Dynamically Sabotage-Proof Tournaments

May 10, 2006

Junichiro Ishida

Associate Professor, Osaka School of International Public Policy (OSIPP)

Sabotage; Tournament; Fast track; Late selection; Supermodularity; Submodularity

This paper examines a two-period tournament where agents may possibly engage in destructive sabotage activities. Under plausible circumstances, sabotage proves to be an effective tool for low-ability agents, especially when they are faced with high-ability agents. The possibility of sabotage then gives rise to a dynamic concern, similar to the Ratchet effect, because an agent runs a risk of becoming the target of sabotage by signaling his high ability in early stages. In this dynamic setting, we first establish an impossibility result where the mere possibility of sabotage makes it impossible to implement the first-best effort due to this dynamic concern. Given this result, we then offer two distinct incentive schemes, fast track and late selection, to circumvent this problem. The fast-track scheme is likely to prevail when the production process values diversity in inputs (submodular technologies) while the late-selection scheme is to prevail when it values homogeneity (supermodular technologies). The present model thus offers a mechanism through which both fast track and late selection arise in a unified framework, which can explain the difference in managerial practices between the US and Japan.

Correspondence to: Junichiro Ishida, OSIPP, Osaka University, 1-31 Machikaneyama Toyonaka, Osaka 560-0043, Japan. E-mail: jishida@osipp.osaka-u.ac.jp

1 Introduction

Relative compensation schemes in general, or tournaments in particular, are a prevalent form of evaluation adopted in many situations. Despite several virtues,[1] however, it is also well known that relative compensation schemes have some drawbacks that are inherent to their relative nature. One particular form of these drawbacks, as suggested by Lazear (1989), is the fact that there may arise an incentive to engage in destructive sabotage activities, instead of more productive counterparts, because reducing the opponent's output can be a close substitute for improving one's own output under relative compensation schemes. Since it is an essential task for any managers to build and sustain a cooperative, or at least non-hostile, work environment, it is of critical interest to see in what ways sabotage activities can be mitigated.

In this paper, we address issues that stem from the possibility that agents may resort to destructive sabotage activities in a dynamic context. Consider a two-period tournament between two agents who may differ in innate ability where each agent has three alternatives to choose from: exerting productive effort, sabotage effort or no effort at all. To appreciate the dynamic inefficiency that arises in this setup, we start the analysis by showing that sabotage tends to be an effective tool for low-ability agents, especially when they are faced with high-ability opponents. That is, when the perceived difference in innate ability is sufficiently large, it becomes the preferred option for the less able agent to sabotage the opponent to fill this gap. This fact gives rise to a serious dynamic implication when agents share some information about each other over time. Suppose that, at the end of each period, each agent can observe the opponent's productivity, which reflects both ability and effort. In such a situation, a high-ability agent essentially runs a risk of becoming the target of sabotage by signaling his high ability in early stages. There then arises an incentive for high-ability agents to control their effort in order to conceal their private information and appear less able.

---

[1] There are several reasons why the principal uses relative compensation schemes, or tournaments in particular, as a way to motivate agents. One reason is that with a relative compensation scheme, the principal can fix the total amount of prizes paid to the agents, which is especially important when the agents' performances are not verifiable to a third party. Also, relative compensation schemes become more effective when the agents are subject to common stochastic shocks. Finally, it is often argued that relative performance evaluation is cheaper and easy to obtain in many regards than absolute performance evaluation.

This dynamic concern yields a critical implication for the design of optimal incentive schemes. Suppose that the cost of productive effort is so small that it is always optimal to induce productive effort, whenever it is feasible to do so. Within this setup, we first show that any tournament that can implement the first-best effort (always inducing productive effort) is not sabotage-proof under fairly plausible conditions: in other words, there exists no contract that can implement the first-best effort in both periods when sabotage is a viable option. The logic behind this result is fairly simple. When the first-best effort (everyone exerting productive effort) is implemented in the first period, any difference in the productivity must be attributed to the difference in innate ability. There then inevitably arises a situation where the perceived difference in innate ability is so large that it is optimal for the less able agent to resort to sabotage activities in the second period. This result thus indicates that although the costs arising from sabotage activities are well recognized in the static setting,[2] the possibility of sabotage invites more serious problems in the dynamic setting than previously recognized.

Given this impossibility result, we then explore ways in which to mitigate sabotage in search of the second best, with a particular focus on how much weight to place on the ranking in each period. The main issue here is how to device a tournament that can mitigate the incentive for low-ability agents to exert sabotage effort in the second period. In general, there are two distinct ways to achieve this goal. More precisely, we show that the initial tournament that can implement the first-best effort in the absence of sabotage can be made sabotage-proof by shifting the weight in either direction, i.e., either towards the ranking in the first period or in the second. While these two schemes can equally prevent sabotage from actually taking place, each comes at a cost with different implications.

First, consider a tournament which places more weight on the ranking in the first period: that is, high-powered incentives are provided in the first period, followed by low-powered incentives in the second. The key aspect of this scheme is the pay compression to make sabotage less effective in the second period, as Lazear (1989) points out. Since less is at stake in the second period, this reduces the incentive for low-ability agents to exert sabotage effort in that period. There is a cost associated with

---

[2] An obvious cost of sabotage is that it substitutes for more productive effort. This also leads to another inefficiency, as pointed out by Chen (2003), that the principal may fail to select the most deserving agent in the presence of sabotage activities.

this scheme, however, because with low-powered incentives, it also fails to induce desirable productive effort from low-ability agents. An illuminating property of this type of incentive scheme is that it always induces productive effort from high-ability agents at a potential cost that low-ability agents exert no effort in the second period. For expositional purposes, we refer to this as the fast-track scheme since this scheme rewards the first-period winner more heavily.

Besides the fast-track scheme, there is another, totally opposite, way to mitigate sabotage in this setting. Now consider a tournament which places more weight on the ranking in the second period: that is, low-powered incentives are provided in the first period, complemented with high-powered incentives in the second. Since more is at stake in the second period, high-ability agents have an incentive to conceal their private information in order not to get too much ahead of others because they run a risk of becoming the target of sabotage in later stages by doing so. An illuminating property of this type of incentive scheme is that it always induces productive effort from low-ability agents at a potential cost that high-ability agents exert no effort in the first period. For expositional purposes, we refer to this as the late-selection scheme since this scheme rewards the second-period winner more heavily.

The present model thus provides a potential explanation for both fast track and late selection in a unified framework, from a previously unexplored perspective. At the same time, the explanation is also consistent with the difference in managerial practices between the US and Japan. It is argued that the US firms often adopt the fast-track scheme by selecting promising candidates early on (also known as the early selection of 'stars'). This draws clear contrast to many Japanese firms which tend to adopt the late-selection scheme where they do not differentiate workers for a substantial period of time, roughly 10-15 years. We argue that the single most important factor in choosing between the two schemes is the nature of the production process. The fast-track scheme makes sure that high-ability agents always exert optimal effort. Since effort is positively related to ability, i.e., more able agents exert more effort, effective inputs from each agent are highly diversified under this scheme. This implies that the fast-track scheme becomes the optimal choice when the production process

values diversity, rather than homogeneity, in inputs. It is argued that this feature, often referred to as submodularity, is more common in industries such as software, fashion and entertainment which value new ideas and creativity, and are more prevalent in the US. In the late-selection approach, on the other hand, effort is negatively related to ability, which makes effective inputs from each agent fairly homogenized. The late-selection scheme is then more likely to be efficient when the production process values homogeneity in inputs. It is argued that this feature, often referred to as supermodularity, is more common in industries such as automobile and consumer electronics which require careful and precise implementation of tasks, and are more prevalent in Japan.[3]

While the argument thus far lays out an explanation for why fast track or late selection is more often adopted given technological factors, we can also make an argument from the other way around. For instance, many sociologists and historians alike point out that harmony (called 'wa' in Japanese) is one of the most important and salient concepts that prevail in the Japanese society: as the old saying goes, "a tall tree catches much wind," which is believed to be one of the principals guiding and regulating people's behavior in Japan. The cultural tendency that emphasizes harmony is believed to stem originally from Confucianism, which has had enormous influences on the Japanese culture in its formative period. Many Japanese firms seem to inherit this cultural tendency where maintaining harmony or homogeneity within organizations once again appears to be the objective of utmost concern. In this sense, compensation schemes which tend to minimize diversity (or preserve homogeneity) are simply a natural fit for the Japanese society from the beginning. The main logic here can then be regarded as an explanation for why industries with supermodular technologies have flourished and been so successful in Japan from a cultural perspective.

The paper is related to several strands of literature. First, there are many works that examine the optimal timing of promotion or compensation with different approaches and focuses.[4] Just to name a few, one of the most influential approaches to explain delayed compensations is the incentive

---

[3]Grossman and Maggi (2000) argue that this difference in the nature of technologies arises from the distribution of talent within each country. It is argued that submodular technologies are more prevalent in the US because the distribution of talent is more diverse in the US.

[4]For more extensive surveys, see Gibbons and Waldman (1999a) and Prendergast (1999).

approach by Lazear (1979), which posits that compensations should be delayed to maintain career incentives. The signaling approach, most notably by Waldman (1984), states that promotions in early stages tend to be inefficiently few because a promotion signals the worker's ability, which in turn raises the retention wage.[5] As for the difference in managerial practices between the US and Japan, Prendergast (1992) argues that Japanese firms are able to delay promotion since workers are limited in their mobility in the labor market. As an explanation for fast track, on the other hand, Meyer (1982) constructs a two-period model and shows that it is optimal to bias the second-period tournament in favor of the first-period winner.[6] Gibbons and Waldman (1999b, 2006) show that workers who receive large wage increases early in their stay at one level of the hierarchy are promoted quickly to the next level because workers who receive large wage increases are likely to be those with high ability.

The paper is also related to works that focus on negative aspects of relative compensation schemes. A seminal paper on sabotage in a tournament is Lazear (1989). The paper is most closely related to Chen (2003) who shows in a static tournament that able members are likely to be subject to sabotage attacks, illustrating an inefficiency that the most able member might not have the best chance of being promoted. The focus of the present paper differs from Chen (2003) as it is placed on dynamic interactions between the agents and the ways in which to mitigate sabotage activities.[7] The paper is also related to a literature that deals with collusion under relative performance evaluation.[8] Along this line, Ishiguro (2004) shows that the principal can prevent collusion by offering asymmetric contracts, even though agents are symmetric with respect to productive abilities. This paper is similar in spirit as it seeks for sabotage-proof contracts, instead of collusion-proof, where agents have no incentive to exert sabotage effort.

The paper is organized as follows. The basic model is outlined in the next section and analyzed in section 3. Section 4 characterizes the equilibrium without sabotage activities as a benchmark. Section

---

[5]Owan (2004) applies this approach to explain the US-Japanese difference by complementing it with firm-specific human capital. Bernhardt (1995) extends this signaling approach to account for fast-track promotions as well as other stylized facts of internal labor markets.

[6]In this paper, we consider a situation where the principal is unable to bias the tournament in order to focus out attention.

[7]Chen (2003) discusses several schemes to mitigate sabotage activities, although not in a formal analysis.

[8]Mookherjee (1984) is one of the first to point out that tournaments are vulnerable to collusion.

5 then extends the analysis to incorporate potential sabotage activities and obtain the main results. Section 6 discusses key properties of the model and section 7 makes some concluding remarks.

2 The model

2.1 Environment

Consider a two-period model in which a principal (female) hires two agents (male), denoted by $i \in \{1, 2\}$, to produce output. Each agent differs in his innate ability $a_i \in \{L, H\}$. We say that agent i possesses high (low) ability if $a_i = H$ ($a_i = L$). The prior distribution of the ability type is given by

$$\text{prob}\{a_i = H\} = \theta \in (0, 1).$$

The ability type is the agent's private information while the prior distribution is common knowledge. Let $\mu_{i,t} \equiv \text{prob}\{a_{-i} = h \mid \Omega_{i,t}\}$, $i \neq -i$, denote agent i's belief about the other agent's ability type at the beginning of period t, based on his information set $\Omega_{i,t}$. By construction, $\mu_{i,1} = \theta$, $i = 1, 2$.

2.2 Production

In each period t, each agent must decide his effort levels $(e_{i,t}, d_{i,t}) \in \{0, 1\}^2$ where $e_{i,t}$ indicates the level of productive effort while $d_{i,t}$ indicates the level of (destructive) sabotage effort. We assume that each effort level is a binary choice and moreover that each agent can exert at most one type of effort, i.e., $e_{i,t} + d_{i,t} \leq 1$. The cost of productive effort is c while that of sabotage effort is $(1 + \lambda)c$.

The individual productivity of each agent, denoted by $y_{i,t}$, depends on his ability and effort as well as the other agent's sabotage effort, if it is positive. We assume that the role of sabotage effort is to negate the other agent's productive effort. More precisely, the productivity of each agent is given by $y_{i,t} = f(a_i, e_{i,t} - d_{-i,t})$ where

$$f(H, 1) = h, \ f(H, 0) = f(L, 1) = m, \ f(H, -1) = f(L, 0) = l, \ f(L, -1) = 0.$$

We assume that the marginal value of productive effort is larger for high-ability agents.

Assumption 1 $h - m > m - l > l > 0$, *i.e., productive effort is complementary to ability.*

6

2.3 Information

The crux of the model is its information structure, i.e., who can observe what. In this model, we consider a situation where the agents have access to more precise information regarding their productivity. More precisely, each agent i can observe the other agent's productivity $y_{-i,t}$ in each period (although not the ability type nor the effort level). The principal, on the other hand, can only observe the relative ranking of each agent $r_{i,t} \in \{0,1\}$ where $r_{i,t} = 1$ if agent i outperforms the other agent in period t. The relative ranking is an imperfect signal of the productivity and is given by

$$\text{prob}\{r_{i,t} = 1 \mid y_{i,t}, y_{-i,t}\} = G(y_{i,t} - y_{-i,t}).$$

We assume that the distribution function G has the following properties: (i) it is strictly increasing in its argument; (ii) $G(0) = 0.5$; (iii) $\lim_{z \to -\infty} G(z) = 0$ and $\lim_{z \to \infty} G(z) = 1$; and (iv) the corresponding density g is single-peaked and symmetric around zero, i.e., $\text{argmax}_z g(z) = 0$ and $g(z) = g(-z)$ for all z.[9] These properties imply that $G(z) + G(-z) = 1$ and $G(x) - G(0) > G(x + z) - G(z)$ for any $x > 0$ and $z > 0$, which we repeatedly use in the subsequent analysis.

If the relative cost of sabotage is too small, it is always optimal to exert sabotage effort; if it is too large, it is never to optimal to do so. To focus our attention to more interesting cases, therefore, we make the following assumption regarding the relative size of these costs.

Assumption 2 $\dfrac{G(h - m) - G(m - l)}{G(m - l) - G(0)} > \lambda > 0.$

2.4 Contracts

Under the current information structure, the relative ranking is the only available measure of performances for the principal.[10] Since the principal must rely on the relative ranking to motivate the agents, sabotage evidently becomes a serious issue for all parties involved in the transactions. Besides

---

[9]All those properties hold when each agent's productivity is subject to a shock drawn from the same distribution.

[10]When agents are asymmetric with respect to the productivity, it is often optimal for the principal to set up a tournament with a handicap. See Lazear and Rosen (1984). Meyer (1992) also shows that the possibility of biased tournaments have important dynamic implications. In order to exclude this effect and focus our attention, we assume that it is not feasible to bias the tournament one way or the other because the principal can only observe the relative ranking.

this, we place two more restrictions on the class of contracts we consider in this paper. First, the agents face a liquidity constraint so that the wages paid to the agents must be nonnegative. Second, the relative ranking is not verifiable to a third party so that the principal cannot offer asymmetric contracts contingent on each agent's identity. Under those restrictions, the loser always receives the minimum, which is zero, and a feasible contract can generically be written as (w_1, w_2) where w_t is the wage paid to the winner in period t. Alternatively, if this is a promotion tournament, $(w_1 r_{i,1} + w_2 r_{i,2})/(w_1 + w_2)$ can be regarded as the promotion probability with the total prize $w_1 + w_2$, and $w_t/(w_1 + w_2)$ as the weight given to each period t.

2.5 Payoffs

Since each agent has three alternatives, i.e., productive effort, sabotage effort, and no effort, there are potentially many equilibria. To focus on more interesting cases, throughout the analysis, we restrict our attention to sabotage-proof equilibria in pure strategies where the agents never exert destructive effort by assuming that the principal's payoff is prohibitively negative when sabotage activities ever take place. This implies that each agent's equilibrium strategy profile is completely summarized by $(e_t^H(\mu_{i,t}), e_t^L(\mu_{i,t}))$, or (e_t^H, e_t^L) for short, where $e_t^a(\mu) \in \{0, 1\}$ is the level of productive effort in period t, conditional on the ability type a and the belief μ. We also define (p_t^H, p_t^L) where $p_t^a \in [0, 1]$ is the *ex ante* probability with which an agent with ability a exerts productive effort in period t.[11]

For simplicity, we assume that the principal's (gross) payoff is specified exclusively as a function of (p_t^H, p_t^L). Let $\pi(p_t^H, p_t^L)$ denote the principal's payoff. The payoff is not necessarily the sum of $y_{i,t}$ as it also captures the degree of externalities (or complementarities) between the agents. We assume that productive effort is sufficiently valuable for the principal so that it is always optimal (and socially efficient) to implement $(p_t^H = 1, p_1^L = 1)$, $t = 1, 2$, at any finite cost whenever it is feasible.

---

[11]The *ex ante* probability refers to the one at the beginning of period 1, that is, before any additional information (namely, $y_{i,1}$) is observed.

2.6 Timing

Notice that the relative ranking observed by the principal reveals no relevant information about the actual realization of the ability type, and there is no incentive for the principal to restructure the initial contract at the interim stage. For this reason, without loss of generality, we consider the following timing of the model.

1. The principal offers a contract (w_1, w_2), which may or may not be accepted by the agents. If the contract is rejected, both parities receive zero and the game ends at this point.

2. In each period $t = 1, 2$, each agent determines the effort levels $(e_{i,t}, d_{i,t})$ and the outcome $(y_{i,t}, r_{i,t})$ is realized.

3. At the end of period 2, the principal pays the wage $w_1 r_{i,1} + w_2 r_{i,t}$ as specified by the contract.

Both the principal and the agents simply maximize the sum of the net payoffs with no discounting. More precisely, the principal maximizes $\pi(p_t^H, p_t^L) - w_t$ over the two periods, subject to various incentive compatibility constraints. Similarly, each agent maximizes the expected wage minus the cost of effort over the two periods.

In this model, fast track is regarded as a compensation scheme which emphasizes early performances and thus places more weight on the ranking in period 1 ($w_1 > w_2$). In the late selection approach, on the other hand, late performances become more important so that a compensation scheme places roughly equal weight ($w_1 \approx w_2$) or even more weight on the ranking in period 2 ($w_2 > w_1$). We show that both situations can be optimal, depending largely on the nature of the production process.

3 Incentives for productive and sabotage effort

3.1 The second-period problem

To solve the model backward, we begin with the second-period problem. The agents' incentives are relatively straightforward in period 2 as it is virtually a static problem with no future concerns.

Let $(e_{-i,t}^H(\mu_{-i,t}), e_{-i,t}^L(\mu_{-i,t}))$, or $(e_{-i,t}^H, e_{-i,t}^L)$ for short, denote the opponent's strategy profile in period t, conditional on the ability type. Since each agent have three alternatives (productive

effort, sabotage effort and no effort), three constraints are sufficient to characterize the preferences among them. First, given the opponent's strategy $(e_{-i,2}^H, e_{-i,2}^L)$ and the belief $\mu_{i,2}$, productive effort is preferred to no effort if

$$[\mu_{i,2}G(f(a_i,1) - f(H, e_{-i,2}^H)) + (1 - \mu_{i,2})G(f(a_i,1) - f(L, e_{-i,2}^L))]w_2 - c$$

$$\geq [\mu_{i,2}G(f(a_i,0) - f(H, e_{-i,2}^H)) + (1 - \mu_{i,2})G(f(a_i,0) - f(L, e_{-i,2}^L))]w_2. \qquad (1)$$

Second, productive effort is preferred to sabotage effort if

$$[\mu_{i,2}G(f(a_i,1) - f(H, e_{-i,2}^H)) + (1 - \mu_{i,2})G(f(a_i,1) - f(L, e_{-i,2}^L))]w_2 - c$$

$$\geq [\mu_{i,2}G(f(a_i,0) - f(H, e_{-i,2}^H - 1)) + (1 - \mu_{i,2})G(f(a_i,0) - f(L, e_{-i,2}^L - 1))]w_2 - (1+\lambda)c. \,(2)$$

Finally, no effort is preferred to sabotage effort if

$$[\mu_{i,2}G(f(a_i,0) - f(H, e_{-i,2}^H)) + (1 - \mu_{i,2})G(f(a_i,0) - f(L, e_{-i,2}^L))]w_2$$

$$\geq [\mu_{i,2}G(f(a_i,0) - f(H, e_{-i,2}^H - 1)) + (1 - \mu_{i,2})G(f(a_i,0) - f(L, e_{-i,2}^L - 1))]w_2 - (1+\lambda)c. \,(3)$$

To induce productive effort, a contract w_2 must satisfy both (1) and (2). Moreover, any sabotage-proof contract must satisfy at least one of either (2) or (3).

We first consider incentives for sabotage effort. Since productive effort is complementary to ability and the role of sabotage effort is to negate productive effort, it is easy to imagine that the incentive to engage in sabotage activities is stronger for low-ability agents. The following statement establishes that when productive effort is less costly than sabotage effort, high-ability agents never exert sabotage effort.

Proposition 1 *Productive effort always dominates sabotage effort for high-ability agents.*

PROOF: See Appendix.

Even though productive effort is less costly, the same statement cannot be made for low-ability agents because the return to productive effort is lower for them. For $a_i = L$, (2) can be written as

$$\lambda c \;\geq\; \{\mu_{i,2}[G(l - f(H, e_{-i,2}^H - 1)) - G(m - f(H, e_{-i,2}^H))]$$

10

$$+ (1 - \mu_{i,2})[G(l - f(L, e^L_{-i,2} - 1)) - G(m - f(L, e^L_{-i,2}))]\}w_2. \tag{4}$$

Given this, two remarks about incentives for sabotage are in order. First, as Lazear (1989) points out, one way to inhibit sabotage activities is to compress the pay scale. Second, notice that the left-hand side is positive when $e^H_{-i,2} = 1$ and $\mu_{i,2}$ is sufficiently large. Since this is from the viewpoint of low-ability agents, this implies that sabotage activities are more of a problem when the perceived difference in ability is sufficiently large. This aspect of the model yields a critical dynamic implication.

In order to induce productive effort, the principal must satisfy both (1) and (4) simultaneously. Examining these constraints closely, however, one can show that there is a range of the belief for which this is not feasible. This yields the following statement.

Proposition 2 *There exists some threshold $\bar{\mu} \in (0, 1)$ such that, given some belief $\mu_{i,2}$ and $e^H_{-i,2} = 1$, low-ability agents never exert productive effort for $\mu_{i,2} > \bar{\mu}$.*

PROOF: See Appendix.

3.2 The first-period problem

The problem becomes more complicated in period 1 since the situation is now dynamic and what each agent does in period 1 may influence the opponent's belief, and the expected payoff, in period 2. Define $u^{a_i}(e_{i,1}, d_{i,1})$, $a_i = L, H$, denote agent i's expected payoff in period 2 as a function of his own effort levels in period 1.

The incentive compatibility constraints need minor modifications in this dynamic setting. As above, three constraints characterize the dominance relationship among the three alternatives. First, given the opponent's strategy $(e^H_{-i,1}, e^L_{-i,1})$, productive effort is preferred to no effort if

$$[\theta G(f(a_i, 1) - f(H, e^H_{-i,1})) + (1 - \theta)G(f(a_i, 1) - f(L, e^L_{-i,1}))]w_1 - c + u^{a_i}(1, 0)$$

$$\geq [\theta G(f(a_i, 0) - f(H, e^H_{-i,1})) + (1 - \theta)G(f(a_i, 0) - f(L, e^L_{-i,1}))]w_1 + u^{a_i}(0, 0). \tag{5}$$

11

Second, productive effort is preferred to sabotage effort if

$$[\theta G(f(a_i,1) - f(H, e^H_{-i,1})) + (1-\theta)G(f(a_i,1) - f(L, e^L_{-i,1}))]w_1 - c + u^{a_i}(1,0)$$

$$\geq [\theta G(f(a_i,0) - f(H, e^H_{-i,1} - 1)) + (1-\theta)G(f(a_i,0) - f(L, e^L_{-i,1} - 1))]w_1 - (1+\lambda)c + u^{a_i}(0,1). \quad (6)$$

Finally, no effort is preferred to sabotage effort if

$$[\theta G(f(a_i,0) - f(H, e^H_{-i,1})) + (1-\theta)G(f(a_i,0) - f(L, e^L_{-i,1}))]w_1 + u^{a_i}(0,0)$$

$$\geq [\theta G(f(a_i,0) - f(H, e^H_{-i,1} - 1)) + (1-\theta)G(f(a_i,0) - f(L, e^L_{-i,1} - 1))]w_1 - (1+\lambda)c + u^{a_i}(0,1). \quad (7)$$

The incentive compatibility constraints are now dynamic in that they include not only w_1 but also w_2.

4 Equilibrium with no sabotage activities: a benchmark

We first consider as a benchmark a case where sabotage effort is not a viable option in order to single out the impact of potential sabotage activities. We in particular seek for an equilibrium where the agents always exert productive effort in both periods, i.e., $\{(p_1^H = 1, p_1^L = 1), (p_2^H = 1, p_2^L = 1)\}$. Assuming that effort is sufficiently valuable, the optimal contract in this benchmark case is the one that implements this effort profile, if such a contract exists.

With no sabotage activities, the agents' behavior is completely characterized by a single constraint in each period, (1) and (5). To induce productive effort with probability one, both of these constraints must be satisfied for any possible contingency. The following statement presents a contract that can achieve this.

Proposition 3 *Suppose that sabotage effort is not a viable option. The optimal contract is then given by*

$$w_1 = w_1^* \equiv \frac{c}{\theta[G(h-l) - G(h-m)] + (1-\theta)[G(m-l) - G(0)]}, \quad w_2 = w_2^* \equiv \frac{c}{G(h-l) - G(h-m)},$$

which implements $\{(p_1^H = 1, p_1^L = 1), (p_2^H = 1, p_2^L = 1)\}$.

PROOF: See Appendix.

Notice that $w_2 > w_1$, i.e., the principal must provide stronger incentives to induce productive effort in period 2. This is because the agents possess more information about each other after a period of production. Since the additional information sometimes discourages low-ability agents to exert productive effort, stronger incentives are needed to overcome this.[12]

5 Equilibrium with sabotage activities

We now introduce sabotage effort into the model and see how the mere possibility of sabotage activities alters the optimal structure of incentives. What is especially important in this respect is proposition 2, which implies the following impossibility result.

Proposition 4 *Suppose that sabotage effort is a viable option. Then, there exists no contract that can implement $\{(p_1^H = 1, p_1^L = 1), (p_2^H = 1, p_2^L = 1)\}$.*

PROOF : If $(p_1^H = 1, p_1^L = 1)$, the agents are able to identify the opponent's ability type with probability one and, hence, $\mu_{i,2} \in \{0, 1\}$. The situation described in proposition 2 is then bound to arise with some positive probability.

Q.E.D.

The proposition indicates that the principal is unable to implement the first-best effort and must instead settle for the second-best. The main problem here is the possibility of sabotage activities in period 2. We now consider two distinct incentive schemes that can suppress sabotage activities.

One possible scheme is to provide low-powered incentives in period 2 and directly suppress sabotage activities, as suggested by Lazear (1989). The well-known cost associated with this scheme is that with low-powered incentives, the principal also fails to induce desirable productive effort. An illuminating property of this type of incentive scheme is that it always induces productive effort from high-ability

---

[12]To be more precise, the marginal value of productive effort (the marginal increase in the winning probability) for low-ability agents decreases as the belief $\mu_{i,2}$ increases, i.e., the opponent becomes more able.

agents. We refer to this incentive scheme as the fast-track scheme since more weight is placed on early performances.

The fast-track scheme $(p_1^H = p_2^H = 1)$: Sabotage activities are circumvented by offering high-powered incentives in period 1, followed by low-powered incentives in period 2. The compensation scheme places more weight on early achievements so that there are not sufficient returns for low-ability agents to exert sabotage effort in period 2.

In this dynamic setting, there is another way to circumvent sabotage activities in period 2. Suppose that the principal provides low-powered incentives in period 1, complemented with high-powered incentives in period 2. Since more is at stake in period 2, high-ability agents would rather choose to exert no effort in order to conceal their ability type. If $\bar{\mu} > \theta$, this can prevent sabotage activities because the productivity reveals no relevant information and hence $\mu_{i,2} = \theta$ with probability one when $(e_1^H = 0, e_1^L = 1)$. An illuminating property of this type of incentive scheme is that it always induces productive effort from low-ability agents, in contract to the fast-track scheme. We refer to this incentive scheme as the late-selection scheme since more weight is placed on late performances.

The late-selection scheme $(p_1^L = p_2^L = 1)$ Sabotage activities are circumvented by offering low-powered incentives in period 1, complemented with high-powered incentives in period 2. The compensation scheme places less weight on early achievements so that more is at stake in period 2.

5.1 The optimal fast-track contract

In this equilibrium, the principal implements $(p_1^H = 1, p_1^L = 1)$ so that the ability type is perfectly identifiable in period 2. To prevent any sabotage activities, incentives must be weak enough to satisfy (4). This obviously comes at a cost: satisfying (4) necessarily implies the violation of (1) for low-ability agents with positive probability.

Proposition 5 *The optimal fast-track contract is given by*

$$w_1 = w_1^{\text{FT}} \equiv \frac{c}{\theta[G(h-l) - G(h-m)] + (1-\theta)[G(m-l) - G(0)]}, \quad w_2 = w_2^{\text{FT}} \equiv \frac{c}{G(h-l) - G(m-l)}.$$

14

which implements $\{(p_1^H = 1, p_1^L = 1), (p_2^H = 1, p_2^L = 0)\}$.

PROOF: See Appendix.

Compared to the case without sabotage activities, the optimal first-track contract provides weaker incentives in period 2, i.e., $w_2^* > w_2^{\mathrm{FT}}$. This is because the principal must compress the pay structure, or minimize the difference between the winner and the loser, in order to mitigate potential sabotage activities, at an expense that it fails to induce productive effort from low-ability agents. The incentive scheme places more weight on early performances, i.e., the ranking in period 1 $r_{i,1}$, as a way to protect high-ability workers from sabotage attacks from colleagues by designating the winner early on: in other words, fast track is used as a safeguard for sabotage activities.

5.2 The optimal late-selection contract

Things are somewhat more complicated in the late-selection scheme. Under this scheme, the principal implements $(p_1^H = 0, p_1^L = 1)$ so that the ability type is not identifiable in period 2. Although it is not possible to induce productive effort only from low-ability agents in period 2, this may be possible in period 1 because high-ability agents face different dynamic incentives.

Proposition 6 *If θ is sufficiently small, the optimal late-selection contract exists and is given by*

$$w_1 = w_1^{\mathrm{LS}} \equiv \frac{c}{G(m-l) - G(0)}, \quad w_2 = w_2^{\mathrm{LS}} \equiv \frac{c}{\theta[G(h-l) - G(h-m)] + (1-\theta)[G(m-l) - G(0)]},$$

which implements $\{(p_1^H = 0, p_1^L = 1), (p_2^H = 1, p_2^L = 1)\}$.

PROOF: See Appendix.

In this late-selection approach, the principal starts with weaker incentives ($w_1^* > w_1^{\mathrm{LS}}$) and raises the stake later on ($w_2^{\mathrm{LS}} > w_2^*$). Since more is at stake in period 2, high-ability agents have an incentive to conceal their true type by exerting no effort even though it certainly lowers the probability of winning in period 1. As a result, agents are not initially differentiated for some time, maintaining harmony within organizations.

6   Discussion: diversity vs homogeneity

The previous section illustrates two distinct schemes that can suppress sabotage activities. The question is then which scheme the principal should adopt. The principal adopts the fast-track scheme over the late-selection scheme if

$$\pi(1,0) - w_1^{\mathrm{FT}} - w_2^{\mathrm{FT}} \geq \pi(0,1) - w_1^{\mathrm{LS}} - w_2^{\mathrm{LS}}, \tag{8}$$

which can be written as

$$\Delta\pi \equiv \pi(1,0) - \pi(0,1) \geq \frac{c}{G(h-l) - G(m-l)} - \frac{c}{G(m-l) - G(0)}. \tag{9}$$

Among other factors, a critical determinant turns out to be the nature of the production function or, more specifically, $\Delta\pi$.

Our interpretation of $\Delta\pi$ is as follows. The first term $\pi(1,0)$ indicates the payoff when only high-ability agents exert productive effort. Since effort is positively related to ability in this case, the distribution of effective inputs is more diversified. This term is then likely to be large when the productivity depends more crucially on the best idea or the luckiest draw, disproportionately reflecting the input of a few highly talented individuals. It is argued that this feature, often referred to as submodularity, is more common in industries such as software, fashion and entertainment which value new ideas and creativity, and are more prevalent in the US.

The situation is totally opposite, on the other hand, when only low-ability agents exert productive effort. Since effort is negatively related to ability in this case, the distribution of effective inputs is more compressed and homogenized. The second term $\pi(0,1)$ is then likely to be large when there are strong complementarities between the tasks and each task needs to be done equally well. It is argued that this feature, often referred to as supermodularity, is more common in industries such as automobile and consumer electronics which require careful and precise implementation of tasks, and are more prevalent in Japan.

In light of this interpretation, the fast-track (late-selection) scheme is likely to prevail if the underlying production technology exhibits submodularity (supermodularity) and hence values diversity

16

(homogeneity) in inputs. The present framework then provides an explanation for the difference in managerial practices between the US and Japan by focusing on the nature of production technologies. It is often pointed out that Japanese firms typically do not differentiate workers for a substantial period, roughly 10-15 years, which draws clear contrast to the early selection of 'stars' in the US. Since Japanese firms are willing to evaluate workers over a long span of time, their workers are in no hurry to show off their talent and potential, and hence it is wise for those with high ability to keep pace with others in early stages of their career. This pattern of behavior reflects a belief widely pervasive in the Japanese society that "a tall tree gets much wind," which is believed to be one of the most important principles guiding and regulating people's behavior in Japan. We argue that this late-selection scheme contributes in important ways to corporate culture that emphasizes harmony and homogeneity in workplaces – one of the defining features of Japanese firms.

7 Conclusion

This paper constructs a two-period model of a tournament to illustrate dynamic inefficiencies that arise from that possibility that agents may engage in sabotage activities. We first show that when sabotage is a viable option, it is impossible to implement the first-best effort under fairly plausible circumstances. Given this result, we then show that the initial contract that can implement the first-best effort without sabotage can be made sabotage-proof by shifting the weight in either direction, i.e., either towards the first-period winner or towards the second-period winner. The critical determinant of the optimal scheme turns out to be the nature of the production process. The fast-track scheme which rewards the first-period winner more heavily is optimal when the production process values diversity in inputs; the late-selection scheme which rewards the second-period winner more heavily is optimal when it values homogeneity. We argue that this result provides a mechanism through which both fast track and late selection arise in a unified framework, which can explain the difference in managerial practices between the US and Japan.

Appendix: the proofs

PROOF OF PROPOSITION 1: For $a_i = H$, (2) can be written as

$$
\begin{aligned}
\lambda c \;\geq\; & \{\mu_{i,2}[G(m - f(H, e^H_{-i,2} - 1)) - G(h - f(H, e^H_{-i,2}))] \\
& + (1 - \mu_{i,2})[G(m - f(L, e^L_{-i,2} - 1)) - G(h - f(L, e^L_{-i,2}))]\}w_2.
\end{aligned}
\tag{A.1}
$$

Productive effort is then preferred to sabotage effort if

$$
\mu[G(m - f(H, e^H - 1)) - G(h - f(H, e^H))] + (1 - \mu)[G(m - f(L, e^L - 1)) - G(h - f(L, e^L))] \leq 0, \tag{A.2}
$$

for any μ and (e^H, e^L). Alternatively, we need to show that

$$
G(m - f(H, e^H - 1)) - G(h - f(H, e^H)) \leq 0, \tag{A.3}
$$

$$
G(m - f(L, e^L - 1)) - G(h - f(L, e^L)) \leq 0, \tag{A.4}
$$

both of which hold for any given (e^H, e^L).

Q.E.D.

PROOF OF PROPOSITION 2: Solving (1), we can show that productive effort is preferred to no effort if $w_2 \geq \overline{w}(\mu_{i,2}; e^H_{-i,2}, e^L_{-i,2})$ where

$$
\overline{w}(\mu; e^H, e^L) \equiv \frac{c}{\mu[G(m - f(H, e^H)) - G(l - f(H, e^H))] + (1 - \mu)[G(m - f(L, e^L)) - G(l - f(L, e^L))]}.
$$

Similarly, it follows from (4) that productive effort is preferred to sabotage effort if $\tilde{w}(\mu_{i,2}; e^H_{-i,2}, e^L_{-i,2}) \geq w_2$ where

$$
\tilde{w}(\mu; e^H, e^L) \equiv \frac{\lambda c}{\mu[G(l - f(H, e^H - 1)) - G(m - f(H, e^H))] + (1 - \mu)[G(l - f(L, e^L - 1)) - G(m - f(L, e^L))]}.
$$

Given that $e^H_{-i,2} = 1$, there exists no contract that can induce productive effort from low-ability agents if $\overline{w}(\mu; 1, e^L) > \tilde{w}(\mu; 1, e^L)$. The existence of the threshold $\bar{\mu}$ is guaranteed if $\overline{w}(1; 1, e^L) > \tilde{w}(1; 1, e^L)$, which can be written as

$$
\frac{G(h - m) - G(m - l)}{G(h - l) - G(h - m)} > \lambda. \tag{A.5}
$$

18

Note that this condition holds because

$$\frac{G(h-m)-G(m-l)}{G(h-l)-G(h-m)} > \frac{G(h-m)+G(m-l)}{G(m-l)-G(0)} > \lambda. \tag{A.6}$$

The first inequality holds since $G(m-l) - G(0) > G(h-l) - G(h-m)$, which hold when the distribution is symmetric and single-peaked around zero. The second inequality holds by assumption 1.

<div align="right">Q.E.D.</div>

PROOF OF PROPOSITION 3: It follows from (1) that productive effort is preferred to no effort for high-ability agents if $w_2 \geq \underline{w}(\mu_{i,2}; e^H_{-i,2}, e^L_{-i,2})$ where

$$\underline{w}(\mu; e^H, e^L) \equiv \frac{c}{\mu[G(h-f(H,e^H)) - G(m-f(H,e^H))] + (1-\mu)[G(h-f(L,e^L)) - G(m-f(L,e^L))]}.$$

Similarly, as can be seen in the proof of proposition 2, productive effort is preferred to no effort for low-ability agents if $w_2 \geq \overline{w}(\mu_{i,2}; e^H_{-i,2}, e^L_{-i,2})$. Note that $\overline{w} > \underline{w}$ for any $\mu_{i,2}$ and $(e^H_{-i,2}, e^L_{-i,2})$ so that $w_2 \geq \underline{w}$ holds if $w_2 \geq \overline{w}$. To induce productive effort with probability one, therefore, we need to focus on $\overline{w}$.

If $(p^H_1 = 1, p^L_1 = 1)$, then $\mu_{i,2} \in \{0, 1\}$. To implement $(p^H_2 = 1, p^L_2 = 1)$, therefore, a second-period contract must satisfy both $w_2 \geq \overline{w}(1; 1, 1)$ and $w_2 \geq \overline{w}(0; 1, 1)$. Note that $\overline{w}(1; 1, 1) > \overline{w}(0; 1, 1)$. This implies that the optimal contract in period 2 is given by

$$w_2 = \overline{w}(1; 1; 1) = \frac{c}{G(h-l) - G(h-m)}. \tag{A.7}$$

We now shift attention to the first-period problem. To implement $(p^H_1 = 1, p^L_1 = 1)$, the following condition must be satisfied:

$$w_1 \geq \frac{c + u^{a_i}(0,0) - u^{a_i}(1,0)}{\theta[G(f(a_i,1) - h) - G(f(a_i,0) - h)] + (1-\theta)[G(f(a_i,1) - m) - G(f(a_i,0) - m)]}. \tag{A.8}$$

The first-period constraint thus hinges critically on $u^{a_i}(0,0) - u^{a_i}(1,0)$.

We first consider incentives for low-ability agents. For those agents, $\mu_{-i,2} = 0$ no matter what they do in period 1. This implies that $u^{a_i}(0,0) = u^{a_i}(1,0)$, and the problem is totally identical to that in period 2. Low-ability agents exert productive effort if $w_2 \geq \overline{w}(\theta; 1, 1)$.

<div align="center">19</div>

Now consider incentives for high-ability agents. The equilibrium expected payoff for high-ability agents under the candidate contract can be computed as

$$u^H(1,0) = [\theta G(0) + (1-\theta)G(h-m)]\overline{w}(1;1;1) - c. \tag{A.9}$$

Suppose that a high-ability agent unilaterally deviates and chooses $e_{i,1} = 0$. This signals to the opponent that the agent is of the low-ability type. This does not influence the opponent's behavior, though, because the candidate contract always induce productive effort for any given belief. Again, the problem is reduced to that in period 2. High-ability agents exert productive effort if $w_1 \geq \underline{w}(\theta; 1, 1)$.

Since $\overline{w}(\theta; 1, 1) > \underline{w}(\theta; 1, 1)$, the optimal contract in period 1 is given by

$$w_1 = \overline{w}(\theta; 1, 1) = \frac{c}{\theta[G(h-l) - G(h-m)] + (1-\theta)[G(m-l) - G(0)]}. \tag{A.10}$$

Q.E.D.

PROOF OF PROPOSITION 5: Note that the ability type is not identifiable only when $(p_1^H = 0, p_1^L = 1)$; for any other profile, the ability type is perfectly identifiable. Provided that effort is sufficiently valuable, it is optimal for the principal to implement $(p_1^H = 1, p_1^L = 1)$, if not $(p_1^H = 0, p_1^L = 1)$.

Given this, we first obtain the following result.

Lemma 1 If $(p_1^H = 1, p_1^L = 1)$, then $p_0^L = 0$.

PROOF: If $(p_1^H = 1, p_1^L = 1)$, then $\mu_{i,2}$ takes either zero or one. We thus need to show that $e_2^L(0) = e_2^L(1) = 0$. Suppose first that $e_2^L(1) = 1$. This implies that $w_2 \geq \overline{w}(1; e^H, e^L)$, can be written as

$$w_2 \geq \frac{c}{G(m - f(H, e^H)) - G(l - f(H, e^H))}, \tag{A.11}$$

for some given $e_2^H(0) = e^H \in \{0, 1\}$. To implement $e_2^L(1) = 1$, however, it is also necessary that $\tilde{w}(1; e^H, e^L) \geq w_2$, which can be written as

$$w_2 \leq \frac{\lambda c}{G(l - f(H, e^H - 1)) - G(m - f(H, e^H)))}. \tag{A.12}$$

Note that these two conditions do not hold simultaneously if $\overline{w}(1; e^H, e^L) > \tilde{w}(1; e^H, e^L)$. When $e^H = 1$, this becomes

$$\frac{c}{G(h-l) - G(h-m)} > \frac{\lambda c}{G(h-m) - G(m-l)}, \tag{A.13}$$

20

which holds because

$$\frac{c}{G(h-l)-G(h-m)} > \frac{c}{G(m-l)-G(0)} > \frac{\lambda c}{G(h-m)-G(m-l)}. \tag{A.14}$$

The second inequality holds by assumption 1. This implies that a necessary condition for $e_2^L(1) = 1$ is $e_2^H(0) = 0$ or, more precisely,

$$w_2 < \underline{w}(0; e^H, 1) = \frac{c}{G(h-m)-G(0)}. \tag{A.15}$$

Also, given that $e_2^H(0) = 0$, (A.11) becomes

$$w_2 \geq \frac{c}{G(m-l)-G(0)}. \tag{A.16}$$

This is a contradiction since (A.15) and (A.16) do not hold simultaneously. We can thus conclude that $e_2^L(1) = 0$.

Now suppose that $e_2^L(0) = 1$. Similarly as above, this implies that $w_2 \geq \overline{w}(0; e^H, 1)$, can be written as

$$w_2 \geq \frac{c}{G(m-l)-G(0)}. \tag{A.17}$$

For this contract to be sabotage-proof, we must also have $\tilde{w}(1; e^H, 1) \geq w_2$. These two conditions do not hold simultaneously if

$$\frac{c}{G(m-l)-G(0)} > \tilde{w}(1; e^H, 1) = \frac{\lambda c}{G(l-f(H, e^H-1))-G(m-f(H, e^H))}. \tag{A.18}$$

Again, this condition holds only when $e^H = 0$. We can then apply the same argument as above and show that $e_2^L(0) = 0$.

$$\text{Q.E.D.}$$

Given that $p_2^L = 0$, the best the principal can do is $(p_1^H = 1, p_1^L = 0)$. We now construct a contract which implements this. Note first that $\underline{w}(0; 1, 0) > \underline{w}(1; 1, 0)$. To implement $(p_2^H = 1, p_2^L = 0)$, therefore, it is necessary that $w_2 \geq \underline{w}(0; 1, 0)$. Besides this, any optimal contract must also satisfy (3) for low-ability agents (for any given $\mu_{i,2} \in \{0, 1\}$). Given $(p_1^H = 1, p_1^L = 0)$, the condition for this can be written as

$$\frac{(1+\lambda)c}{\mu[G(h-l)-G(m-l)]+(1-\mu)[G(l)-G(0)]} \geq w_2, \tag{A.19}$$

21

for some given $\mu_{i,2} = \mu$. It immediately follows from this that this condition is the most stringent when $\mu = 1$. Sabotage-proofness thus requires that

$$\frac{(1+\lambda)c}{G(h-l) - G(m-l)} > \underline{w}(0; 1, 0) = \frac{c}{G(h-l) - G(m-l)}, \tag{A.20}$$

which holds by assumption 1.

We now construct a contract that implements $(p_1^H = 1, p_1^L = 1)$. First, low-ability agents cannot influence the opponent's belief no matter what they do. This again implies that the problem for low-ability agents in period 1 is identical to that in period 2, and the same constraints apply for this case. To implement $(p_1^H = 1, p_1^L = 1)$, it is necessary that

$$w_1 \geq \overline{w}(\theta; 1, 1) = \frac{c}{\theta[G(h-l) - G(h-m)] + (1-\theta)[G(m-l) - G(0)]}. \tag{A.21}$$

Now suppose that a high-ability agent deviates and chooses $e_{i,1} = 0$. This signals to the opponent that he possesses low ability. The agent then exerts productive effort only if

$$\theta G(0) + (1-\theta)G(h-m) + u^{\mathrm{FT}}(1) - c \geq \theta G(m-h) + (1-\theta)G(0) + u^{\mathrm{FT}}(0), \tag{A.22}$$

which is reduced to

$$w_1 \geq \frac{c + u^{\mathrm{FT}}(0) - u^{\mathrm{FT}}(1)}{G(h-m) - G(0)}. \tag{A.23}$$

Given that the optimal second-period contract implements $(p_2^H = 1, p_2^L = 0)$, the deviation does not change the opponent's behavior: high-ability agents always exert productive effort while low-ability agents always exert no effort, regardless of the belief. This implies that $u^{\mathrm{FT}}(0) = u^{\mathrm{FT}}(1)$, and that the problem is again identical to that in period 1. Since $\overline{w}(\theta; 1, 1) > \underline{w}(\theta; 1, 1)$, the constraint for high-ability agents is not binding, and the optimal first-period contract is given by $w_1 = \overline{w}(\theta; 1, 1)$.

Q.E.D.

PROOF OF PROPOSITION 6: If $(p_1^H = 0, p_1^L = 1)$, then the ability type is not identifiable, i.e., $\mu_{i,2} = \theta$. To implement $(p_2^H = 1, p_2^L = 1)$, it is necessary that

$$w_2 \geq \overline{w}(\theta; 1, 1) = \frac{c}{\theta[G(h-l) - G(h-m)] + (1-\theta)[G(m-l) - G(0)]}. \tag{A.24}$$

22

For this to be sabotage-proof, we must also have

$$\tilde{w}(\theta; 1, 1) = \frac{\lambda c}{\theta[G(h - m) - G(m - l)]} \geq w_2. \tag{A.25}$$

The second-period contract must satisfy both of these conditions.

Given this, we now examine the first-period problem. We first consider incentives for low-ability agents. If a low-ability agent deviates from the equilibrium and chooses $e_{i,1} = 0$, this signals to the opponent his true type. Given the optimal contract in period 2, however, this does not change anyone's behavior because $\overline{w}(\theta; 1, 1) > \overline{w}(0; 1, 1) > \underline{w}(0; 1, 1)$. This implies that the problem for low-ability agents in period 1 is identical to that in period 2, and the same constraints apply for this case. To implement $(p_1^H = 0, p_1^L = 1)$, it is necessary that

$$w_1 \geq \overline{w}(\theta; 0, 1) = \frac{c}{G(m - l) - G(0)}. \tag{A.26}$$

The situation is more complicated for high-ability agents. If a high-ability agent deviates and chooses $e_{i,1} = 1$, this signals to the opponent his true type. This may lead the opponent to exert sabotage effort in period 2 if he is of the low-ability type. Taking this dynamic aspect into account, the incentive compatibility constraint needs some modification. It follows from (5) that productive effort is preferred to no effort for high-ability agents if

$$w_1 \geq \frac{c + u_2^H(0, 0) - u_2^H(1, 0)}{\theta[G(h - f(H, e^H)) - G(m - f(H, e^H))] + (1 - \theta)[G(h - f(L, e^L)) - G(m - f(L, e^L))]}. \tag{A.27}$$

It is easy to obtain the expected payoff on the equilibrium path:

$$u_2^H(0, 0) = [\theta G(0) + (1 - \theta)G(h - l)]w_2 - c. \tag{A.28}$$

It is, on the other hand, more complicated to see what happens off the equilibrium path. To this end, we first establish the following result.

Lemma 2 *Suppose that a high-ability agent deviates and chooses $e_{i,1} = 1$, and moreover that the second-period contract satisfies (A.24). Then, in period 2, (i) the opponent exerts sabotage effort if he is of the low-ability type and productive effort if he is of the high-ability type; (ii) the deviating agent exerts productive effort.*

23

PROOF: We examine what happens off the equilibrium path, when a high-ability agent deviates. Suppose first that the opponent is of the low-ability type. The deviation then induces sabotage effort because

$$\frac{c}{\theta[G(h-l) - G(h-m)] + (1-\theta)[G(m-l) - G(0)]} > \tilde{w}(1;1,1) = \frac{\lambda c}{G(h-m) - G(m-l)}, \tag{A.29}$$

by assumption 1. Now suppose that the opponent is of the high-ability type. In this case, the opponent still exerts productive effort because $\overline{w}(\theta;1,1) > \underline{w}(1;1,1)$.

Given this, the deviating agent still chooses productive effort if

$$[\theta G(0) + (1-\theta)G(m-l)]w_2 - c \geq [\theta G(m-h) + (1-\theta)G(0)]w_2, \tag{A.30}$$

$$[\theta G(0) + (1-\theta)G(m-l)]w_2 - c \geq [\theta G(0) + (1-\theta)G(l)]w_2 - (1+\lambda)c. \tag{A.31}$$

The first condition holds because

$$w_2 \geq \overline{w}(\theta;1,1) > \frac{c}{\theta[G(h-m) - G(0)] + (1-\theta)[G(m-l) - G(0)]}. \tag{A.32}$$

The second condition always holds for any nonnegative w_2.

Q.E.D.

This implies that the expected payoff is given by

$$u_2^H(1,0) = [\theta G(0) + (1-\theta)G(m-l)]w_2 - c. \tag{A.33}$$

The incentive compatibility constraint (A.27) can then be written as

$$w_1 \geq \frac{c + (1-\theta)[G(h-l) - G(m-l)]w_2}{\theta[G(h-f(H,e^H)) - G(m-f(H,e^H))] + (1-\theta)[G(h-f(L,e^L)) - G(m-f(L,e^L))]}. \tag{A.34}$$

To implement $(p_1^H = 0, p_1^L = 1)$, therefore, the following condition must be satisfied:

$$\frac{c + (1-\theta)[G(h-l) - G(m-l)]w_2}{G(h-m) - G(0)} > w_1. \tag{A.35}$$

To sum up, any contract that implements $(p_1^H = 0, p_1^L = 1)$ must satisfy (A.24), (A.25), (A.26), and (A.35). These conditions are summarized as

$$\frac{\lambda c}{\theta[G(h-m) - G(m-l)]} > w_2 \geq \frac{c}{\theta[G(h-l) - G(h-m)] + (1-\theta)[G(m-l) - G(0)]}. \tag{A.36}$$

24

$$\frac{c + (1-\theta)[G(h-l) - G(m-l)]w_2}{G(h-m) - G(0)} > w_1 \geq \frac{c}{G(m-l) - G(0)}. \tag{A.37}$$

An obvious candidate is

$$w_1 = \frac{c}{G(m-l) - G(0)}, \ w_2 = \frac{c}{\theta[G(h-l) - G(h-m)] + (1-\theta)[G(m-l) - G(0)]}.$$

This is optimal if it satisfies (A.37):

$$\frac{c + (1-\theta)[G(h-l) - G(m-l)]w_2}{G(h-m) - G(0)} > \frac{c}{G(m-l) - G(0)}, \tag{A.38}$$

which can be written as

$$\frac{(1-\theta)[G(h-l) - G(m-l)]}{\theta[G(h-l) - G(h-m)] + (1-\theta)[G(m-l) - G(0)]} > \frac{G(h-m) - G(m-l)}{G(m-l) - G(0)}. \tag{A.39}$$

This condition holds when θ is sufficiently small.

<div align="right">Q.E.D.</div>

References

Bernhardt, Dan, 1995, Strategic Promotion and Compensation, *Review of Economic Studies*, 62, 315-39.

Chen, Kong-Pin, 2003, Sabotage in Promotion Tournaments, *Journal of Law, Economics, and Organization*, 19, 119-40.

Gibbons, Robert and Waldman, Michael, 1999a, Careers in Organizations: Theory and Evidence, in Handbook of Labor Economics, vol. 3, eds., Ashenfelter, Orley C. and Card, David, Amsterdam: North Holland.

Gibbons, Robert and Waldman, Michael, 1999b, A Theory of Wage and Promotion Dynamics inside Firms, *Quarterly Journal of Economics*, 114, 1321-58.

Grossman, Gene M. and Maggi, Giovanni, 2000, Diversity and Trade, *American Economic Review*, 90, 1255-75.

Ishiguro, Shingo, 2004, Collusion and Discrimination in Organizations, *Journal of Economic Theory*, 116, 357-69.

Lazear, Edward P., 1979, Why is There Mandatory Retirement?, *Journal of Political Economy*, 87, 1261-84.

Lazear, Edward P., 1989, Pay Equality and Industrial Politics, *Journal of Political Economy*, 97, 561-80.

Lazear, Edward P. and Rosen, Sherwin, 1981, Rank-Order Tournaments as Optimum Labor Contracts, *Journal of Political Economy*, 89, 841-64.

Meyer, Margaret A., 1992, Biased Contests and Moral Hazard: Implications for Career Profiles, *Annales d'Economie et de Statistique*, 25/26, 165-87.

Mookherjee, Dilip, 1984, Optimal Incentive Schemes with Many Agents, *Review of Economic Studies*, 51, 433-46.

Owan, Hideo, 2004, Promotion, Turnover, Earnings, and Firm-Sponsored Training, *Journal of Labor Economics*, 22, 955-78.

Prendergast, Canice, 1992, Career Development and Specific Human Capital Collection, *Journal of the Japanese and International Economies*, 6, 207-27.

Prendergast, Canice, 1999, The Provision of Incentives in Firms, *Journal of Economic Literature*, 37, 7-63.

Waldman, Michael, 1984, Job Assignments, Signalling and Efficiency, *Rand Journal of Economics*, 15, 255-67.