its private information on his/her OJT ability. The firm's revealing its private information on the worker's OJT ability to him/her (thereafter we call it information revelation) affects his/her working hours. In general, information revelation can result in an increase or a decrease in the incentive for the high-type workers with high cost of long working hours to mimic the low-cost workers and work long hours. It depends on the nature of training. If the training programs are designed to target high-performers and thus complement worker ability, information revelation is likely to motivate high-ability workers to work longer hours. In contrast, if the training programs mainly target low-ability workers and develop uniform skill distribution among workers, training and ability will be substitutes in the production of promotion. In this case the information revelation policy will discourage high-ability workers from working longer hours.

When training and OJT ability are complementary inputs in the production of promotion, the high-ability worker, once his/her high ability is revealed by the firm, finds the *ex ante* expected benefit of working long hours (and hence getting training) greater, and is thereby more likely to work long hours. Information revelation generates an important trade-off. On the one hand, information revelation intensifies competition for promotion which leads to more inefficiently long working hours for qualified workers. Here, information revelation hampers the ex-ante efficiency in terms of how much effort the workers put forth in order to be selected as managers. On the other hand, information revelation induces unqualified workers to stop working inefficiently long hour, which is Pareto improving. It follows that total changes in working hours (and hence changes in the *ex ante* efficiency) caused by information revelation depends on the balance of the increase in working hours of high-ability workers and the decrease in working hours of low-ability workers. Such a balance in turn depends on the distribution of the cost of long working hours.[1]

Suppose the distribution of the cost of long working hours is highly right-skewed (a

[1]Another factor that affects the firm's decision involving the trade-off is overtime hourly wage. If overtime pay is zero, the firm will always prefer to maximize working hours whereas, if overtime pay is very high, the firm will choose the information revelation policy that minimizes overtime work hours.

majority of workers have very high cost of long working hours). As a result of the right-skewed distribution of the cost of working long hours, the increase in working hours of high-ability workers is likely to be more restrained by their high cost of working long hours, while the decrease in working hours of low-ability workers is apt to be more forthcoming due to their high cost of working long hours. As such, in terms of the *ex ante* efficiency loss due to excessively long working hours---or intensified rat race, information revelation is more efficient when the distribution of the cost of long working hours is more right-skewed.

We then use the model and derive an important implication for gender differences in promotion. In so doing, we consider a case that the distribution of the cost of working long hours is more right-skewed for women than for men---the case appears to be quite plausible, in light of social norm placing more demand on women for household production than men. In this plausible case, the model shows that it can be efficient for the firm to adopt the information revelation strategy for women and the information concealment strategy for men. Such gender-based use of differential HRM strategies will be prohibitively costly if it is difficult for the firm to use an implicit contract to carry out the gender-based use of differential HRM strategies, and gender discrimination in the workplace is vigorously litigated. However, implicit contracts for employment relations have been pervasive in Japan,[2] and it is quite plausible that Japanese firms can use an implicit contract to apply the information revelation strategy for women, while applying the information concealment strategy for men.

We further demonstrate that in our model the use of the information revelation

[2] See, for instance, Moriguchi (2002) for a historical account for the continued pervasiveness of implicit employment contracts in Japan.

strategy for women and the information concealment strategy for men will result in an empirically testable hypothesis---the relationship between working hours and the odds of subsequent promotion is stronger for women than for men. The model predicts that in the plausible case of the right-skewed distribution of the cost of long working hours for women, the firm finds it efficient to reveal its private information on each female worker's OJT ability to her, while concealing its private information on each male worker's OJT ability from him. Intuitively the revelation of ability for women is more likely to lead to a separation of high-ability women from the rest, and high-ability women are more likely to work long hours, receive training, and get promoted. A strong positive correlation between long working hours and the odds of promotion for women will follow. In contrast, for men, such information is concealed, and the firm treats all workers equally as if they were of equal OJT ability. Since everyone thinks that he has some chance of getting promoted, everyone works long hours to signal their low cost and commitment to the firm. But, in fact, only those with high OJT ability are chosen for managerial training and get promoted to management positions. Therefore, the correlation between working hours and the odds of subsequent promotion is weaker for men.

Finally, in order to test the hypothesis, we obtain detailed longitudinal personnel records from a large manufacturing firm in Japan with about 6,000 employees. In addition to usual biographical data, the personnel records contain detailed information on each employee's job assignment annually for 2005-2013. The longitudinal job assignment data, combined with our institutional knowledge of the firm's promotion ladders, allow us to define promotion (our key dependent variable) with little measurement error. Furthermore, the personnel records also include accurate data on actual working hours (our primary explanatory variable) for all employees over the same time period. We find consistently for all specifications that the employee's number of hours worked in the previous year is positively associated with his/her odds of promotion, and more importantly that such positive associations between working hours and the probability of promotion are significantly stronger for women than for men, supporting the hypothesis.

In the next section we place our paper in the relevant literature and articulate the paper's contributions. In section 3 we present the model, and demonstrate its utility by deriving numerical examples from the model which will help providing a coherent interpretation of stylized facts about international differences in the gender gap in the labor market. In section 4, from the model (signaling regime), we derive a prediction concerning the gender difference in the hours-promotion linkage and provide econometric evidence that is consistent with the prediction. Concluding remarks are given in section 5.

2. Related Literature

This paper builds on the "rat race" adverse selection models with heterogeneous costs of work speed, effort or work hours. Akerlof (1976) shows that when workers are given the choice of choosing among assembly lines that run at different speeds, they end up choose those that run too fast, for by doing so they can minimize their odds of working with low-productive workers and sharing output with such low-productivity workers and maximize their odds of benefitting from working with more productive workers. Miyazaki (1977), Stiglitz (1975) and Sampson and Simmons (2000) formalizes or generalizes Akerlof's model. Landers, Robitzer, and Taylor (1996) and

Bardsley and Sherstyuk (2006) extends this idea to a dynamic setting using overlapping generation models where a successful worker becomes a partner (co-owner) in the next period and write a contract for new workers. For example, Landers, Robitzer, and Taylor (1996) find that the equilibrium is characterized by low wage and inefficiently high effort levels. Our model demonstrates that a “rat race” could arise in a signaling model where employees voluntarily choose working hours—many workers work inefficiently long hours in order to signal their commitment to the firm (the worker’s private information) and maximize their chances of receiving managerial training and promotion.

We then incorporate into the rat race model a key feature of another important related literature, the inefficient promotion literature (e.g., Waldman 1984, Prendergast, 1992, Owan 2004, Ishida 2004). The firm can possess private information about the employee’s productivity and use this information strategically. For instance, according to Prendergast (1992), it can be in the interest of the firm to slow down the speed of promotion of its employees and avoid revealing its private information on the worker’s promotion prospect, for such a late promotion policy can perverse low-ability workers’ motivation to work hard and long for an extended period of time. In sum, by exploring an intersection between the rat race literature and the inefficient promotion literature, we develop a rat race model with a novel feature--- the firm’s strategic decision on whether to reveal its private information on the employee’s OJT ability and hence the firm’s intention to promote him/her (in addition to the worker’s signaling decision regarding whether to work long hours to signal his/her private information on the employee’s commitment to the firm).

Using our novel rat race adverse selection model we makes contributions to the literature on the gender gap in the labor market in general and the gender gap in promotion in particular. Relative to the vast literature on the gender pay gap, the literature on the gender gap in promotion is relatively small. Most early works focus on documenting that promotion rates are lower for women than for men with similar observed characteristics (e.g., Cabral, Ferber, and Green 1981, Cannings 1988, Cobb-Clark 2001, Paulin and Mellor 1996, Pekkarinen and Vartianinen, 2004).

Two competing explanations for such “unexplained” gender differences in promotion were proposed. First, there may be unobserved productivity differences or female preferences for different job characteristics (e.g. jobs with less supervisory responsibilities and/or shorter working hours).[3] Such differences may be caused, for example, by gendered division of labor in the family—married women allocate more effort to child care and housework, and less effort to market work (Becker 1985). More recent variants of this line of theorizing include behavioral models of the gender gap in promotion tournament with particular emphasis on the gender difference in preferences for competition and risk.[4]

The second explanation attributes the differences to taste-based discrimination or statistical discrimination (Becker 1957, Phelps 1972, Arrow 1973, Lazear and Rosen 1990). Later works extend the statistical discrimination model to focus on more specific

[3] The sociological literature emphasizes that the disadvantages associated with being in an occupation dominated by women persistently exists simply because bureaucratization and rationalization institutionalize the disadvantage in formal job description, job ladders, and patterns of pay progression. See Barnett, Baron, and Stuart (2000) for example.

[4] There is a growing body of experimental evidence suggesting that women shy away from competitive work environment while men embrace it (see, for example, Niederle and Vesterlund, 2007; Booth and Nolen, 2012; Datta Gupta et al., 2013; Garratt et al., 2013; and Sutter and Rutzler, 2015). Moreover, some experimental evidence suggests that in such competitive environment, women tend to perform worse than men (see, for instance, Gneezy et al., 2003; Gneezy and Rustichini, 2004). Perhaps most importantly a recent experimental study by Kuhn and Villeval (2015) provided evidence pointing to women’s preference for jobs in cooperative team environment over jobs in non-cooperative competitive environment.

aspects of the gender gap in promotion. For instance, Booth, Francesconi, and Frank (2003) focus on the limited outside job opportunities for female managers due to their demanding obligations at home and show that promotion is accompanied by smaller pay raise for women than for men. Most recent works explore the dynamics of the statistical discrimination model and derive somewhat more mixed predictions on the gender gap in promotion (Fryer, 2007 and Bjerk, 2008).

Empirical evidence on the gender gap in promotion is rather limited mostly due to the scarcity of the data suitable for testing alternative hypotheses concerning the gender gap in promotion. Blau and DeVaro (2006), using the Multi-City Study of Urban Inequality, find that gender differences in the odds of promotion still remain even after accounting for job performance, occupation, and detailed firm characteristics. Their work reinforces the view that some form of discrimination may be at work behind the gender differences. Similar results are also found by McCue 1996; and Cobb-Clark 2001. More recently Smith, Smith, and Verner (2013), Frederiksen and Kato (2016), and Cassidy, DeVaro, and Kauhanen (2016) take advantage of detailed and reliable registry data from Denmark and Finland, and confirm that the significant gender gap in the odds of promotion to top management is still pervasive and that such gender gap cannot be explained fully by a variety of individual and firm characteristics.

As mentioned, in contrast to the literature on the gender gap in promotion, the literature on the gender pay gap is immense. We refer to only a few recent contributions to the literature that are of particular relevance to our study. Among other things, a number of researchers started to pay particular attention to the importance of the gender difference in working hours as a major culprit for the persistent gender wage gap. For example, Bertrand, Goldin and Katz (2010) use a panel of MBAs from the University of

Chicago find that much of the gender pay gap can be accounted for by career interruptions due to parenting and short working hours. More recently similar evidence on the importance of the gender difference in working hours for the gender pay gap is found for a more general sample of workers (Cha and Weeden, 2013 for the U.S. and Frederisen, Kato, and Smith 2015 for Denmark) and for the population of all employees at a large manufacturing firm in Japan (Kato, Kawaguchi and Owan, 2013). Goldin (2014) also shows that occupations characterized by high returns to overwork are also those with the largest gender gap in earnings. Taking a step forward, Cortes and Pan (2014) have shown that low-skilled immigration leads to a reduction in the gender gap in weekly hours worked, as well as the gender pay gap, particularly in occupations that disproportionately reward longer hours of work.

On our reading of the literature, no rigorous attempt has been made to examine the gender difference in working hours as a possible culprit for the gender gap in promotion although we are certainly not the first to suggest the possibility that the gender gap in hours may account for the gender difference in promotion. For instance, Bardsley and Sherstyuk (2006)'s theoretical work considers possible linkage between rat race and glass ceiling in an overlapping generation model of rat race with the agent becoming the principal after promoted successfully in an continuous organization such as law firms in the U.S.. More recently Gicheva (2013), after presenting rigorous evidence on the relationships between working hours and wage growth (and promotion) in general, provide additional insight on the gender gap in pay and promotion in the context of her dynamic labor supply model. In fact, in developing our empirical strategy to yield evidence that is consistent with our model, we draw on her approach.[5]

3. The Theory

3.1 Set up

There are many firms and workers in a competitive labor market. Firms can enter the market without cost. All workers are employed for two periods. We assume the discount rate to be zero. Workers are different in their commitment level and their ability—learning capability.

A worker's commitment level b determines his/her disutility of working long hours with $b \in B = \{b_1, b_2, b_3, \ldots, b_n\}$ where $b_1 < b_2 < .. < b_n$. More precisely, Type-i's utility is given by $w_1 + w_2 - \frac{h}{b_i}$ where w_t is the wage in period *t* and *h* is the overtime working hours. Let q_i be the probability that $b = b_i$. A worker's ability or learning capability is denoted by $a \in \{a_H, a_L\}$ where $a_L < a_H$. The ability affects the probability that he/she gets promoted to a manager $\mu(a, \delta)$ and the speed of learning-by-doing $f(a)$ where δ is the amount of training provided by the employer. Let *p* be the share of the workers with ability a_H.

A worker become privately informed about his/her commitment level b_i (or type *i*) after the worker enters the firm and its distribution $\{q_i\}$ is public knowledge. A worker's OJT ability is initially unknown for all participants (even for the worker

[5] Gicheva (2013) finds that the relationship between hours and wage growth is non-linear—for workers who put in 48 hours per week or more, working 5 extra hours per week increases annual wage growth by 1 percent, but when hours are less than 48, the average effect is zero. She also shows that working five extra hours per week increases the probability of receiving promotion by more than 2.5%. We will identify similar non-linear relationships between hours and promotion odds. However unlike Gicheva (2013), we study the gender differences in the relationships between hours and promotion odds explicitly and uncover significant gender differences.

himself/herself) but is revealed to the firm soon after the entry but not to the worker. The key assumption in the model is that the firm can share this private information with the worker in order to influence his/her choice of hours of work. We assume that, when private information on the worker's OJT ability is shared with the worker, the employer can verify the accuracy of the information with sufficient persuasive evidence. In other words, although concealing the information is feasible, presenting fabricated information is not allowed. Let $s \in \{R, N\}$ be the information revelation strategy where $s = R$ indicates that the firm truthfully signals a_H or a_L while $s = N$ indicates that the firm sends no signal. We assume that the firm can commit to its information revelation policy.

When workers are employed, the workers produce output by working in each period. In the first period, they are assigned to easy jobs and their marginal productivity is constant and independent of their ability and the same across firms. Then, their output depends only on the number of hours worked. We assume workers produce $y_1 = x_1 + mh$, where x_1 is the productivity during regular hours which is constant, $m > 0$ is the marginal productivity of overtime work, and h is the overtime working hours in period 1.

In the second period, the worker gets promoted with the probability $\mu(a, \delta)$ and produces $y_2 = x_2 + 2f(a)h + 2R(b)$ whereas those who do not get promoted (with the probability 1- $\mu(a, \delta)$) produce $y_2 = x_2 + 2f(a)h$. x_2 is again constant. Note that overtime work done by a young worker positively affect the productivity when he/she gets older. So, we assume learning-by-doing here. Their ability is fully revealed to the workers themselves perhaps as a result of training before the production starts in the second period. We assume $f(a_H) > f(a_L)$, $\mu(a_H, \delta) > \mu(a_L, \delta), \mu_\delta >$

$0, \mu_{\delta\delta} < 0$, and $R(b)$ is increasing in b. Simply said, workers who have high learning capability are more likely to be qualified for managerial jobs (μ is increasing in a) and more committed managers generate more value (R is increasing in b).

Note that the worker's learning ability and his/her overtime working hours in period 1 are complementary in the second-period productivity, which gives the employer an incentive to reveal the worker's ability to the worker himself/herself in period 1. The firm can increase the probability that a worker is qualified for a managerial position by training the worker (*i.e.* $\mu_\delta > 0$) with the cost $\kappa(\delta)$ where $\kappa(\delta)$ is increasing and convex in δ. We argue that workers who receive managerial training experience a wider range of tasks and thus become more capable of identifying the cause of a problem and better coordinating with other workers (for theory and evidence on the importance of job rotation and broadening the scope of human capital for appointments to top management, see Gibbons and Waldman. 2004, Lazear, 2012, and Frederiksen and Kato, 2016)

In our model, the long-term contract is not feasible. When hiring workers, a firm can only offer a first-period base wage w^B. We assume that the regular working hours $\underline{h}$ is determined exogenously by working hour regulations, whereas the overtime hourly wage is also legally determined by $w^{OT} = \rho \cdot w^B/\underline{h}$. Therefore, the first-period wage is given by $w_1 = w^B + w^{OT}h$. Note that the training level δ is not part of the initial contract and we assume that the firm cannot commit to a specific level of δ. Also that, since the workers' commitment type and ability type are unknown for all participants before the firm hires them, it is impossible to screen out the workers through designing the base wage and information revelation policy. Then, in competitive labor market the equilibrium base wage is simply pinned down as to make the expected profit of the firm to be zero.

In the second period, the firm and the worker bargain over wage. Note that at this point, there is no asymmetric information between the firm and the worker. We assume that skills acquired through learning-by-doing and managerial training are firm-specific and those who quit at the beginning of the second period produce x_2 by working for other firms. Assuming the Nash bargaining solution, the second period wage is determined by $w_2 = x_2 + f(a)h + R(b)$ for managers and $w_2 = x_2 + f(a)h$ for regular workers.

As we state earlier, the worker's ability is the employer's private information in the first period. The firm has to decide whether to feedback this ability information to the worker or not. Our key assumption is that the firm can commit to its feedback policy—either information revelation (early selection) or non-information revelation (late selection). We postulate that the reputation mechanism works until some parameter change makes it profitable for the firm to switch. Otherwise, the firm often has an incentive to reveal the information to a high-ability worker and hide it from low-ability one (or vice versa).

The game flow is depicted in Figure 1:

1. Firms offer base wage w^B to workers in the labor market and the workers decide whether to accept it or not. A worker who accepts the contact enters the firm. Firms also decide on their feedback policy *s* (either information revelation, i.e. $s =$ R, or non-information revelation, $s =$ N).
2. In the first period:
   (a) each worker recognizes his/her commitment level *b*;

(b) the firm observes each hired worker's ability to acquire human capital on the job and reveals (or does not reveal) this assessment to the worker if information revelation (non-information revelation) is chosen at the beginning; and

(c) the worker chooses hours of work and earns a wage according to the contract.

3. In the second period:

(a) the firm provides each worker with a certain amount of training δ and the size of δ is determined by the firm, based on its assessment of the worker ability and the hours of work in the first period;

(b) workers are promoted with the probability $\mu(a, \delta)$;

(c) the firm and the worker bargain over the wage and hours of work after the worker's ability and commitment type are known to both parties; and

(d) the worker works and the wage is paid out.

Note that the worker does not make any decisions in the second period.

The solution concept we use is the notion of perfect Bayesian equilibrium, which requires the following elements:

(i) the firm's feedback policy *s* and its strategy in training $\delta^*(a, h)$ maximize the firm's expected profit given the belief function $\lambda(b|a, h, s)$;

(ii) a worker's strategy in choosing working hours $\sigma(h|s, a, b)$ maximizes the worker's expected utility, where $\sigma(h|s, a, b)$ is a probability distribution over working hour *h* for each commitment type *b*;

and

(iii) the belief function $\lambda(b|a, h, s)$ is consistent with $\sigma(h|s, a, b)$; for any *i*,

$$\lambda(b_i|a,h,s) = \frac{q_i\sigma(h|s,a,b_i)}{\sum_{b_j\in B} q_j\sigma(h|s,a,b_j)} \ if \ \sum_{b_j\in B} q_i\sigma(h|s,a,b_j) > 0,$$

and $\lambda(b|a,h,s)$ is any probability distribution on B if $\sum_{b_j\in B} q_i\sigma(h|s,a,b_j) = 0$.

As we explain in more details later, perfect Bayesian equilibrium cannot rule out some unreasonable equilibria. We refine the set of equilibria by further imposing D1-Criterion.

3.2 Results

Preliminary

Let $u(a,b,h,\delta)$ be the utility of the worker with ability a and the commitment type b when he/she chooses the hours of work h in the first period and the employer provides the amount of training δ. Then,

$$u(a,b,h,\delta) = w^B + h\left(w^{OT} + f(a) - \frac{1}{b}\right) + x_2 + \mu(a,\delta)R(b) \qquad (1)$$

Our baseline scenario is that nobody wants to work overtime if there is no information asymmetry. The following assumption precisely describes such situation.

Assumption 1: $w^{OT} + f(a) - \frac{1}{b} < 0$ for any a and b

Let $\pi(a,b,h,\delta)$ be the firm profit when the worker's ability and commitment type is (a,b), he/she chooses the hours of work h in the first period and the employer provides the level of training δ. Then,

$$\pi(a,b,h,\delta) = x_1 - w^B + (m + f(a) - w^{OT})h + \mu(a,\delta)R(b_i) - \kappa(\delta)$$

(2)

Note that, if $m + f(a_k) - w^{OT} > 0$, overtime work is profitable. We consider both cases when it is profitable and unprofitable.

We solve the model backward. Since the worker does not make any decisions in the second period, we first solve for the firm's choice of training. If the firm can observe the worker's commitment type b (we later show that a separating equilibrium exist), it solves

$$\delta^*(a,b) = \underset{\delta}{arg\max}\,\mu\,(a,\delta)R(b) - \kappa(\delta)$$

From the concavity assumed, there is a unique solution. Although a is binary, we treat $\mu(a,\delta)$ as a continuous and differentiable function of a in order to simplify the notation in our comparative statics analysis. Then, there are some comparative statics results as follows:

- When $\mu_{a\delta} > (<) 0$, $\frac{\partial\delta^*}{\partial a} = -\frac{\mu_{a\delta}R(b)}{\mu_{\delta\delta}R(b) - \kappa''(\delta)} > (<) 0$.
- $\frac{\partial\delta^*}{\partial b} = -\frac{\mu_{\delta}R'(b)}{\mu_{\delta\delta}R(b) - \kappa''(\delta)} > 0$ because $R(b)$ is increasing in b.
- When $\mu_{a\delta} > (<) 0$, $\frac{\partial^2\delta^*}{\partial a\partial b} = \frac{\mu_{a\delta}R'(b)\kappa''(\delta)}{(\mu_{\delta\delta}R(b) - \kappa''(\delta))^2} > (<) 0$.

To simplify our notation, let us redefine $\mu^*(a,b) \equiv \mu(a,\delta^*(a,b))$. Then,

- $\frac{\partial\mu^*}{\partial a} = \mu_a(a,\delta^*) + \mu_\delta\frac{\partial\delta^*}{\partial a} > 0$ if $\mu_{a\delta} > 0$.
- $\frac{\partial\mu^*}{\partial b} = \mu_\delta\frac{\partial\delta^*}{\partial b} > 0$.
- $\frac{\partial^2\mu^*}{\partial a\partial b} = \mu_{a\delta}\frac{\partial\delta^*}{\partial b} + \mu_{\delta\delta}\frac{\partial\delta^*}{\partial a}\frac{\partial\delta^*}{\partial b} + \mu_\delta\frac{\partial^2\delta^*}{\partial a\partial b}$

  $= \frac{(1+\mu_\delta)\mu_{a\delta}R'(b)\kappa''(\delta)}{(\mu_{\delta\delta}R(b) - \kappa''(\delta))^2} > (<) 0$ if $\mu_{a\delta} > (<) 0$.

These analyses show that the sign of $\mu_{a\delta}$ determines the key comparative statics results. We consider both cases, $\mu_{a\delta} > 0$ (Case 1) and $\mu_{a\delta} < 0$ (Case 2). Before solving for the equilibrium, let us discuss the interpretation of the two cases.

First, $\mu_{a\delta} < 0$ is a reasonable assumption when worker homogeneity is desirable in the firm. For example, during Japan's post-war period of high growth, coordinating activities across firms and across businesses in growing markets was a primary task for managers rather than devising innovations and formulating strategies. The main focus of managerial training was placed on expanding cross-functional knowledge and developing relations with colleagues and customers. Training had an effect of offsetting ability differences and those who got promoted were not necessarily those who were smartest or those with highest leadership skills.

In contrast, the case of $\mu_{a\delta} > 0$ is a more realistic assumption when there are gains from worker heterogeneity. For example, in a world with a lot of uncertainty, having a small number of very innovative employees or those with high leadership skills will be better than having a large number of mediocre individuals. According to our interview with a senior manager of a leading training company in Japan, there has been a trend toward more human resources development budget spent on more selective training programs targeting future leaders aimed at developing strategies and leadership skills over the past decade or two. As such, the same training benefits high-ability workers more than low-ability workers. In other words, training and ability have now become complementary (rather than substitutable) inputs for promotion in recent years.

Working hours in separating equilibria

We next solve for a Perfect Bayesian equilibrium given the firm's feedback policy. In both cases, a separating equilibrium exists in which the commitment type is fully revealed to the employer. We state our first main results in Proposition 1 and 2.

Proposition 1 (information revelation case): *Suppose the firm has revealed the worker's ability to the worker himself/herself. In this subgame, there is a unique Perfect Bayesian equilibrium that satisfies D1 criterion. In this equilibrium, for k = H, L,*

- *types are fully separating,*
- ***e**ach commitment type chooses the minimum working hours that are not mimicked by another. Namely,* $\sigma(h_i^k|R, a_k,b_i) = 1$ *where* h_i^k *satisfies,*

$$\left(w^{OT} + f(a_k) - \frac{1}{b_{i-1}}\right) h_i^k + \mu^*(a_k, b_i) R(b_{i-1})$$
$$= \left(w^{OT} + f(a_k) - \frac{1}{b_{i-1}}\right) h_{i-1}^k + \mu^*(a_k, b_{i-1}) R(b_{i-1}) \quad (3)$$

- *the firm forms the belief* $\lambda(b_i|a_k, h, R) = 1$ *if* $h = [h_i^k, h_{i+1}^k)$ *for* $i = 1,2, \ldots n - 1$ *and* $\lambda(b_n|a_k, h, R) = 1$ *if* $h \geq h_n^k$,
- *and the firm provides training* $\delta^*(a_k, b_i)$ *if* $h = [h_i^k, h_{i+1}^k)$ *for* $i = 1,2, \ldots n - 1$ *and* $\delta^*(a_k, b_n)$ *if* $h \geq h_n^k$.

By iteration, we obtain

$$h_i^k = \sum_{j=2}^{i} \frac{\Delta_j^k}{A_j^k} R(b_{j-1})$$

where $h_1^K = 0$, $\Delta_j^k \equiv \mu^*(a_k, b_j) - \mu^*(a_k, b_{j-1})$ *is the marginal probability gain from mimicking,* $A_j^k \equiv \frac{1}{b_{j-1}} - w^{OT} - f(a_k) > 0$ *is the net cost of working overtime.*

There is a quite similar result under the non-information revelation policy.

Proposition 2 (non-information revelation case): *Suppose the firm does not reveal the worker's ability to the worker himself/herself. In this subgame, there is a unique Perfect Bayesian equilibrium that satisfies D1 criterion. In this equilibrium, for k = H, L,*

- types are fully separating,
- *the worker with the commitment type* b_i *chooses the minimum working hours that cannot be mimicked by the worker with* b_{i-1} *and the working hour is independent of the ability type. Namely,* $\sigma(h_i^N \middle| N, a_H, b_i) = \sigma(h_i^N \middle| N, a_L, b_i) = 1$ *where* h_i^N *satisfies,*

$$\left(w^{OT} + pf(a_L) + (1-p)f(a_H) - \frac{1}{b_{i-1}}\right) h_i^N + \left(p\mu^*(a_L, b_i) + (1-p)\mu^*(a_H, b_i)\right)R\left(b_{j-1}\right) = \left(w^{OT} + pf(a_L) + (1-p)f(a_H) - \frac{1}{b_{i-1}}\right) h_{i-1}^N + \left(p\mu^*(a_L, b_{i-1}) + (1-p)\mu^*(a_H, b_{i-1})\right)R\left(b_{j-1}\right) \qquad (4)$$

- *the firm forms the belief* $\lambda(b_i|a_k, h, N) = 1$ *if* $h = [h_i^N, h_{i+1}^N)$ *for* $i = 1,2, \ldots n-1$ *and* $\lambda(b_n|a_k, h, N) = 1$ *if* $h \geq h_n^N$,
- *and the firm provides training* $\delta^*(a_k, b_i)$ *if* $h = [h_i^N, h_{i+1}^N)$ *for* $i = 1,2, \ldots n-1$ *and* $\delta^*(a_k, b_n)$ *if* $h \geq h_n^N$.

By iteration, we obtain

$$h_i^N = \sum_{j=2}^{i} \frac{\Delta_j^N}{A_j^N} R\left(b_{j-1}\right)$$

where $h_1^N = 0, \Delta_j^N \equiv p\left(\mu^*\left(a_L, b_j\right) - \mu^*\left(a_L, b_{j-1}\right)\right) + (1-p)\left(\mu^*\left(a_H, b_j\right) - \mu^*\left(a_H, b_{j-1}\right)\right)$ *is the marginal probability gain from mimicking,* $A_j^N \equiv pA_j^L + (1-p)A_j^H = \frac{1}{b_{j-1}} - w^{OT} - pf(a_L) - (1-p)f(a_H) > 0$ *is the net cost of working overtime.*

The proofs are presented in the Appendix.

D1-criterion is necessary to rule out equilibria that are supported by unreasonable beliefs. For example, consider a pooling equilibrium where all types choose $h = 0$. This could constitutes a perfect Bayesian Nash equilibrium with the belief that anybody who deviates from this action has the lowest commitment. However, suppose somebody deviates from this equilibrium strategy (*i.e.* $h = 0$) and work overtime. The type of workers who are most likely to benefit from this deviation is the one who has the highest commitment (i.e. b_n or lowest disutility). If the employer believes that only those with b_n deviate, they are better off by deviating than in the equilibrium if the overtime work hours are not excessively high. Hence, the above pooling equilibrium does not survive D1-criterion.

How D1 criterion works can be illustrated by Figure 2. Consider three commitment types b_1, b_2, b_3, and b_4, and suppose that their equilibrium strategies are to choose h_1^k, h_2^k, h_3^k, and h_4^k, respectively. The indifference curves for each type are drawn in the chart. Note that the slopes of the indifference curves are decreasing in the worker's commitment type because, from equations (3) and (4), $\frac{d\delta}{dh} = \frac{\frac{1}{b_i} - w^{OT} - f(a_k)}{\frac{\partial\mu}{\partial\delta} R(b_i)}$, which is decreasing in b_i. Suppose a worker whose commitment type is unknown for the employer has worked *x* hours where $h_2^k < x < h_3^k$. Which type of workers are more likely to deviate from their equilibrium strategy and choose *x*? Note that the commitment type *i* will deviate if the firm's resultant choice of training exceeds P_i, which is the point on type *i*'s indifference curve corresponding to the hours worked *x*. Namely, the commitment type *i* is indifference between $\delta^* \,(a, b_i)$ and P_i. As the figure shows, if some types find it beneficial to deviate to working *x* hours, it must be type 2 because $P_2 < P_3 < P_1$. This means that the commitment type 2 is most likely to deviate.

If deviation comes only from the commitment type 2, the firm will offer the original level of training $\delta^*(a, b_2)$. Given that, deviating to *x* from h_2^k is not beneficial for type 2, because the training the firm will offer is not changed even though he/she increases working hour. In the equilibrium in Proposition 1 and 2, the same as above holds true for any off-equilibrium working hour, which means that the equilibrium in Proposition 1 and 2 survive D1-criterion.

Next, we characterize the working hours in the equilibrium.

Lemma 1 When Assumption 1 holds, $h_i^L < h_i^N < h_i^H$ for any i in case 1 (i.e. $\mu_{a\delta} > 0$) and $h_i^H < h_i^N < h_i^L$ for any i if $f(a_H) - f(a_L)$ is small enough in case 2 (i.e. $\mu_{a\delta} < 0$).

Proof: We first show $\frac{\Delta_j^L}{A_j^L} - \frac{\Delta_j^H}{A_j^H} < 0$ for any j in case 1.

- $A_j^L > A_j^H$ from the assumption $f(a_L) < f(a_H)$.
- $\Delta_j^L < \Delta_j^H$ if $\frac{\partial^2 \mu^*}{\partial a \partial b} > 0$ (case 1). Then, $\frac{\Delta_j^L}{A_j^L} - \frac{\Delta_j^H}{A_j^H} < 0$. Thus,

$$h_i^L - h_i^N = \sum_{j=2}^{i} \left[\frac{\Delta_j^L}{A_j^L} - \frac{\Delta_j^N}{A_j^N}\right] R(b_{j-1}) = (1-p) \sum_{j=2}^{i} \frac{A_j^H}{A_j^N} \left[\frac{\Delta_j^L}{A_j^L} - \frac{\Delta_j^H}{A_j^H}\right] R(b_{j-1}) < 0$$

$$h_i^H - h_i^N = \sum_{j=2}^{i} \left[\frac{\Delta_j^H}{A_j^H} - \frac{\Delta_j^N}{A_j^N}\right] \Delta_R(b_{j-1}) = p \sum_{j=2}^{i} \frac{A_j^L}{A_j^N} \left[\frac{\Delta_j^H}{A_j^H} - \frac{\Delta_j^L}{A_j^L}\right] R(b_{j-1}) > 0$$

Similarly, in case 2, when $f(a_H) - f(a_L)$ is small enough, $\frac{\Delta_j^L}{A_j^L} - \frac{\Delta_j^H}{A_j^H} > 0$. This concludes the proof. Q.E.D.

An intuition of Lemma 1 is as follows: If training and worker ability are complementary inputs for promotion (i.e., case 1), the workers with high ability receive a greater amount of training and enjoy a higher chance of promotion than the workers with low ability. Furthermore, and the marginal effect of worker commitment (b) on the promotion prospect is higher for the high-ability workers than the low-ability workers (i.e., $\frac{\partial^2 \mu^*}{\partial a \partial b} > 0$) This implies that the workers who become aware of their high ability through information revelation have greater incentives to mimic the higher-commitment workers. Then, to prevent lower-commitment workers from mimicking their working hour, high-commitment workers with high ability have to work longer hours than the high-commitment workers with low ability. In case when the workers are not informed of their ability, the hours of work are chosen between the levels for the high-ability workers and the low-ability ones when the worker ability is revealed.

The opposite is true if training and worker ability are substitutes (i.e., case 2) but there is a countervailing effect. Note that the worker with high ability can learn more from working when he/she is young than the low ability worker. Therefore, the worker who becomes aware of his/her high ability through information revelation has greater incentive to work long hours, though the worker who becomes aware of his/her low ability has smaller incentive to do so. If the difference in the speed of learning-by-doing is very large, this effect from heterogeneous learning-by-doing on the working hours could more than offset the other effect through the firm's training policy that is designed to level ability differences in case 2. The next proposition compared the average working hours between the information revelation and non-information revelation policies.

Proposition 3: Under Assumption 1, $p \sum_{i=1}^{n} q_i \, h_i^L + (1-p) \sum_{i=1}^{n} q_i \, h_i^H > \sum_{i=1}^{n} q_i \, h_i^N$ holds in case 1. In case 2, on the other hand, $p \sum_{i=1}^{n} q_i \, h_i^L + (1-p) \sum_{i=1}^{n} q_i \, h_i^H <$

$\sum_{i=1}^{n} q_i\, h_i^N$ if $f(a_H) - f(a_L)$ is small enough.

Proof: In case 1, since $A_j^H - A_j^L = f(a_L) - f(a_H) < 0$ and $\frac{\Delta_j^L}{A_j^L} - \frac{\Delta_j^H}{A_j^H} < 0$ for any j,

$$ph_j^L + (1-p)h_j^H - h_j^N = p(1-p)\sum_{j=2}^{i} \frac{A_j^H - A_j^L}{A_j^N}\left[\frac{\Delta_j^L}{A_j^L} - \frac{\Delta_j^H}{A_j^H}\right] R(b_{j-1}) > 0$$

Therefore, $p\sum_{i=1}^{n} q_i\, h_i^L + (1-p)\sum_{i=1}^{n} q_i\, h_i^H > \sum_{i=1}^{n} q_i\, h_i^N$. The proof is similar for case 2.

Q.E.D.

This result comes from the complementarity among the ability, training, commitment, and working hours. When information on worker ability is revealed, knowing that a greater amount of training is provided, more capable workers work longer to signal their commitment so that the firm further increases the amount of training securing the opportunity of promotion. Less capable workers work less but the average working hours are still higher than when the ability information is not revealed. The result changes when $\mu_{a\delta} < 0$ under which ability and training are substitutes in producing promotion. Although the worker still has an incentive to work longer to signal their commitment, this signaling has a lower return for capable workers. Therefore, the average working hours are shorter when the ability information is revealed.

Optimal Feedback Policy

Finally, we examine the optimal feedback policy for the firm. Let $\pi^s(b_i)$ be the expected equilibrium profit of the firm when the firm employs type-b_i worker under information revelation policy *s*. Comparing $\pi^R(b_i)$ with $\pi^N(b_i)$, we get

$$\pi^R(b_i) - \pi^N(b_i) = p(m + f(a_L) - w^{OT})(h_i^L - h_i^N) + (1-p)(m + f(a_H) - w^{OT})(h_i^H - h_i^N)$$

$$= p(1-p)(f(a_H) - f(a_L)) \times \sum_{j=2}^{i} \frac{2w^{OT} - m - \frac{1}{b_{j-1}}}{\frac{1}{b_{j-1}} - w^{OT} - pf(a_L) - (1-p)f(a_H)} \left[\frac{\Delta_j^L}{A_j^L} - \frac{\Delta_j^H}{A_j^H}\right] R(b_{j-1})$$

The following proposition is straightforward from this equation.

Proposition 4: $\sum_i q_i \pi^R(b_i) < \sum_i q_i \pi^N(b_i)$ if $2w^{OT} - m - \frac{1}{b_i} > 0$ for any i, in case 1; or if $2w^{OT} - m - \frac{1}{b_{i-1}} < 0$ for any i and $f(a_H) - f(a_L)$ is small enough, in case 2.

Note that $2w^{OT} - m - \frac{1}{b_i} > 0$, the condition in the first case does not necessarily violate Assumption 1 ($w^{OT} + f(a) - \frac{1}{b} < 0$ for any a and b), which requires the overtime hourly wage not to be too high, if both the first-period marginal productivity and the learning-by-doing effect are sufficiently low.

Our results are summarized in Table 1. Note that there are two cases in which non-information revelation is the optimal feedback policy for the firm. First, it is so when: (1) the training program is equal and inclusive focusing more on leveling individual performance differences, resulting in offsetting ability differences (*i.e.* $\mu_{a\delta} < 0$); (2) learning by doing is not drastically different between more capable and less capable workers (*i.e.* $f(a_H) - f(a_L)$ is small); and (3) the overtime hourly wage w^{OT} is sufficiently low. Another case is when: (1) the training program is selective focusing more on developing capable leaders (*i.e.* $\mu_{a\delta} > 0$); and (2) the overtime hourly wage w^{OT} is sufficiently high. In all other quadrants, revealing the worker ability is optimal. As we have argued earlier, during the post-war period of high growth, training opportunities were offered equally to all employees and helped to improve the standard and to level performance differences. Furthermore, many workers voluntarily worked without overtime pay because such efforts were rewarded by promotion. According to our model, non-information revelation was presumably optimal in those days.

In recent years, we believe that there have been two major changes. First, our interview with a senior manager at a leading training company and our casual conversations with Japanese HR managers reveal that Japanese firms are reducing the budget for periodical cohort training programs and increasing more selective training programs designed to make high-performers more productive. If our interpretation is correct, we now have $\mu_{a\delta} > 0$. Second, overtime regulations are being more strictly enforced and fewer non-managerial workers are willing to work voluntarily without overtime pay because of limited promotion opportunities. According to Table 1, these changes may still imply that non-information revelation is optimal.

There is another view, however. Overtime pay is not calculated for an expanding set of workers: (1) discretionary working system that applies to professionals (i.e. similar to exempts in the US but limited to much narrower job categories such as managers and researchers); and (2) deemed overtime pay system under which fixed overtime pay is paid out to those whose working hours are hard to monitor such as sales people. White-collar exemption, which is comparable to exempts in the US, is also being considered to exempt high-income white-collar workers from overtime pay regulations. If so, at least for those whose marginal pay for overtime work is zero, information revelation may be optimal lately.

Efficiency

Note that, if there is no asymmetric information, the working hours should be minimal $h(a, b_i, s) = \underline{h}$ for all i. Then, the efficient level of training is obtained by solving

$$\max_{\delta} 2\mu\,(a, \delta)R(b) - \kappa(\delta).$$

Given the information asymmetry, the second best can be achieved by a long-term contract. The efficient information revelation policy minimize the total cost of working hours $\sum_{i=1}^{n} q_i \frac{h_i}{b_i}$ with the incentive compatibility constraints for each type. The employer should solve $\delta^*(a, b) = arg\max_{\delta} 2\mu\,(a, \delta)R(b) - \kappa(\delta).$ Without long-term contracts feasible, the total cost of working hours won't be minimized implying that working hours are always excessive.

Heterogeneity among work groups

In the rest of the discussion, we assume $\mu_{a\delta} > 0$ because we believe that the inequality best describes the current situation of human resource development policy among Japanese firms. With this condition, revealing the firm's private information on the worker's ability to himself/herself encourages high-ability workers work longer but low-ability workers work shorter.

Proposition 4 implies that $\pi^R(b_i) > \pi^N(b_i)$ is more likely for low commitment types whereas $\pi^R(b_i) < \pi^N(b_i)$ is more likely for high commitment types. We illustrate this comparison in Figure 3. Therefore, if low-commitment (high-commitment) types have a high share of the total population, information revelation (non-information revelation) is optimal. Suppose it is feasible to employ different information revelation policies for demographically clearly distinct groups: men and women. The pattern we observe in the gendered division of household labor in Japan appears to suggest that the distribution of the cost of working long hours is quite different between men and women. A number of studies have reported that among industrialized nations Japanese husbands share housework with their wives the least, and provided evidence that social contexts determined at the national level such as gender ideology play significant roles in explaining cross-national variations in gender division of household labor (Davis and Greenstein, 2004; Fuwa, 2006; Greenstein, 2009; Qian and Sayer, 2015). Thus, it is plausible that male workers in Japan are more likely to enjoy greatly reduced cost of long working hours due to social norm favoring the gendered division of household labor. The other side of the coin is that female workers in Japan are more likely to face greatly elevated cost of long working hours due to the same social norm. As such, it is probably safe to assume that work hour flexibility (i.e. parameter b) has larger density for higher values with men whereas it has larger density for lower values with women—Japan is subject to great inequality in the cost of working long hours between men and women.[6] In sum, low-commitment types with higher cost of working long hours are likely to have a higher share among women than among men in Japan.

It follows that adopting information concealment for women is likely to be suboptimal because there are more "low commitment" types among women than among

[6] For stylized facts about the status of Japanese women in the labor market in international perspective, see Steinberg and Nakane, 2012

men. If firms can adopt different information revelation policies for men and women, they might adopt information revelation for women while information concealment for men. This is especially likely in the current policy environment where the Abe administration is pressing Japanese firms to increase the number of female managers. Then, we would observe distinctly different promotion patterns between male and female workers. Figure 4 illustrates such a likely pattern. If women are informed of their ability, more capable women work longer hours, knowing that the return to doing so is higher than otherwise while less able women work shorter hours, knowing that their chance of promotion is lower than otherwise (thus any voluntary overtime work is likely to be a waste of time). There will be a substantial difference in the promotion rate between the two groups. In contrast, there will be little dichotomy between high- and low-ability men since there is no information revelation for men, and men decide on their working hours without knowing their ability types. To the extent to which commitment types and ability types are uncorrelated, it follows that the slope of the hours-promotion profile should be flatter for men than for women.

The above discussion generates the following testable implication.

Hypothesis: The incidence of promotion should be more highly correlated with working hours for women than for men.

4. Evidence from an Econometric Case Study of a Large Japanese Firm

We conduct an econometric case study of a large Japanese manufacturing firm, MfgJapan,[7] and provide evidence in support of the above hypothesis—the relationship

[7] Our confidentiality agreement with the firm prohibits us from revealing the actual name of the firm as well as further details on their product lines.

between working hours and promotion rates is stronger for women than for men. MfgJapan is a large manufacturing firm in Japan and is listed in the first section of Tokyo Stock Exchange. It employs 20,000 workers worldwide (of which 6,000 are employed within Japan).[8] We are granted full access to detailed personnel records between FY2005 and FY2013 on all domestic employees directly employed by the firm. It is one of the first firms which participated in our Industry-Academia-Government Personnel Data Repository (IAGPDR) project.[9]

The personnel records include standard biographical data such as gender, date of birth, date of hire, nationality, education, and marital status among others. Most importantly the personnel records provide detailed job assignment information on each employee annually for 2005-2013. As described in some detail below, the longitudinal job assignment data, combined with our institutional knowledge of the firm's promotion ladder, allow us to define promotions (our key dependent variable) unusually precisely. A potentially serious measurement error concerning promotions has been a significant limitation of prior studies. For instance, as pointed out by Blau and DeVaro (2006), it is plausible that men regard certain job changes as promotions, while women may not consider the same job changes as promotions. By using an institutionally-informed definition of promotion, we minimize such measurement error--- all employees at JfgJapan, including both men and women, are fully aware of their promotion ladder and

[8] Non-standard employees such as part-time and contract workers are not included in our data, for they are off the career ladder.

[9] Our partner, Works Applications, Co. is a major ERP software package provider in Japan with approximately 300 listed firms in their user network. The Research Institute of Economy, Trade and Industry (RIETI), a think tank established by the Ministry of Economy, Trade, and Industry for policy-oriented research provided us with technical resources to store the data in a secure environment. In this project, almost all personnel records stored in the firm's human resource management system were deposited into RIETI's high-security server (with the exception of sensitive identity information such as names and addresses). Researchers analyze the data remotely using RIETI's virtual private network.

there is clear consensus among all employees concerning which job assignment changes are promotion).[10]

Furthermore, the personnel records also include accurate data on actual working hours (our primary explanatory variable) for all employees over the same time period. Especially we make use of unusually detailed information on who takes a parental leave and when, and when he/she returns from his/he parental leave, and calculate actual working hours which account for gender differences in lost working hours due to parental leave.

Before describing MfgJapan's promotion ladder in detail, we provide some key background statistics on the labor force at MfgJapan. Table 2 shows the proportion of single men, married men, single women, and married women with and without young children over 2005-2013.[11] The number of children and their birth years are not included in the personnel data directly. However, the Japanese government mandates every pregnant female worker to take a maternity leave of eight weeks immediately following the delivery of her child. Fortunately the incidence of such a government-mandated maternity leave is recorded in the personnel data. We use this information and define a female worker with young children as a female worker who

---

[10] Many prior studies define promotion as a discrete wage increase and/or a change in standard occupational codes (e.g., Blau and DeVaro, 2006; McCue 1996; Cobb-Clark 2001; Smith, Smith, and Verner, 2013; Frederiksen and Kato 2016; and . Most recently

[11] We include a very small number of single women with young children in this category, for they will be subject to at least the same level of motherhood/ immobility penalty as married women with young children. In principle if a woman with young children is entering the firm, our definition of married women with young children will fail to count her unless she has another child birth after entering the firm and hence take a mandatory maternity leave. Based on our interviews with the firm's HR managers, however, we are reasonably confident that our definition of married women with young children captures most if not all of married women with young children at the firm.

took at least one mandatory maternity leave since FY1999, the year in which the Japanese government introduced a law mandating the maternity leave.[12]

The proportion of female employees hovered around 12-13% during the period under investigation, which is lower than the 16-18% average of large manufacturing firms with 1,000 or more employees in Japan, according to the Basic Survey on Wage Structure conducted by the Ministry of Health, Labor and Welfare. As in the case of most large Japanese firms in manufacturing, overall separation rates at the firm are low, and we found little significant correlation between separation rates and individual worker characteristics.[13]

Table 3 shows the raw (unadjusted for observable individual characteristics) gender pay gap. Compared to married men, single men, single women, married women without young children, and married women with young children earn 26%, 31%, 27%, and 40% less, respectively, on average during the 2005-2013 period.[14] Since married men are 5-10 years older than the other groups on average, a substantial portion of the gaps may come from differences in the amount of human capital accumulated through working in the firm. Furthermore, women, especially those who are married with young children, work substantially fewer hours than men. Single women, married women without young children, and married women with young children work 7%, 10%, and 21% fewer hours, respectively, than married men, on average during the 2005-2013 period, while single men do not work significantly different hours than married men.

---

12 For example, a pregnant woman who delivered her child in 1999 is considered having a young child of age 5 in 2004. In 2013, she is considered having a young child of age 13 (incidentally this child is the oldest "young child" in our dataset).

13 These results as well as all other unreported results are available upon request from the corresponding author.

14 There is a companion paper which analyzes the same data and studies the nature and scope of the gender pay gap of the firm (Kato, Kawaguchi, and Owan, 2013)

Obviously if the worker takes a maternity or parental leave, her working hours during her leave are zero. Our calculation of annual working hours exclude those who are not working in any month in the year. As such, low working-hour numbers for women with young children are not the result of her maternity/parental leave.[15]

MfgJapan's promotion ladder is based on a job grade system under which each job is assigned a specific job grade level, based on an evaluation of the job content. Figure 5 shows MfgJapan's promotion ladder with solid lines indicating typical paths, and dotted lines possible yet less typical paths. As in the case of most Japanese firms in manufacturing, there are two tracks in MfgJapan's promotion ladder: (i) the white-collar track for college graduates; and (ii) the blue-collar track for high-school graduates. Initially both college graduates on the white-collar track and high-school graduates on the blue-collar track start at the same entry level, J1. Both groups then move to J2. Since a move from J1 to J2 is automatic for most workers, it is typically not considered promotion. The next level is SA for college graduates and J3 for high-school graduates, followed by SB for college graduates and J4 for high-school graduates. The management rank begins at G6 for college graduates and JH for high-school graduates. Typically JH is the highest level that high-school graduates on the blue-collar track can reach although it is not impossible for high-school graduates to switch to the management track by moving to SA, SB, and G6. Beyond G6, a select group of college graduates climb up the ladder to G5. G6/JH and G5 are supervisory management positions with no formal authority to evaluate subordinate performance. A further select

[15] MfgJapan uses a policy of Transition Period Part-Time Schedule for female workers with young children which allows such female employees with young children to reduce their work hours to 6 hours per day for a fixed time period. The provision of this policy or an equivalent substitute is required by the 2002 Amendment to the 1992 Child Care and Family Care Leave Act. See Asai (2015) for the efficacy of parental policy reforms in Japan.

group of those in G5 are promoted to G4, G3, G2, and G1 (middle and top management positions with formal authority to evaluate subordinate performance). Finally administrative support staff members who typically have high-school or two-year college degrees can move up to J2 and J3 levels, but it is extremely rare for them to get promoted to managerial positions.

Note that in 2011 MfgJapan changed its job grade system to a job skill system under which pay is linked to the manager's assessment of the worker's skill level rather than the job content. The change applied only for non-managerial workers, and the same job grade system continued for managerial workers. Based on our extensive interviews with HR managers at MfgJapan and our observation of the actual transition pattern, we were able to link the job skill system to the job grade system for non-managerial workers.[16] However, FY2011, being the transition period from the job grade system to the job skill system, was a somewhat confusing year for all employees at MfgJapan, and it is plausible that the 2011 data may be contaminated by the change and subject to significant measurement errors. To this end, for the following econometric analysis, we exclude the 2011 data.[17]

Table 4 shows the distribution of the workforce across job levels, gender and marital status in FY2005, FY2009 and FY2013. Women in higher level positions are still quite limited in number despite some improvement from FY2005.[18] For example,

---

[16] Specifically skill grades 83-85 correspond to J1 and J2; skill grades 42,43, and 73 to SA and J3; skill grades 41 and 63 to SB and J4; and skill grade 54 to G6 and JH.

[17] Reassuringly we repeated the same analysis, including the 2011 data, and found little change in our key results. Alternatively we repeated the same analysis by excluding not only 2011 but also all subsequent years (essentially using only data under the job grade system, 2005-2010), and again reassuringly we found similar results although the estimates are somewhat less precise as expected from a decrease in the number of observations.

[18] The lack of notable improvement from FY2009 to FY2013 is partly due to the spin-off in early 2010s of its R&D division which had many female managers.

as of FY2013, only 4.5% of managers (G6 and beyond) are women. Given that 13.3% of its workforce is women, this share of women in managerial positions is disproportionately low.

Given the job ladder at the firm presented in Figure 5, we can estimate a logit model of the probability of promotion, conditional on the current job level, as a function of a number of explanatory variables. Specifically, we begin with estimating a baseline logit model:

$$\ln[\Pr(PROMOTION_{it})/1-\Pr(PROMOTION_{it})]= \alpha + \beta female_i + \gamma marriage_{it} + \delta female_i * marriage_{it} + \eta \ln(hours_{it-1}) + \sigma \ln(hours_{it-1}) * female_i + \mathbf{Z_{it}\mu} + u_{it} \quad (1)$$

where $PROMOTION_{it}$=1 if worker *i* is promoted to a higher job level in year *t*, zero otherwise; $female_i$=1 if worker *i* is female, zero otherwise; $marriage_{it}$=1 if worker *i* is married in year *t*, zero otherwise; and $hours_{it-1}$=the number of hours worked for worker *i* in year *t*-1. For $\mathbf{Z_{it}}$, we control for the worker's age, tenure, education, current job level, and whether or not the worker currently work in the headquarters. Since we pool data for different years, year dummy variables are also included to control for common shocks that affect all employees (e.g., macroeconomic shocks).

Of prime interest is the sign and significance of the estimated coefficient on $\ln(hours_{it-1})*female_i$. The positive and statistically significant coefficient supports our hypothesis. Columns (1) and (3) of Table 5 presents the logit estimates of Eq. (1) for all employees and college-graduates respectively.

We find consistently for all specifications that the estimated coefficients on $\ln(hours_{it-1})$ are positive and statistically significant at the 1 percent level, indicating that hours worked in the previous year are positively associated with the odds of promotion. Most important (for the objective of the paper) is that the estimated coefficients on the interaction term, ln($hours_{it-1}$)*$female_i$, are also positive and statistically significant at the 1 percent level, supporting our testable hypothesis developed in the previous section---the positive associations between hours worked and the odds of promotion are stronger for women than for men,.

Using the U.S. National Longitudinal Survey of Youth, Gicheva (2013) finds that the relationship between hours worked and wage growth is non-linear—only workers who put forth more hours than a certain threshold see the positive impact of working additional hours. It is possible that Eq (1) may be misspecified, and that the interaction term involving ln($hours_{it-1}$) and $female_i$ may be capturing such nonlinearity. To rule out this possibility, we re-estimate the model, replacing ln($hours_{it-1}$) with a set of dummy variables – $hours18less_{it-1}$ takes a value of one if worker i's hours worked in year t-1 is less than 1800 hours per year, zero otherwise ($hours18_{it-1}$ is omitted as a reference category); $hours18_{it-1}$ takes a value of one if worker i's hours worked in year t-1 is 1800 or more but less than 2000, zero otherwise; $hours20_{it-1}$ takes a value of one if worker i's hours worked in year t-1 is 2000 or more but less than 2200, zero otherwise; $hours22_{it-1}$ takes a value of one if worker i's hours worked in year t-1 is 2200 or more but less than 2400, zero otherwise; and $hours24_{it-1}$ takes a value of one if worker i's hours worked in year t-1 is 2400 or more, zero otherwise.

Columns (2) and (4) of Table 5 present the logit estimates of Eq. (1) which allows for a non-linear relationship between the odds of promotion and working hours. Reassuringly all of the interaction terms involving $female_i$ and each dummy variable indicating a range of working hours are still positive and statistically significant at the 1 percent level.

To demonstrate the economic significance of the estimated coefficients on a set of working hours dummy variables and their interaction with gender, we calculate for men and women separately the predicted odds of promotion for each range of working hours in the previous year by integrating over the remaining covariates. The resulting hours-promotion profiles along with their 95 percent confidence intervals are shown in Figure 6. As expected, the hours-promotion profiles for all employees as well as for college graduates are indeed steeper for women than for men. Moreover, the figure suggests a non-linear relationship between working hours and the odds of promotion for women---a considerable jump in the odds of promotion when women's working hours exceed 2400 hours per year. No such threshold is detected for men.

It is worth noting that the predicted odds of promotion are generally lower for women than for men for those whose hours worked are less than 2000 hours per year, and that 79 percent (62 percent) of women (college graduate women) and 42 percent (40 percent) of men (college graduate men) belong to these categories. The predicted odds of promotion for women, however, are significantly higher than for men when hours worked exceed 2400 hours per year although only 1.5 percent (3.7 percent) of women (college graduate women) work such long hours compared with 7.9 percent (12.1 percent) of men (college graduate men) who do so.

There is, however, an alternative interpretation of the observed stark difference between men and women in the association between hours and the subsequent promotion rates. Many scholars including Blau and Kahn (2000) and Goldin (2014) have reported that women were concentrated in low-paying jobs such as administrative support and service occupations.[19] Some college-educated women may be assigned to

[19] Historically in Japan, women were typically assigned to administrative support and clerical jobs called *ippan-shoku* (general job), while college educated men were assigned to

jobs with little training and promotion prospects while others join college-educated men for management trainee jobs. If women in the former group have no reasonable promotion prospects, they will not work long hours. Combining members from the two groups may result in a high correlation between the odds of promotion and hours worked.

To rule out the possibility of this alternative interpretation of job segregation by gender, we further estimate the model with organizational unit fixed effects. At MfgJapan, a typical organization unit is a section of 3-9 people for white-collar workers and a group of people working on the same production line or support function (up to 90) for blue-collar workers. The organizational unit fixed effects can account for the gender gap in promotion caused by such job segregation within the firm, to the extent to which good jobs and bad jobs are separated into different organizational units.

As an effective way to control for such organizational unit fixed effects in the logit framework we employ a fixed effects logit model in which organizational unit fixed effects are accounted for by the number of promotion within each unit. Reassuringly the results change little when we use such a fixed effects logit model instead of a standard logit model. The observed high correlation between the odds of promotion and hours worked among women does not appear to be driven by job segregation by gender within the firm.

4. Conclusions

This paper has presented a novel model of promotion within the firm which sheds

management trainee positions called *sogo-shoku* (composite job). Although such differentiation was prohibited by the EEOA, a similar pattern of task assignment may still remain.

new light on the interplay between working hours and the odds of subsequent promotion. The key feature of our model is the coexistence of two different sources of asymmetric information: (i) the worker's cost of long working hours: and (ii) the worker's OJT ability (the worker's ability to accumulate valuable human capital on the job through learning by doing). On the one hand, the worker's cost of working long hours is known to the worker but not to the firm at the beginning of his/her tenure at the firm---the worker's cost of working long hours is his/her private information. On the other hand, the worker's OJT ability is initially unknown to both parties when he/she starts working at the firm, yet it is revealed to the employer earlier than the worker himself/herself at a relatively early stage of his/her tenure.

The firm prefers to train employees who are willing to work long hours because a higher rent is generated by promoting such workers. Thus, the firm provides the worker with managerial training only after observing the employee's hours worked. The firm's decision to provide training also depends on its private information about the worker's OJT ability, which also affects the second period productivity when the worker gets promoted. Upon completion of training, the firm then promotes the worker.

Of primary interest is the firm's decision to reveal or conceal its private information on the worker's OJT ability, and we have shown that the firm's optimal choice of information revelation policy crucially depends on three factors. First, the nature of training matters. If the training programs are designed to target high-performers and thus complement worker ability in raising the odds of promotion, information revelation of worker ability motivates high-ability workers to work longer. In contrast, if the training programs mainly target the left tail of the ability distribution of workers, aimed at developing uniform skill distribution among workers, training and ability will be substitutable inputs in raising the odds of promotion. It follows that information revelation will discourage high-ability workers from working longer.

Second, overtime hourly pay has been shown to determine both the signaling cost for the worker and the marginal return to information revelation for the firm. If overtime pay is zero, the firm will always want to maximize working hours, and long working hours will become a more effective signal. In contrast, if overtime pay is very high, the firm will try to minimize overtime work hours, and the low-commitment worker with higher cost of long working hours will have a stronger incentive to mimic the high-commitment workers with lower cost of long working hours.

Third, we have demonstrated that the distribution of the cost of long working hours also significantly affects the marginal return to information revelation. Assuming that training complements ability in raising the odds of promotion, when the firm adopts the information revelation strategy, the high-ability worker, having being informed of his/her high ability, will be more likely to send a signal of his/her commitment to the firm by working long hours which will earn him/her promotion. In contrast, the low-ability worker, having been made aware of his/her low ability, will be less likely to do so. In other words, information revelation will lead to an increase in working hours for high-ability workers and a decrease in working hours for low-ability workers. Total changes in working hours (and hence the *ex ante* efficiency) caused by information revelation then depends on the balance of the increase in working hours of high-ability workers and the decrease in working hours of low-ability workers.

When social norm imposes greater demand on women for household production than on men, the distribution of the cost of long working hours is likely to be more right-skewed for women than for men. As a result of the right-skewed distribution of the cost of working long hours, as compared to male workers, the increase in working hours of high-ability female workers is likely to be more restrained by their high cost of working long hours, while the decrease in working hours of low-ability female workers is apt to be more forthcoming due to their high cost of working hours. As such, in terms of the *ex ante* efficiency loss due to excessively long working hours, or intensified rat race, information revelation is more efficient when applied to women than to men. Being cognizant of such a gender-based difference in the distribution of the cost of long working hours, the firm may find it optimal to use the information revelation strategy for women, and the information concealment strategy for men.

Using this model, we have demonstrated that, under a reasonable set of conditions, the correlation between working hours and the odds of subsequent promotion will be stronger for women than for men. Our analysis of longitudinal personnel data from a large Japanese manufacturing firm has yielded rigorous econometric evidence that is consistent with this prediction.

Appendix

We will provide the proof of proposition 1. Preparing for the proof, we introduce some notations and provide useful lemmas.

Fix the worker's type k. Consider the set of working hour $H = \{h_1, \dots, h_m\}$ such that $h_1 < h_2 < \dots < h_m$ which are chosen by ability-a_k worker in an equilibrium with strictly positive probability and the amount of training $\delta^k(h)$ that the firm provides to the worker who choses $h \in H$. For $h, h' \in H$, it have to be satisfied that if $h' < h$ then $\delta^k(h') < \delta^k(h)$ (otherwise no one chooses h).

We say an equilibrium is a partition equilibrium if both type-b' and type-b'' worker choose the same working hour h and $b'' > b'$ then any type-b worker such that $b' < b < b''$ also chooses h. The following lemma says any equilibrium is either completely separating equilibrium or partition equilibrium.

Lemma A1. Any pooling equilibrium is the partition equilibrium.

proof. Consider equilibrium working hours h, h' and h'' such as $h' < h < h''$. Because $\frac{\partial^2 u}{\partial \delta \partial b} > 0$ and $\frac{\partial^2 u}{\partial h \partial b} > 0$, if type-$b'$ prefers $(h, \delta^k(h))$ to $(h', \delta^k(h'))$, for $b > b'$ type-b also prefer $(h, \delta^k(h))$ to $(h', \delta^k(h'))$. Also that, if type-b'' prefers $(h, \delta^k(h))$ to $(h'', \delta^k(h''))$, for $b < b''$ type- b also prefer $(h, \delta^k(h))$ to $(h'', \delta^k(h''))$. Thus, if type-b' and type-b'' choose h, type-b chooses h. □

Consider an off-path working hour h_d. Let x_i be the equilibrium working hour which type-b_i worker chooses. Because $\frac{\partial u}{\partial \delta} > 0$ and $\frac{\partial u}{\partial h} < 0$, there exist an unique training $\delta_i^{id}(h_d)$ such that type-b_i worker is indifferent between his/her equilibrium outcome $(x_i, \delta(x_i))$ and $(h_d, \delta_i^{id}(h_d))$. The following lemma is useful in applying D1 criterion.

Lemma A2. (i) For $h < h_d < h^{'}$, $\delta_i^{id}(h_d) < \delta_j^{id}(h_d)$ for $i,j \in (l|x(l) \leq h)$ and $i > j$, and $\delta_j^{id}(h_d) < \delta_i^{id}(h_d)$ for $i,j \in (l|x(l) \geq h^{'}$) and $i > j$, (ii) for $h_d < h_1$, $\delta_1^{id}(h_d) < \delta_i^{id}(h_d)$ for any $i \neq 1$, and (iii) for $h_m < h_d$, $\delta_n^{id}(h_d) < \delta_i^{id}(h_d)$ for any $i \neq n$.

proof. The former part of (i) holds from the following; for $i,j \in \{l|h(l) \leq h\}$ and $i > j$,

$$0 = u(a_k, b_i, h_d, \delta_i^{id}(h_d)) - u(a_k, b_i, x_i, \delta^k(x_i))$$

$$> u(a_k, b_j, h_d, \delta_i^{id}(h_d)) - u(a_k, b_j, x_i, \delta^k(x_i))$$

$$\geq u(a_k, b_j, h_d, \delta_i^{id}(h_d)) - u(a_k, b_j, x_j, \delta^k(x_j))$$

$$= u(a_k, b_j, h_d, \delta_i^{id}(h_d)) - u(a_k, b_j, h_d, \delta_j^{id}(h_d)).$$

The first inequality follows from $\delta_i^{id}(h_d) > \delta(x_i)$, $\frac{\partial^2 u}{\partial\delta\partial b} > 0$ and $\frac{\partial^2 u}{\partial h\partial b} > 0$. Because u is strictly increasing in δ, we obtain $\delta_j^{id}(h_d) > \delta_i^{id}(h_d)$. (iii) also hold from above.

The later part of (i) holds from the following; for $i,j \in \{l|h(l) \geq h'\}$ and $i > j$,

$$0 = u(a_k, b_j, h_d, \delta_j^{id}(h_d)) - u(a_k, b_j, x_j, \delta^k(x_j))$$

$$> u(a_k, b_i, h_d, \delta_j^{id}(h_d)) - u(a_k, b_i, x_j, \delta^k(x_j))$$

$$\geq u(a_k, b_i, h_d, \delta_j^{id}(h_d)) - u(a_k, b_i, x_i, \delta^k(x_i))$$

$$= u(a_k, b_i, h_d, \delta_j^{id}(h_d)) - u(a_k, b_i, h_d, \delta_i^{id}(h_d)).$$

The first inequality follows from $\delta_j^{id}(h_d) < \delta^k(x_j)$, $\frac{\partial^2 u}{\partial\delta\partial b} > 0$ and $\frac{\partial^2 u}{\partial h\partial b} > 0$. Then, we obtain $\delta_i^{id}(h_d) > \delta_j^{id}(h_d)$. (ii) also hold from above. □

We check an equilibrium whether or not to survive D1 criterion as the following two steps. First, we specify how the firm makes a belief when the firm observes an off-path working hour. Let $MBR_k(\lambda_k, h)$ be the set of the mixed best-response training strategy to the worker whose working hour is h given the belief λ_k. Abusing notation, $\lambda_k(T|h)$ also represents the posterior probability with which the firm believes the worker is in set T. Applying D1 criterion, the firm makes the belief that the deviator is NOT type-b_i worker on off-path working hour h_d if there exist $j \neq i$ such that

$$\{D_k(b_i, B, h_d) \cup D_k^0(b_i, B, h_d)\} \subset D_k\big(b_j, B, h_d\big)$$

where

$$D_k(b_i, B, h_d) = \bigcup_{\{\lambda_k|\lambda_k(B|h)=1\}} \{\delta \in MBR_k(\lambda_k, h_d) | u(a_k, b_i, x_i, \delta^k(x_i))$$

$$< u(a_k, b_i, h_d, \delta)\},$$

which represents the set of mixed best-response training strategy to working hour h_d and beliefs concentrated on B that make type-b_i worker strictly to prefer to the equilibrium strategy in the sense of expectation and

$$D_k(b_i, B, h_d) = \bigcup_{\{\lambda_k|\lambda_k(B|h)=1\}} \{\delta \in MBR_k(\lambda_k, h_d) | u(a_k, b_i, x_i, \delta^k(x_i))$$

$$= u(a_k, b_i, h_d, \delta)\},$$

which represent the set of expected mixed-strategy best-response training strategy that make type-b_i worker exactly indifferent. Note that $\{D_k(b_i, B, h_d) \cup D_k^0(b_i, B, h_d)\} = [\delta_i^{id}(h_d), \delta^*(a_k, b_n)]$, then $\{D_k(b_i, B, h_d) \cup D_k^0(b_i, B, h_d)\} \subset D_k(b_j, B, h_d)$ holds if $\delta_i^{id}(h_d) > \delta_j^{id}(h_d)$.

Second, we check the type of the worker who is considered as the deviator whether prefers the deviation to the equilibrium outcome. We define,

$$B^*(h_d) = \{i | \{D_k(b_i, B, h_d) \cup D_0^k(b_i, B, h_d)\} \not\subset D_k(b_j, B, h_d) \forall j \neq i\}$$

as the set of the type which the firm consider as the deviators and

$$\Delta^*(B^*(h_d), h_d) = \bigcup_{\{\lambda_k | \lambda_k(B^*(h_d)|h_d)=1\}} \{\delta \in BR_k(\lambda_k, h_d)\}$$

as the set of expected best-response training strategy to h_d for belief $\lambda_k(\cdot\,|h_d)$ such that $\lambda_k(B^*(h_d)|h_d) = 1$. The equilibrium does not survive D1 criterion if there exist a type of worker b_i such that, for some $h_d \notin H$,

$$\min_{\in\Delta^*(B^*(h_d),h_d)} u\,(a_k, b_i, h_d, \delta) > u(a_k, b_i, x_i, \delta^k(x_i)).$$

Proof of Proposition 1

proof. We show that the equilibrium characterized in Proposition 1 is the only equilibrium that survive D1 criterion under $s = R$.

We first show that, in the equilibrium that survive D1 criterion, the longest hour chosen in the equilibrium, denoted by h_m, is chosen by only type-b_n worker. Suppose not. If more than two types of worker choose h_m, then $\delta^k(h_m) < \delta^*(a_k, b_n)$. Consider a deviation strategy $h_m + \epsilon$, $\epsilon > 0$. From Lemma A2(iii), $B^*(h_m + \epsilon) = \{b_n\}$ then $\Delta^*(B^*(h_m + \epsilon), h_m + \epsilon) = \{\delta^*(a_k, b_n)\}$. There is small ϵ such that $u(a_k, b_n, h_m + \epsilon, \delta^*(a_k, b_n)) > u(a_k, b_n, h_m, \delta^k(h_m))$ then type- b_n worker deviates from the equilibrium. This is contradiction.

Next, we show that, in the equilibrium that survive D1 criterion, the second longest hour chosen in the equilibrium, denoted by h_{m-1}, is chosen by only type-b_{n-1} worker and that worker is indifferent between h_m and h_{m-1}. Suppose he/she is not indifferent, that is, $u(a_k, b_{n-1}, h_m, \delta^k(h_m)) < u(a_k, b_{n-1}, h_{m-1}, \delta^k(h_{m-1}))$. Then, there exist small $\epsilon' > 0$ such that $u(a_k, b_{n-1}, h_m - \epsilon', \delta^k(h_m)) < u(a_k, b_{n-1}, h_{m-1}, \delta^k(h_{m-1}))$

and $u(a_k, b_n, h_m - \epsilon', \delta^k(h_m)) > u(a_k, b_n, h_m, \delta^k(h_m))$. This implies $\delta_n^{id}(h_m - \epsilon') < \delta_{n-1}^{id}(h_m - \epsilon')$. Together with Lemma A2(i), we obtain $B^*(h_m - \epsilon') = \{b_n\}$ and $\Delta^*(B^*(h_m - \epsilon'), h_m - \epsilon') = \{\delta^*(a_k, b_n)\}$, then $u(a_k, b_n, h_m - \epsilon', \delta^*(a_k, b_n)) > u(a_k, b_n, h_m, \delta^k(h_m))$. Thus, type-$b_n$ worker deviates from the equilibrium, this is contradiction.

Suppose he/she is indifferent but h_{m-1} is chosen by more than two types. Since type-b_n chooses h_m, if more than two types of worker choose h_{m-1}, $\delta^k(h_{m-1}) < \delta^*(a_k, b_{n-1})$. Consider a deviation strategy $h_{m-1} + \epsilon'' < h_m$, $\epsilon'' > 0$. Comparing $\delta_n^{id}(h_{m-1} + \epsilon'')$ and $\delta_{n-1}^{id}(h_{m-1} + \epsilon'')$,

$$
\begin{aligned}
0 &= u(a_k, b_{n-1}, h_m, \delta^k(h_m)) - u(a_k, b_{n-1}, h_{m-1}, \delta^k(h_{m-1})) \\
&= u(a_k, b_{n-1}, h_m, \delta^k(h_m)) - u(a_k, b_{n-1}, h_{m-1} + \epsilon'', \delta_{n-1}^{id}(h_{m-1} + \epsilon'')) \\
&< u(a_k, b_n, h_m, \delta^k(h_m)) - u(a_k, b_n, h_{m-1} + \epsilon'', \delta_{n-1}^{id}(h_{m-1} + \epsilon'')) \\
&= u(a_k, b_n, h_{m-1} + \epsilon'', \delta_n^{id}(h_{m-1} + \epsilon'')) - u(a_k, b_n, h_{m-1} + \epsilon'', \delta_{n-1}^{id}(h_{m-1} + \epsilon''))
\end{aligned}
$$

where the first inequality follows from $\delta^k(h_m) > \delta_{n-1}^{id}(h_{m-1} + \epsilon'')$, $\frac{\partial^2 u}{\partial \delta \partial b} > 0$ and $\frac{\partial^2 u}{\partial h \partial b} > 0$, then we obtain $\delta_n^{id}(h_{m-1} + \epsilon'') > \delta_{n-1}^{id}(h_{m-1} + \epsilon'')$. Together with Lemma A2(i), $B^*(h_{m-1} + \epsilon'') = \{b_{n-1}\}$ and $\Delta^*(B^*(h_{m-1} + \epsilon''), h_{m-1} + \epsilon'') = \{\delta^*(a_k, b_{n-1})\}$. There is small ϵ'' such that $u(a_k, b_{n-1}, h_{m-1} + \epsilon'', \delta^*(a_k, b_{n-1})) > u(a_k, b_{n-1}, h_{m-1}, \delta^k(h_{m-1}))$ then type-b_{n-1} worker deviates from the equilibrium. This is contradiction. Applying the same manner downward, we can show that any equilibrium hour is chosen by only one type of worker and that type is indifferent between his/her equilibrium outcome and the one-rank higher worker'outcome. Finally, since the worker's commitment type is completely separated in the equilibrium, the lowest commitment worker does not need to signal his/her type to the firm and chooses his/her optimal working hour $h = 0$. Therefore, the equilibrium survive D1 criterion is only the equilibrium characterized in the proposition.

□

Since it is straightforward to prove Proposition 2 by applying almost the same manner, we omit the proof of Proposition 2. The only difference from the proof of Proposition 1 is that the worker's utility is expected utility with regard to his/her ability. The firm makes belief on off-path hours depending on the range of expected promotion probability $p\mu(a_H, \delta^H) + (1-p)\mu(a_H, \delta^H)$ which can improve the worker's utility, that is, the deviator the firm believes is the worker whose equilibrium utility can be improved with the lowest promotion probability.

References

Akerlof, George A. "The Economics of Caste and of the Rat Race and Other Woeful Tales." *Quarterly Journal of Economics*, 1976, *90*(4), pp. 599-617.

Arrow, Kenneth J. and Arrow, Kenneth J. "The Theory of Discrimination," *Collected Papers of Kenneth J. Arrow. Volume 6 Applied Economics.* Cambridge, Mass. and London: Harvard Univeristy Press, Belknap Press, 1985, 143-64.

Asai, Yukiko. "Parental Leave Reforms and the Employment of New Mothers: Quasi-Experimental Evidence from Japan." *Labour Economics*, 2015, *36*, pp. 72-83.

Bardsley, Peter and Sherstyuk, Katerina. "Rat Races and Glass Ceilings." *B.E. Journal of Theoretical Economics: Topics in Theoretical Economics*, 2006, *6*(1), pp. 1-35.

Barnett, William P.; Baron, James N. and Stuart, Toby E. "Avenues of Attainment: Occupational Demography and Organizational Careers in the California Civil Service." *American Journal of Sociology*, 2000, *106*(1), pp. 88-144.

Becker, Gary S. *The Economics of Discrimination*. Chicago, IL: University of Chicago Press., 1957.

Bertrand, Marianne; Goldin, Claudia and Katz, Lawrence F. "Dynamics of the Gender Gap for Young Professionals in the Financial and Corporate Sectors." *American Economic Journal: Applied Economics*, 2010, *2*(3), pp. 228-55.

Bjerk, David. "Glass Ceilings or Sticky Floors? Statistical Discrimination in a Dynamic Model of Hiring and Promotion." *Economic Journal*, 2008, *118*(530), pp. 961-82.

Blau, Francine D. and DeVaro, Jed. "New Evidence on Gender Differences in Promotion Rates: An Empirical Analysis of a Sample of New Hires." *Industrial Relations*, 2007, *46*(3), pp. 511-50.

Blau, Francine D. and Kahn, Lawrence M. "Gender Differences in Pay." *Journal of Economic Perspectives*, 2000, *14*(4), pp. 75-99.

Booth, Alison L.; Francesconi, Marco and Frank, Jeff. "A Sticky Floors Model of Promotion, Pay, and Gender." *European Economic Review*, 2003, *47*(2), pp. 295-322.

Booth, Alison and Nolen, Patrick. "Choosing to Compete: How Different Are Girls and Boys?" *Journal of Economic Behavior and Organization*, 2012, *81*(2), pp. 542-55.

Cabral, Robert; Ferber, Marianne A. and Green, Carole A. "Men and Women in Fiduciary Institutions: A Study of Sex Differences in Career Development." *Review of Economics and Statistics*, 1981, *63*(4), pp. 573-80.

Cannings, Kathy. "Managerial Promotion: The Effects of Socialization, Specialization, and Gender." *Industrial and Labor Relations Review*, 1988, *42*(1), pp. 77-88.

Cassidy, Hugh; DeVaro, Jed and Kauhanen, Antti. "Promotion Signaling, Gender, and Turnover: New Theory and Evidence." *Journal of Economic Behavior and Organization*, 2016, *126*, pp. 140-66.

Cha, Youngjoo and Weeden, Kim A. "Overwork and the Slow Convergence in the Gender Gap in Wages." *American Sociological Review (forthcoming)*, 2013.

Cobb-Clark, Deborah A. and Polachek, Solomon W. "Getting Ahead: The Determinants of and Payoffs to Internal Promotion for Young U.S. Men and

Women," *Worker Wellbeing in a Changing Labor Market.* Research in Labor Economics, vol. 20. Amsterdam; London and New York: Elsevier Science, JAI, 2001, 339-72.

Cortes, Patricia, and Jessica Pan. "", . "When Time Binds: Return to Working Long Hours and the Gender Wage Gap among the Highly Skilled," IZA Discussion Paper, No. 9846, 2016.

Datta Gupta, Nabanita; Poulsen, Anders and Villeval, Marie Claire. "Gender Matching and Competitiveness: Experimental Evidence." *Economic Inquiry,* 2013, *51*(1), pp. 816-35.

Davis, ShannonN and Greenstein, TheodoreN. "Cross-National Variations in the Division of Household Labor." *Journal of Marriage & Family,* 2004, *66,* pp. 1260-71.

Frederiksen, Anders and Kato, Takao. "Human Capital and Career Success: Evidence from Linked Employer-Employee Data." *Economic Journal (forthcoming),* 2016.

Fryer, Roland G. "Belief Flipping in a Dynamic Model of Statistical Discrimination." *Journal of Public Economics*, 2007, *91*(5-6), pp. 1151-66.

Fuwa, Makiko. "Macro-Level Gender Inequality and the Division of Household Labor in 22 Countries." *American Sociological Review*, 2004, *69*(6), pp. 751-67.

Garratt, Rodney J.; Weinberger, Catherine and Johnson, Nick. "The State Street Mile: Age and Gender Differences in Competition Aversion in the Field." *Economic Inquiry*, 2013, *51*(1), pp. 806-15.

Gicheva, Dora. "Working Long Hours and Early Career Outcomes in the High-End Labor Market." *Journal of Labor Economics,* 2013, *31*(4), pp. 785-824.

Gneezy, Uri; Niederle, Muriel and Rustichini, Aldo. "Performance in Competitive Environments: Gender Differences." *Quarterly Journal of Economics*, 2003, *118*(3), pp. 1049-74.

Gneezy, Uri and Rustichini, Aldo. "Gender and Competition at a Young Age." *American Economic Review*, 2004, *94*(2), pp. 377-81.

Goldin, Claudia. "A Grand Gender Convergence: Its Last Chapter." *American Economic Review*, 2014, *104*(4), pp. 1091-119.

Greenstein, Theodore N. "National Context, Family Satisfaction, and Fairness in the Division of Household Labor." *Journal of Marriage & Family*, 2009, *71*, pp. 1039-51.

Ishida, Junichiro. "Signaling and Strategically Delayed Promotion." *Labour Economics*, 2004, *11*(6), pp. 687-700.

Kato, Takao; Kawaguchi, Daiji and Owan, Hideo. "Dynamics of the Gender Gap in the Workplace: An Econometric Case Study of a Large Japanese Firm," *Paper presented at the NBER Japan Project meeting.* Tokyo, 2013.

Kuhn, Peter and Villeval, Marie Claire. "Are Women More Attracted to Co-Operation Than Men?" *Economic Journal*, 2015, *125*(582), pp. 115-40.

Landers, Renee M.; Rebitzer, James B. and Taylor, Lowell J. "Rat Race Redux: Adverse Selection in the Determination of Work Hours in Law Firms." *American Economic Review,* 1996, *86*(3), pp. 329-48.

Lazear, Edward P. and Rosen, Sherwin. "Male-Female Wage Differentials in Job Ladders." *Journal of Labor Economics*, 1990, *8*(1), pp. S106-23.

McCue, Kristin. "Promotions and Wage Growth." *Journal of Labor Economics,* 1996,

14(2), pp. 175-209.
Miyazaki, Hajime. "The Rat Race and Internal Labor Markets." *Bell Journal of Economics*, 1977, *8*(2), pp. 394-418.
Moriguchi, Chiaki. "Implicit Contracts, the Great Depression, and Institutional Change: A Comparative Analysis of U.S. And Japanese Employment Relations, 1920-1940." *Journal of Economic History*, 2003, *63*(3), pp. 625-65.
Niederle, Muriel and Vesterlund, Lise. "Do Women Shy Away from Competition? Do Men Compete Too Much?" *Quarterly Journal of Economics*, 2007, *122*(3), pp. 1067-101.
Owan, Hideo. "Promotion, Turnover, Earnings, and Firm-Sponsored Training." *Journal of Labor Economics*, 2004, *22*(4), pp. 955-78.
Paulin, Elizabeth A. and Mellor, Jennifer M. "Gender, Race, and Promotions within a Private-Sector Firm." *Industrial Relations*, 1996, *35*(2), pp. 276-95.
Pekkarinen, Tuomas and Vartiainen, Juhana. "Gender Differences in Promotion on a Job Ladder: Evidence from Finnish Metalworkers." *Industrial and Labor Relations Review*, 2006, *59*(2), pp. 285-301.
Phelps, Edmund S. "The Statistical Theory of Racism and Sexism." *American Economic Review*, 1972, *62*(4), pp. 659-61.
Prendergast, Canice. "Career Development and Specific Human Capital Collection." *Journal of the Japanese and International Economies*, 1992, *6*(3), pp. 207-27.
Qian, Yue and Sayer, Liana C. "Division of Labor, Gender Ideology, and Marital Satisfaction in East Asia." *Journal of Marriage & Family*, 2016, *78*, pp. 383-400.
Sampson, Anthony A. and Simmons, Robert. "Adverse Selection When Jobs Are Hard to Do." *Scottish Journal of Political Economy*, 2000, *47*(3), pp. 325-36.
Smith, Nina; Smith, Valdemar and Verner, Mette. "Why Are So Few Females Promoted into Ceo and Vice President Positions? Danish Empirical Evidence, 1997-2007." *Industrial and Labor Relations Review*, 2013, *66*(2), pp. 380-408.
Steinberg, Chad and Nakane, Masato. "Can Women Save Japan?," IMF Working Paper WP/12/248, 2012.
Stiglitz, Joseph E. "The Theory of 'Screening,' Education, and the Distribution of Income." *American Economic Review*, 1975, *65*(3), pp. 283-300.
Sutter, Matthias and Glatzle-Rutzler, Daniela. "Gender Differences in the Willingness to Compete Emerge Early in Life and Persist." *Management Science*, 2015, *61*(10), pp. 2339-54.
Waldman, Michael. "Job Assignments, Signalling, and Efficiency." *RAND Journal of Economics*, 1984, *15*(2), pp. 255-67.

Table 1. Summary of the results

	w^{OT} sufficiently low	w^{OT} sufficiently high
$\mu_{a\delta} > 0$	Revelation is optimal. $h_i^L < h_i^N < h_i^H$ (ex. US, UK)	Non-revelation is optimal. $h_i^L < h_i^N < h_i^H$
$\mu_{a\delta} < 0$ & $f(a_H) - f(a_L)$ is small	Non-revelation is optimal. $h_i^H < h_i^N < h_i^L$ (ex. Japan before 1990)	Revelation is optimal. $h_i^H < h_i^N < h_i^L$

Table 2 Proportion of single men, married men, single women, and married women with and without young children over 2005-2013

Fiscal year	Single men	Married men	Single women	Married women without young children	Married women with young children*
2005	16.9%	71.3%	5.8%	5.1%	1.0%
2006	18.2%	69.9%	5.7%	4.7%	1.5%
2007	19.4%	68.4%	5.8%	4.7%	1.8%
2008	21.4%	66.4%	5.8%	4.5%	2.0%
2009	21.7%	65.8%	5.8%	4.4%	2.3%
2010	21.0%	66.1%	5.8%	4.4%	2.7%
2011	20.5%	67.0%	5.3%	4.3%	2.9%
2012	20.3%	66.8%	5.5%	4.3%	3.0%
2013	20.6%	66.2%	5.7%	4.3%	3.3%

Source: Personnel data provided by MfgJapan

Note: Includes only standard employees. * shows the number of female workers who have taken parental leave since FY1999. This group also includes a very small number of women who got divorced after having at least one child.

Table 3 Wage, Age, and Hours Worked of single men, married men, single women, and married women with and without young children over 2005-2013

Average Annual Wage

Fiscal year	Single men	Married men	Single women	Married women without young children	Married women with young children*
2005	6,329,606	8,313,333	5,427,290	5,500,951	4,497,976
2006	6,412,564	8,438,916	5,602,173	5,762,287	4,696,728
2007	6,379,339	8,484,262	5,812,861	6,057,774	4,906,125
2008	6,205,874	8,382,596	5,745,866	6,087,957	4,895,578
2009	5,617,625	7,805,416	5,340,310	5,722,504	4,700,844
2010	5,777,784	8,060,655	5,583,306	6,050,426	4,878,955
2011	6,373,215	8,715,721	6,130,749	6,727,495	5,627,406
2012	6,081,554	8,144,137	5,770,398	6,321,167	5,244,958
2013	6,182,902	8,202,954	5,776,894	6,410,150	5,388,207

Average Age

Fiscal year	Single men	Married men	Single women	Married women without young children	Married women with young children*
2005	36.6	46.1	37.5	37.6	34.0
2006	36.1	45.9	38.1	38.6	34.7
2007	35.7	45.6	38.5	39.4	35.4
2008	34.9	45.0	38.7	40.1	36.0
2009	34.9	44.8	39.0	40.9	36.8
2010	35.5	44.6	40.0	41.5	37.4
2011	36.3	44.6	41.5	42.2	38.3
2012	36.6	44.7	42.0	43.0	39.1
2013	37.2	44.9	42.3	43.9	39.7
2014	37.9	45.4	42.9	45.2	40.5

Average Hours Worked per Year

Fiscal year	Single men	Married men	Single women	Married women without young children	Married women with young children*
2005	2,073	2,064	1,931	1,820	1,671
2006	2,088	2,076	1,947	1,838	1,663
2007	2,093	2,089	1,949	1,854	1,586
2008	2,085	2,076	1,922	1,835	1,605
2009	2,049	2,034	1,857	1,804	1,576
2010	2,032	2,037	1,912	1,854	1,610
2011	2,075	2,062	1,909	1,865	1,611
2012	2,066	2,050	1,921	1,872	1,618
2013	2,066	2,042	1,913	1,869	1,616

Source: Personnel data provided by MfgJapan.

Note: The figures in the table only include those who work for twelve months in each fiscal year. Hence, when an employee takes a parental leave in year t, he/she will be excluded from year t.

Table 4 Proportion of Workers in Different Job Grades in 2005, 2009, and 2013

Job Grades	Single men	Married men	Single women	Married women without young children	Married women with young children*	Total
As of 2005						
J1,J2	46.75%	24.01%	75.12%	67.34%	83.65%	33.59%
SA,J3	16.05%	12.74%	14.47%	10.28%	7.69%	13.22%
SB,J4	12.33%	12.77%	2.76%	5.14%	1.92%	11.62%
G6,JH	6.72%	7.17%	2.60%	6.97%	4.81%	6.80%
G5	10.16%	17.73%	3.90%	7.52%	1.92%	14.98%
G4	4.89%	14.57%	0.33%	1.83%	0.00%	11.32%
G3	1.17%	5.66%	0.16%	0.55%	0.00%	4.27%
G1,G2	0.17%	1.29%	0.00%	0.00%	0.00%	0.95%
Total	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Supervisory managers	416	3527	43	92	7	4085
(G6/JH or G5)	10.18%	86.34%	1.05%	2.25%	0.17%	100.00%
Middle and top managers	112	1635	3	13	0	1763
(G4 or higher)	6.35%	92.74%	0.17%	0.74%	0.00%	100.00%
As of 2009						
J1,J2	50.11%	22.50%	60.26%	55.29%	71.84%	33.27%
SA,J3	20.20%	14.32%	24.72%	19.65%	11.43%	16.37%
SB,J4	12.10%	13.50%	5.98%	4.10%	6.12%	12.18%
G6,JH	4.79%	7.86%	3.23%	6.91%	8.16%	6.89%
G5	6.75%	14.85%	4.36%	8.64%	2.45%	11.92%
G4	4.31%	17.72%	1.29%	3.89%	0.00%	12.84%
G3	0.83%	6.57%	0.00%	0.86%	0.00%	4.54%
G1,G2	0.09%	1.22%	0.00%	0.00%	0.00%	0.82%
Total	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Supervisory managers	385	3361	55	94	26	3921
(G6/JH or G5)	9.82%	85.72%	1.40%	2.40%	0.66%	100.00%
Middle and top managers	120	1778	8	22	0	1928
(G4 or higher)	6.22%	92.22%	0.41%	1.14%	0.00%	100.00%
As of 2013						
J1,J2	40.57%	24.75%	52.69%	49.37%	66.11%	32.00%
SA,J3	19.92%	10.58%	22.31%	14.18%	7.97%	13.24%
SB,J4	18.91%	14.10%	11.15%	11.39%	7.97%	14.61%
G6,JH	7.20%	8.97%	5.19%	6.33%	12.96%	8.41%
G5	6.46%	12.64%	5.38%	10.89%	4.65%	10.62%
G4	5.03%	17.28%	1.92%	5.57%	0.33%	12.83%
G3	0.64%	5.23%	0.00%	0.76%	0.00%	3.63%
G1,G2	0.42%	4.92%	0.19%	0.76%	0.00%	3.39%
Total	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Supervisory managers	373	2980	66	96	54	3569
(G6/JH or G5)	10.45%	83.50%	1.85%	2.69%	1.51%	100.00%
Middle and top managers	115	1667	11	28	1	1822
(G4 or higher)	6.31%	91.49%	0.60%	1.54%	0.05%	100.00%

Source: Personnel data provided by MfgJapan
Note: Percentages do not add up to 100% because miscellaneous job grades are excluded.

Table 5: Logit estimates of the relationship between the probability of promotion (conditional on the current job level) and working hours for men and women

	All Employees		College Graduates	
VARIABLES	(1)	(2)	(3)	(4)
age_{it}	0.186***	0.190***	0.222***	0.228***
	(0.026)	(0.026)	(0.050)	(0.050)
age_{it}^2	-0.00268***	-0.00272***	-0.00357***	-0.00364***
	(0.00032)	(0.00032)	(0.00066)	(0.00066)
$tenure_{it}$	0.0507***	0.0503***	0.0222	0.0222
	(0.0093)	(0.0093)	(0.0153)	(0.0154)
$tenure_{it}^2$	-0.00025	-0.00023	0.00200***	0.00201***
	(0.00025)	(0.00025)	(0.00054)	(0.00054)
$marriage_{it}$	0.368***	0.369***	0.368***	0.368***
	(0.041)	(0.041)	(0.048)	(0.048)
$female_i$	-21.79***	-1.330***	-14.84***	-0.937***
	(2.257)	(0.168)	(3.082)	(0.218)
$marriage_{it}$*$female_i$	0.00963	0.0139	-0.128	-0.111
	(0.115)	(0.114)	(0.143)	(0.141)
$\ln(hours_{it-1})$	1.199***		0.775***	
	(0.144)		(0.165)	
$female_i$ * $\ln(hours_{it-1})$	5.384***		3.620***	
	(0.426)		(0.580)	
$hours18_{it-1}$		0.170**		0.038
		(0.072)		(0.088)
$hours20_{it-1}$		0.243***		0.134
		(0.072)		(0.088)
$hours22_{it-1}$		0.483***		0.239***
		(0.076)		(0.092)
$hours24_{it-1}$		0.417***		0.216**
		(0.082)		(0.096)
$female_i$* $hours18_{it-1}$		0.818***		0.490**
		(0.166)		(0.226)
femalei* $hours20_{it-1}$		1.383***		0.969***
		(0.171)		(0.223)
femalei* $hours22_{it-1}$		1.193***		0.774***
		(0.216)		(0.269)
femalei* $hours24_{it-1}$		1.878***		1.530***
		(0.253)		(0.300)
Pseudo R^2	0.089	0.0896	0.0747	0.0751
Observations	60,998	60,998	28,546	28,546

Source: Personnel data provided by MfgJapan

Notes: In addition to the variables in the table, we control for education, job level, location (the headquarters or other offices/plants), and year. College graduates are employees with a bachelor's degree or higher, and include those from *Kouto Senmon Gakkou* (5-year colleges of technology that admits junior high school graduates).

*** p<0.01, ** p<0.05, * p<0.1

Figure 1. Timing of the game

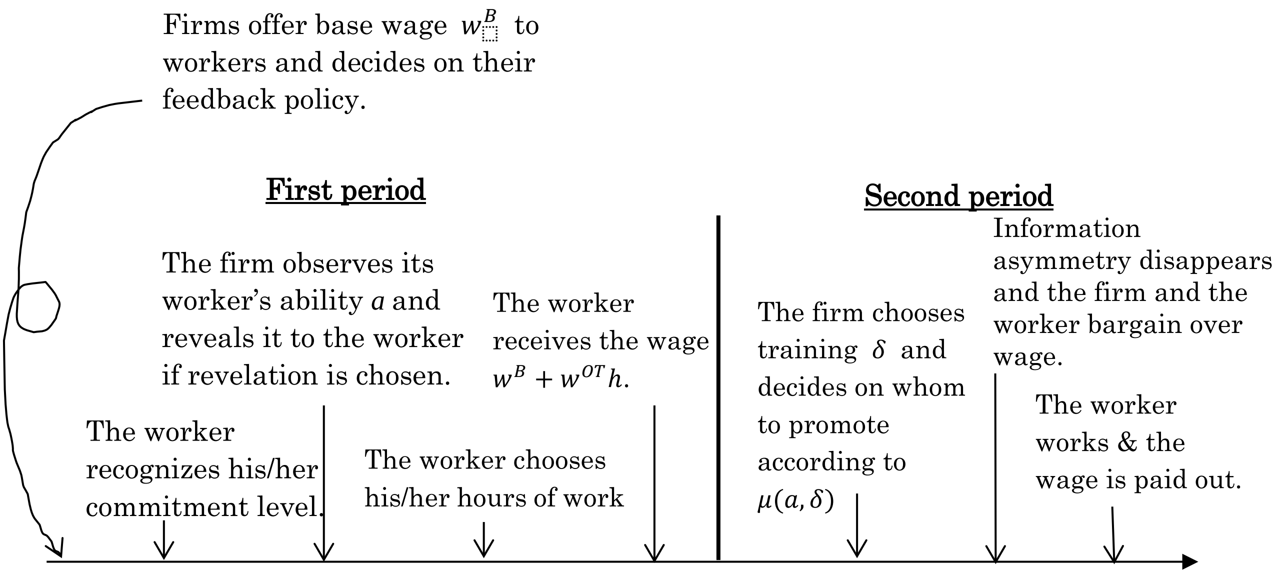


Figure 2. Do our equilibrium survive D1 criterion?

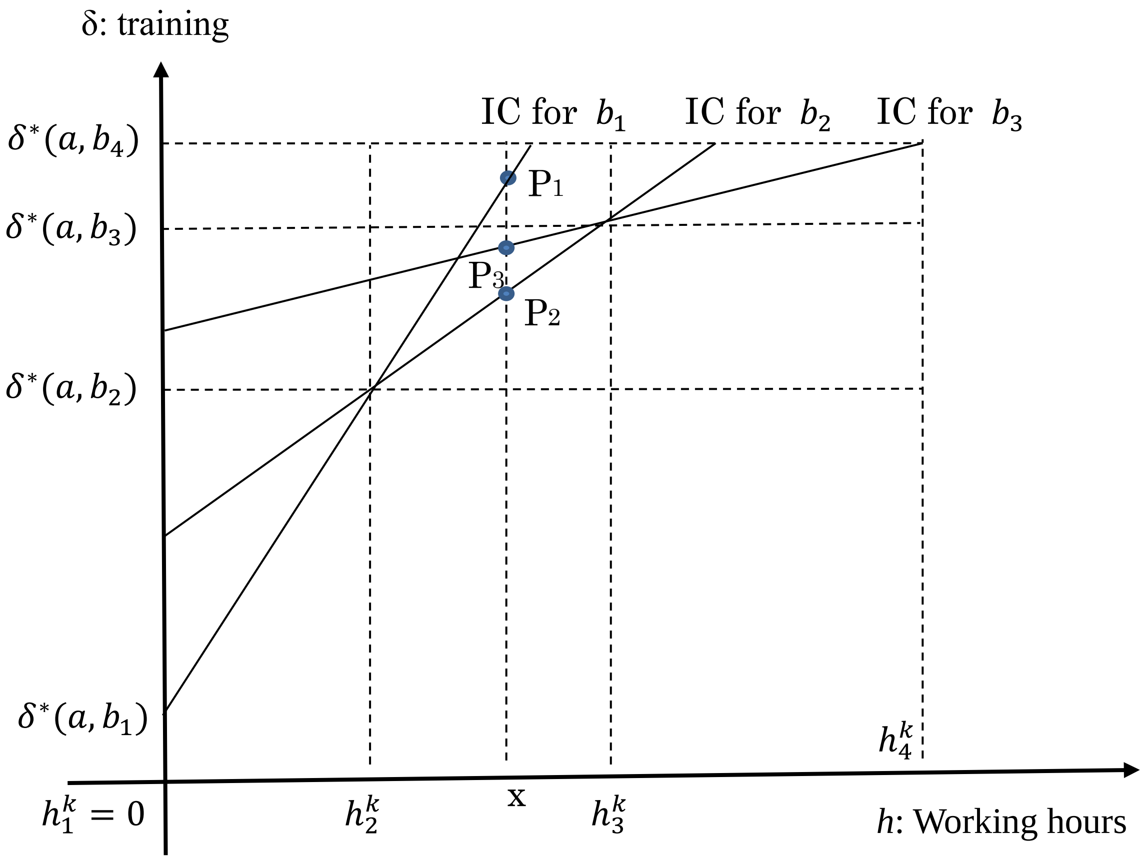


Figure 3 Expected Profit as a Function of b when $\mu_{a\delta} > 0$

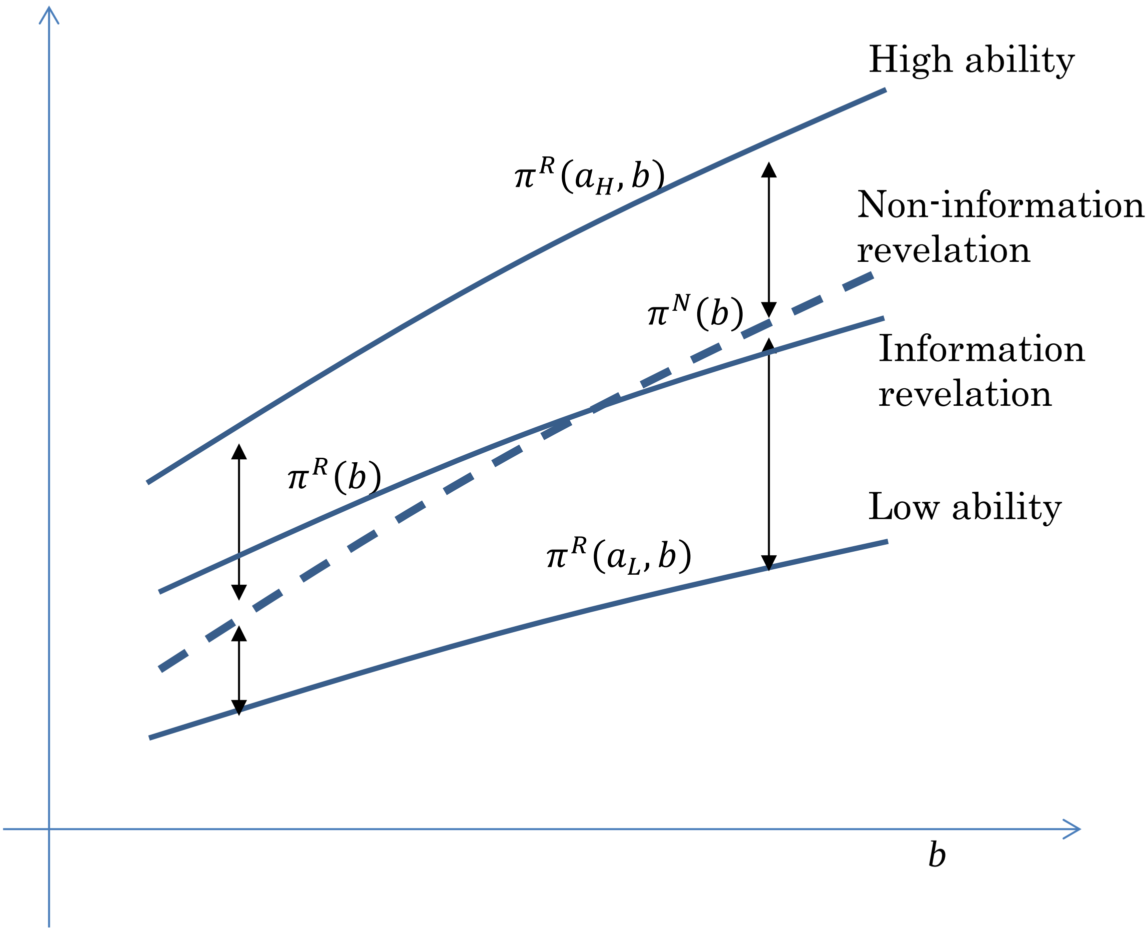


Figure 4. Gender Difference in Promotion Pattern

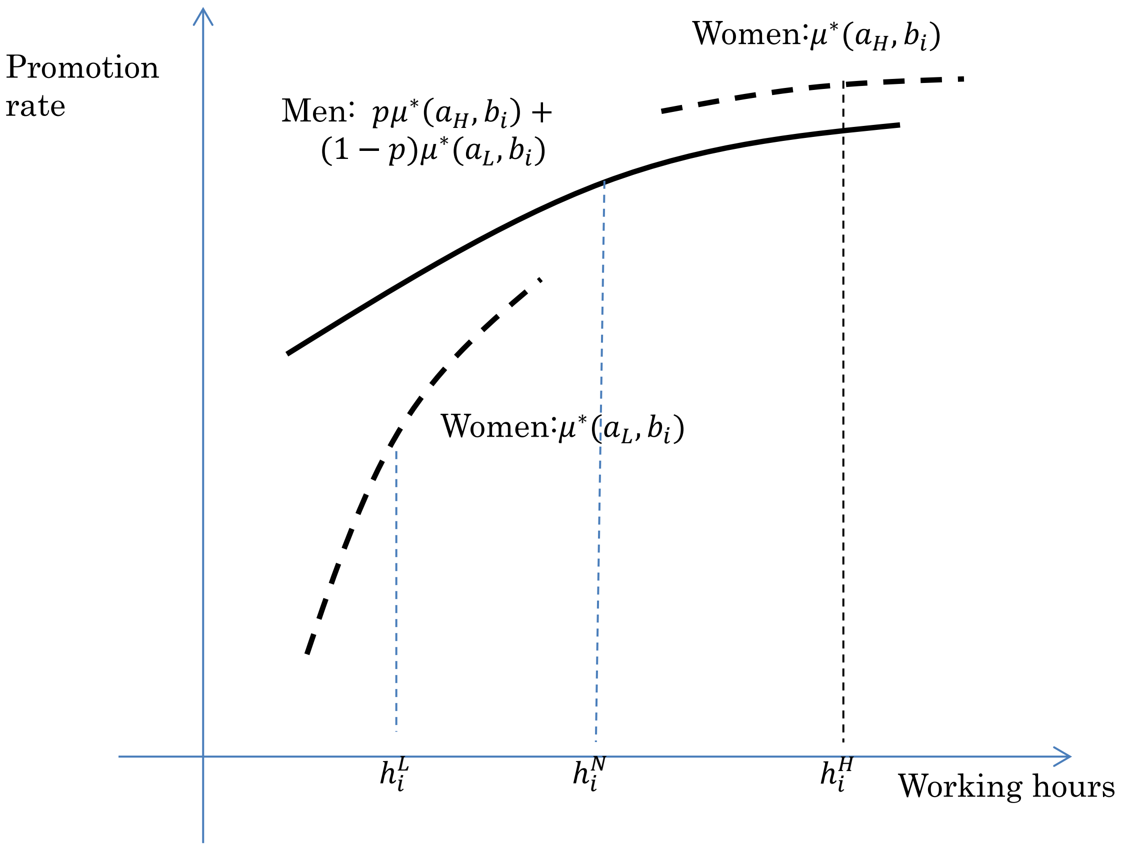


Figure 5 MfgJapan's Promotion Ladder

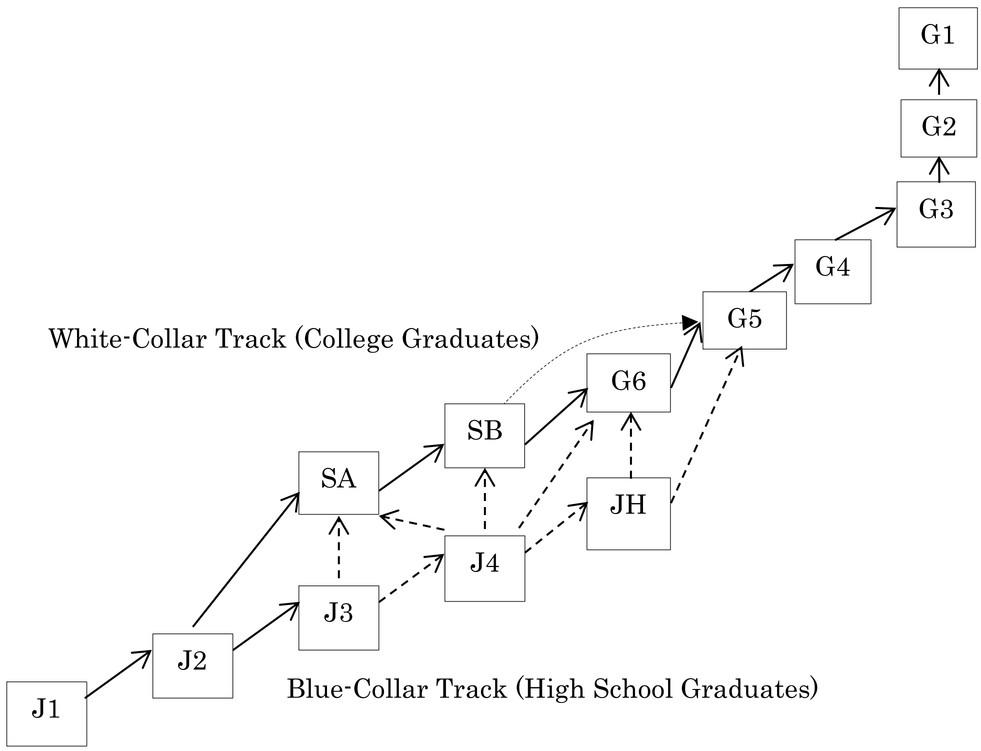
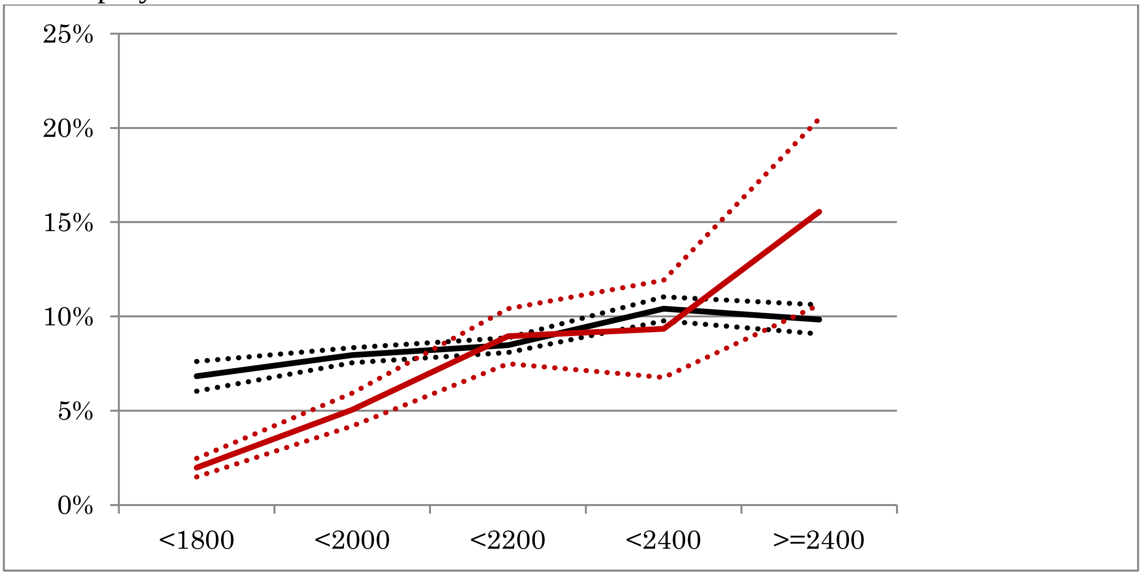
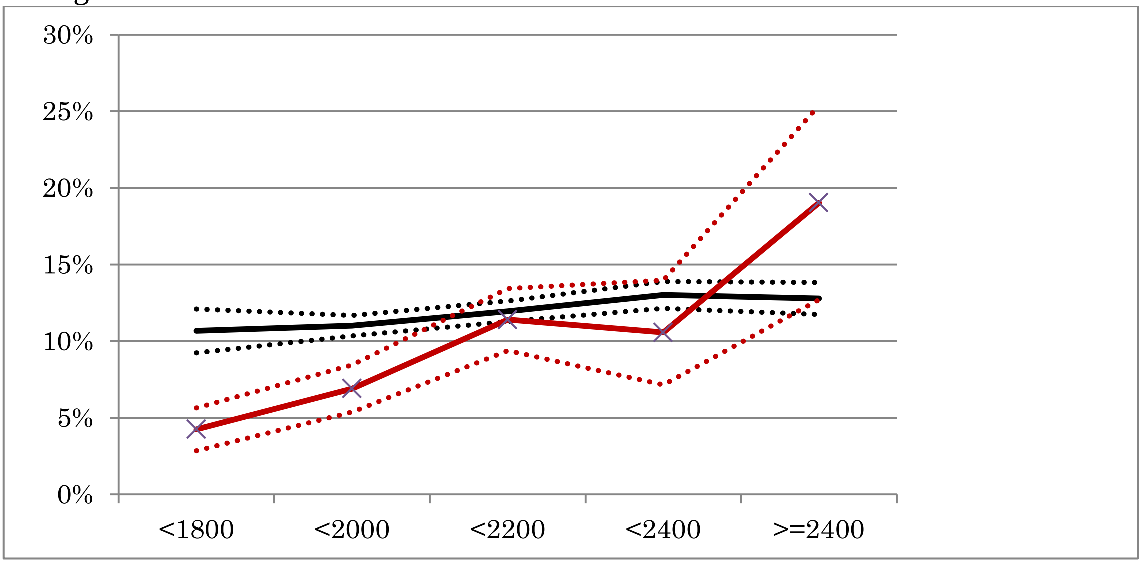


Figure 6 Average Predicted Probability of Promotion for Employees with Different Annual Working Hours: Men and Women

All Employees



College Graduates



Notes: The dotted lines indicate 95% confidence intervals. College graduates are employees with a bachelor's degree or higher, and include those from *Kouto Senmon Gakkou* (5-year colleges of technology that admits junior high school graduates).

The proportion of workers in each hour range by gender is shown in the following tables:

All employees

	<1800	<2000	<2200	<2400	>=2400	Total
Men	8.7%	33.5%	35.5%	14.4%	7.9%	100.0%
Women	43.3%	35.6%	14.9%	4.6%	1.5%	100.0%

College Graduates

	<1800	<2000	<2200	<2400	>=2400	Total
Men	7.8%	32.5%	30.0%	17.6%	12.1%	100.0%
Women	25.2%	37.0%	24.2%	9.8%	3.7%	100.0%