A Theory of Favoritism

Zhijun Chen*

2013-04-04

Abstract

Favoritism prevails in organizations that rely on subjective assessments of employee performance, and its harmful impact on the efficiency is widely recognized. This paper shows that favoritism could benefit the employer when collusion among employees becomes a serious threat in organizations. Favoritism differentiates the incentive constraints for the agents, and adequate favoritism reduces the cost for preventing collusion but excessive favoritism increases the incentive cost.

JEL Classification: C72, D82

Keywords: Favoritism, Collusion, and Tournament

*University of Auckland. Contact e-mail: chenzj1219@gmail.com.

1 Introduction

Favoritism prevails in a variety of organizations. The harmful impacts of favoritism on productivity and efficiency have been widely recognized. For instance, it is argued that favoritism is one of the most important sources of conflicts in organizations,[1] and that it results to the distortion of incentives.[2] This begs the question: why do employers (or supervisors) play favoritism albeit the resulting harmful impacts? One simple answer is that employers may have intrinsic preference over some employees and they can derive a utility from playing favoritism. While this altruistic reason for favoritism might be relevant in organizations where the appraiser is not the residual claimant and may act on his own preference, it does not bite when the employer is the residual claimant of the organization who aims at maximizing its own benefit.

It appears that there must exist some efficiency enhancing motivations for playing favoritism in organizations which could well offset the above-mentioned perverse impacts. To disclose the non-altruistic rationale for favoritism, it is essential to examine the key incentive issues in organizations where favoritism emerges. Favoritism prevails in organizations where objective measurements of employee performance are unavailable and thus incentive schemes are designed on the basis of employers' subjective assessments.[3] Subjectivity of performance assessments then opens a door to favoritism, where employers act on personal preference toward subordinates to favor some employees over others.

In these organizations that rely on subjective assessments of performance, typical incentive contracts often take the form of tournament where a prize is committed to the winner of the contest.[4] The commitment of fixed prize mitigates the employer's opportunism and, when

[1] In a survey of Canadian government workers, Comerford (2002) finds that favoritism is the second most important source of workplace conflict followed by workload; while Albright and Carr (1997) list favoritism as one of the top ten misconducts against workers that mitigate working incentives. In the Alpha Review by Burke Croup Minnesota, Inc., it is even argued that favoritism is a cancer in organizations.

[2] Prendergast and Topel (1993) investigate the phenomena of discretion and bias in performance evaluation in organizations, and argue that favoritism can give rise to inefficiency on two margins. The first is rent seeking by workers, which is usually a waste of time. The second occurs because bias makes it difficult to determine the true talents of workers.

[3] As argued by Prendergast and Topel (1993, 1996), while most of the economics literature on incentives in organizations focuses on situations where compensation schemes can be made based on objective performance measures such as output or sales, it ignores the fact that most compensation arrangements involve superiors' *subjective*, and hence non-contractible, judgements about employee performance.

[4] Prendergast (1999) gives an excellent survey for the provision of incentives in such organizations.

the prize for the contest winner is sufficiently high, competition between employees provides strong incentives to promote high levels of efforts.[5] However, tournaments are not robust to collusion. Because the outcome of the tournament is determined by the relative performance of the employees, which is related to the difference of their effort levels rather than the absolute levels of efforts, when the employees cut their efforts collectively, their expected payoffs are not affected since the outcome of the tournament is unchanged, and employees benefit from saving their effort costs.

The phenomenon of collusion is prominent in organizations.[6] As a response, the design of incentive mechanisms must take into account the possibility that employees collude to manipulate their efforts. This paper shows that, when collusion between employees becomes a serious problem, the employer can benefit from playing favoritism by reducing the cost for preventing collusion, whereas favoritism does not increase the welfare in the absence of collusion. We demonstrate the insights in a stylized model of tournament, where there is one principal (the employer) and two homogenous agents (the employees). Each agent is assigned a project independently, and the output of the project depends on the agent's effort (high or low) and is also affected by a random shock. The principal benefits from the high level of outputs and is the residual claimant of the projects. The agents' efforts are not observable by other parties and, to overcome the moral hazard problem, the principal must provide the proper incentives for the agents. A fixed prize is committed to the winner of the contest, but the selection of the winner is based on the principal's subjective assessment of agent's relative performance as objective measurements of outputs are not available, and the principal may favor one agent over another by overestimate the output of the favored agent.[7] However, the principal does not derive a utility from playing favoritism, since the allocation of the fixed prize does not affect the principal's payoff ex post. In other words, we assume away the altruistic motivation of playing favoritism and focus instead on the non-altruistic incentive effects.

Favoritism here takes the form of bias in the subjective assessment of relative performance and thus changes the probability of winning for the agents, which causes different incentive impacts: it increases the probability of winning for the favored one while on the other hand

[5] For instance, see Green and Stokey (1983), Lazear and Rosen (1981), Nalebuff and Stiglitz (1983).

[6] See Tirole (1986, 1992) for detailed discussion. As collusive behaviors are always conducted in secret, what we have observed is only the tip of the iceberg.

[7] For instance, if the output of the favored agent is 100, the principal can overestimate to 120. Thus the favored agent is more likely to win the prize of the tournament than his peer given other conditions equal.

decreases that for the disfavored one, given other things equal. While it could provide stronger incentives for the favored agent,[8] favoritism appears to suppress the incentives for the disfavored agent to take the high effort, which in turn calls for higher incentive cost (i.e., the tournament prize) to induce the high effort than absent favoritism. As the principal aims to induce the high efforts from both agents, it indeed pays for the principal to play favoritism in this way. Thus, in the absence of collusion, the principal does not gain from favoritism.

Tournaments are vulnerable to collusion: the agents could instead take the low efforts collectively and benefit themselves from saving the effort costs. Sustaining collusion involves some non-juridical enforcement mechanisms such as social norms, reputation concerns, as well as long-term relationship.[9] Since we are not motivated in this paper to investigate the collusion-enforcement mechanisms in tournaments, we would rather take a short-cut in modelling collusion which has been widely used in the literature of collusion-proof mechanism design. Following the modelling approach of Tirole (1992), we assume that collusive agreements between agents could be enforced by a mediator, which can be viewed as a modelling short-cut of some non-judicial enforcement mechanism such as "word of honor", and that enforcing the side contract incurs an efficiency loss for the collusive coalition, which reflects the feature of non-judicial enforcement.

Since agents' efforts are not observable, to mitigate the incentives for deviating to the high effort unilaterally, the side payment from the winner to the loser must be used to reduce the payoff gap between the states of winning and losing.[10] For instance, when agents are able to split the prize equally then no one has incentives to take the high effort since each agent is fully insured with a fixed payment regardless of winning or losing.

The agents are treated unequally under favoritism such that the favored agent could earn a higher expected payoff than the disfavored one given other things equal. This implies that the agents should be also treated asymmetrically in collusion such that the favored agent would be granted the higher stake of collusion, and this can be achieved only by differentiating the side payments for the agents.[11] Moreover, while granting some bias in the performance evaluation

[8] This effect prevails only if the degree of favoritism is not excessive; see the discussion later.

[9] See Tirole (1992) for detailed discussion.

[10] Since collusion is enfored by non-judicial mechanisms, side transfers often involves a dead-weight loss. For instance, a $100-dollar-value of gift from the donnar (the winner of the tournament) may worth only $80 to the receiver (the loser of the tournament).

[11] That is, the favored agent should pay less side payment as a winner than the disfavored one, say, for instance, the favored agent as the winner should give 40% of the prize to the loser, while the disfavored one as the winner should offer 60% of the prize to the loser (the favored one).

would suppress the incentives of the disfavored agent, it does provide stronger incentives for the favored agent to take the high effort. Hence, favoritism differentiates the incentive constraints for the agents to take the high efforts: it relaxes the incentive constraint for the favored agent but on the other hand tightens the constraint for the disfavored one, whereas both agents have the same incentive constraint without favoritism. In the absence of collusion, such differentiation entails higher incentive cost since the principal has to incentivise both agents. Whereas, when agents are able to collude, differentiating the incentive constraints indeed reduces the incentive cost for preventing collusion since the principal only needs to induce one agent (say, the favored one) to deviate for this purpose. That is, the incentive constraint for the deviation of the favored agent is less stringent than that absent favoritism, and as a result the principal needs to pay less for the favored agent to deviate from collusion.

However, the effect of differentiation in incentive constraints arises only for some degree of favoritism. Granting excessively high bias in the performance evaluation would insure the favored agent an excessively high probability of winning even if it takes the low effort, which in turn discourage the favored one to take the high effort.[12] This indicates that excessive favoritism would indeed makes the principal worse off.

Summarizing the above analysis, we have three main results. Favoritism does not benefit the principal in the absence of collusion; however, play some degree of favoritism contributes to reduce the incentive cost for preventing collusion. Finally, excessive favoritism is not desirable. The optimal degree of favoritism thus minimizes the principal's incentive cost for collusion-proofness and can be determined endogenously.

This paper is closely related to the seminal paper of Prendergast and Topel (1996) (hereafter PT), but it departs from their paper in several key aspects. First, the motivations for favoritism are different in two papers. PT studies the organization with the vertical relationship of employer-supervisor-worker, where the supervisor has the authority of discretion on the subjective assessment of the worker's performance and moreover values the power of exercising favoritism. In other words, in their model favoritism takes a form of altruism since the supervisor derives a utility from favoritism. By contrast, we focus the organizations with the relationship of principal-multiagent, where the principal is the residual claimant of the organization and does not derive a utility from exercising favoritism. That is, we study the non-altruistic motivation

[12] Consider for instance the extreme case of favoritism where the principal grants infinite bias in the performance evaluation so that the favored agent will win the prize with probability one. Obviously, the favored one has no incentives to take the high effort in this case.

for favoritism.

Second, the incentive mechanisms at play are also different. In their paper, while the supervisor derives an extra utility by exercising favoritism, it also has to bear the cost of overestimating the worker's performance, thus the optimal bias balances the trade-off. In our model, favoritism differentiates the incentive constraints for the agents, which helps in reducing the incentive cost for preventing collusion, but excessive favoritism mitigates this cost-reducing effect, and the optimal bias minimizes the incentive cost for collusion-proofness.

Third, our paper derives different main implications on favoritism in organizations from theirs. PT shows whether favoritism is harmful or beneficial depends crucially on the existence of distortions in the "market" for favoritism. If the firm can charge supervisors an optimal price for exercising their preferences and if the only cost of favoritism is the risk it imposes on workers, then the incentives will be set as though favoritism did not exist. Whereas, we show that favoritism does not benefit the employer absent collusion but it allows the employer to reduce the cost for collusion-proofness, and that excessive favoritism is not desirable.

Following the altruistic modelling approach of PT, Berger, Herbertz, and Sliwka (2011) also consider the organizations with one manager and two (heterogeneous) agents. The principal benefits from the agents' efforts but also derives a utility from favoring one agent over another. They show that favoritism leads to a lower quality of promotion decisions and in turn lower efforts, but the effect can be mitigated by pay-for-performance incentives for the manager. That is, making the manager the residual claimant could mitigate his incentives of playing favoritism, which coincides with the first result of our paper.

Our paper also relates to Kwon (2006), which shows that favoritism can arise endogenously as an optimal decision rule in a model of strategic delegation of decisions where two experts produce competing ideas with conflicts in preference. In his paper, favoritism is equivalent to the delegation of authority to the favorite, and the altruistic motivation of favoritism is also assumed away.

This research is also related to the literature of collusion-proof mechanism design. In particular, Ishiguro (2004) studies the discriminatory incentive scheme in fighting collusion in tournaments, in the sense that the wage schemes offered to agents depend on their identities. That is, the favored agent could win the prize if and only if his performance is better than his peer (there is no bias in the subjective assessment), but the discriminated one is excluded for winning the prize whatever his performance is. As a result, only the favored one will take the high effort. However, such discriminatory incentive scheme is not robust to the opportunism since the principal has incentives to renege the payment by claiming that the winner of the tournament is the discriminated agent and needs to pay nothing in this case. Moreover, this discriminatory policy is not commonly observed as naked discrimination is illegal under most jurisdictions. In our paper, favoritism takes instead a form of bias in subjective performance evaluation, which is commonly observed in organizations.

We set up the model of tournament in Section 2, and then show in Section 3 that favoritism does not benefit the principal when agents are unable to collude. We discuss the issue of collusion in Section 4 and further show the desirability of introducing favoritism under the threat of collusion. The determination of optimal degree of bias is analyzed in Section 5, and finally Section 6 concludes the paper.

2 The Model

We set up a simple model representing a stylized organization, which consists of an employer (call it the principal) and two employees (call them the agents). The two agents, as denoted by A^1 and A^2, are recruited from a competitive labor market with a reservation payoff equal to 0, and possess the same production skills. For simplicity, we assume that the principal is risk-neutral and the agents are also risk-neutral but are protected by the limited liability.

There are two identical projects and each agent takes one project independently. The output of each agent A^i, $i = 1, 2$, is given by $y^i = e^i + \varepsilon^i$, where e^i is the effort level of agent A^i and ε^i represents a random shock with zero mean; the output y^i can be commonly observed ex post. The random variables ε^i, $i = 1, 2$, are identically and independently distributed with a symmetric distribution function $F(\cdot)$ on $\mathbb{R}$, where $F(\cdot)$ is twice-differentiable and has symmetric properties: $F(0) = 1/2$ and $F(x) = 1 - F(-x)$ for any $x \in \mathbb{R}$. The corresponding density function, as denoted by $f(.)$, thus satisfies $f(x) = f(-x)$ for any $x \in \mathbb{R}$ and we assume further that $f(x)$ is weakly decreasing for $x \geq 0$.[13]

While in principal the agents can choose effort levels continuously, for the tractability of analysis, we adopt the approach of discrete effort levels. Moreover, for the simplicity of exposition, we assume that each agent can choose two effort levels, namely high or low, as denoted by$e^i = h$ or $e^i = 0$ respectively, and that their efforts are not observable by others. We denote by $C(e^i)$ the agents' effort cost, and furthermore normalize the effort costs to $C(0) = 0$ and

[13] This property is satisfied for quite general distributions including uniform distribution and normal distributions.

$C(h) = c > 0$.

The outputs of the agents, y^i, $i = 1, 2$, are observable but not verifiable. In other words, objective measurements of the outputs are not available here, thus incentive contracts cannot rely on the absolute or relative performance. As a result, the principal adopts a tournament mechanism in order to incentivise the agents for taking high efforts. The simple tournament mechanism comprises a fixed prize t for the winner (and only the winner) of the contest and an assessment rule for the selection of the winner.[14] More precisely, an assessment rule specifies conditions under which an agent will win in the tournament. For instance, an unbiased rule commits that the agent A^1 wins the prize if and only if he has a better performance than his peer, that is, $y^1 > y^2$.

Due to subjectivity of performance assessment, the principal has the right of discretion in the selection of the winner. This opens a door for favoritism where the principal might act on personal preference toward the agents to favor one agent. We assume away the motivation of altruism in favoritism and instead focus on the case that favoritism does not increase the principal's utility directly.[15] The principal could overestimate the performance of the favored agent by granting additional value b ($b \geq 0$) in the output, but cannot renege on the prize whoever wins in the tournament. For instance, if the principal favors the agent A^1, he could announce that the output of this agent is $y_1 + b$, albeit that his real output is y_1.

We focus on the case where the principal favors some agent explicitly such that the identity of the favored agent and the (biased) assessment rule are commonly known. Let subscripts "f" and "d" stand for the status of "favored" and "disfavored" respectively. Under the biased assessment rule, the favored agent A_f wins if only if $y_f + b > y_d$. Since the principal does not gain directly from favoring one agent, and the prize for the winner is fixed whoever wins, the principal has no incentives to renege on the (biased) assessment rule ex post.

An agent's output y^i depends on its effort e^i as well as the random shock ε^i, the probability distribution function of y^i is then given by $P_r\{y^i \leq y\} = F(y - e^i)$. However, the probability

[14] The agent can also get a basic wage whatever he wins or loses, which is normalized to zero for simplicity. It is a well-known result that the optimal tournament contract must involve zero basic wage when agents are risk-neutral and protected by the limited liability.

[15] Prendergast and Topel (1996) also investigated the phenomenon of favoritism in organizations where supervisors often impose some bias in their evaluations of workers' performance according to their own preferences. They assume that the supervisor's utility depends on the pay of his subordinate, and the supervisor favors some subordinate simply because such favoritism increases his utility (a kind of altruism). Instead, in our model, favoritism does not improve the principal's welfare directly.

of winning for agent A^i is dependent of relative performance between agents. In the absence of favoritism, the agent A^i wins the prize if and only if $y^i > y^j$, and the probability of winning is thus given by

$$P_r\{y^i > y^j\} = P_r\{\varepsilon^j - \varepsilon^i \leq e^i - e^j\} = G(e^i - e^j),$$

where $G(\cdot)$ is the distribution function of $\varepsilon^i - \varepsilon^j$ and is derived by

$$G(x) \equiv P_r\{\varepsilon^i - \varepsilon^j \leq x\} = \int_{-\infty}^{+\infty} F(x+\varepsilon)f(\varepsilon)d\varepsilon.$$

Note that $G(\cdot)$ has a symmetric property such that $G(x) = 1 - G(-x)$, which follows from

$$1 - G(-x) = P_r\{\varepsilon^i - \varepsilon^j \geq -x\} = P_r\{\varepsilon^j - \varepsilon^i < x\} = G(x),$$

and it is also easy to check that its density function $g(\cdot)$ inherits the properties of $f(\cdot)$.

With favoritism, the biased assessment rule gives the favored agent a bias b in the evaluation of its performance. Denoting by $e \equiv (e_f, e_d)$ the pair of efforts where e_f (resp. e_d) represents the effort of the favored (resp. disfavored) agent, the probability of winning for the favored agent can thus be written as

$$p_f(e;b) \equiv P_r\{y_f + b > y_d\} = P_r\{\varepsilon_d - \varepsilon_f \leq (e_f + b) - e_d\} = G(e_f - e_d + b).$$

By contrast, the disfavored agent is imposed a bias $-b$ in the evaluation, and the probability of winning is thus give by

$$p_d(e;b) = P_r\{y_d > y_f + b\} = 1 - p_f(e;b) = G(e_d - e_f - b).$$

It appears that the probability of winning is dependent of the difference of effort levels plus the bias. Moreover, given that both agents take the same efforts, then the favored agent is more likely to get the prize than disfavored one as $p_f(e;b) = G(b) > p_d(e;b) = G(-b)$ when $e_d = e_f$.

Under this tournament mechanism, the favored agent earns an expected payoff $U_f(e;b) = p_f(e;b)t - C(e_f)$ while the disfavored one obtains $U_d(e;b) = p_d(e;b)t - C(e_d)$. The principal's net benefit can be expressed as $ER(y_f, y_d) - t$, where $ER(y_f, y_d)$ is the expected revenue and is assumed to be increasing in the efforts. We assume that the expected revenue when both agents take high efforts is much higher than that with low efforts, and that the extra gain of the expected revenue is much greater than the effort cost c, so that it is always desirable to induce the high efforts. This allows us to focus on the implementation problem where the principal aims to induce the high efforts with the minimum incentive costs. We assume further that the inverse hazard rate $H(x) \equiv G(x)/g(x)$ is weakly increasing, which holds for a variety of distributions including normal distribution and uniform distribution.

3 Tournaments Absent Collusion: A Benchmark

As a benchmark, we first examine the incentive effects of the tournament when agents are unable to collude. Given that the principal's incentive scheme, which consists of the fixed prize t and the (biased) assessment rule, each agent chooses the effort level independently. The favored agent is willing to exert the high effort in the tournament if the expected payoff (expecting that the other party will also take the high effort) $G(b)t$ exceeds the effort cost c, which implies that the tournament prize must be high enough such that $t \geq c/G(b)$. Similarly, the disfavored agent is willing to take the high effort if $t \geq c/G(-b)$.

However, since the efforts cannot be observed by other parties, the agents may instead take the low effort unilaterally to save the effort costs. Thus, to implement the high efforts in the Nash equilibrium, the agents must be prevented from deviating unilaterally and taking instead the low effort. For the favored agent, taking low effort unilaterally yields an expected payoff equal to $G(b-h)t$, and such deviation reduces the probability of winning from $G(b)$ to $G(b-h)$, which incurs an expected loss equal to $(G(b) - G(b-h))\,t$. Hence, the favored agent is prevented from such unilateral deviation if the expected loss $(G(b) - G(b-h))\,t$ overweighs the gain from cost saving c, which in turn requires that the principal offers sufficiently high incentive prize such that

$$t \geq T_f^a(b) \equiv \frac{c}{G(b) - G(b-h)}.$$

Notice that, since the probability function satisfies $G(x) = 1 - G(-x)$, i.e., the probability of winning for the favored agent is equal to the probability of losing for the disfavored one, it follows that $G(b) - G(b-h) = G(h-b) - G(-b)$. This implies that, when the favored agent takes low effort unilaterally, the decrease of his winning probability is equal the increase of the winning probability for the disfavored one,[16] thus the threshold $T_f^a(b)$ is also equal to $\frac{c}{G(h-b)-G(-b)}$.

Since $T_f^a(b) > c/G(b)$, the principal has to pay extra incentive cost to mitigate the moral hazard problem of the favored agents, this extra cost is known as the information rent in the literature of incentive theory.[17] By analogy, the principal must provide extra incentives to overcome the moral hazard problem from the disfavored agent, which requires that the tournament prize must satisfy[18]

$$t \geq T_d^a(b) \equiv \frac{c}{G(-b) - G(-b-h)} = \frac{c}{G(h+b) - G(b)}.$$

[16] Notice that this property holds only for the case with two agents.

[17] See, for instance, the textbook of Laffont and Martimort (2002).

[18] It is straighforward to see $T_d^a(b) > c/G(-b)$.

Hence, both the favored and disfavored agents will take the high efforts in the Nash equilibrium if and only if

$$t \geq T^a(b) \equiv \max\{T^a_f(b), T^a_d(b)\},$$

and moreover this equilibrium is unique.[19]

Lemma 1 *When agents are unable to collude, they will take the high efforts in the unique Nash equilibrium if and only if $t \geq T^a(b)$.*

Proof. See Appendix A. ■

It is straightforward to see that the two thresholds are "symmetric" in the sense that $T^a_f(b) = T^a_d(-b)$, since the winning probability of agents is determined by the difference of their efforts plus the bias, and the fact that the favored agent receives a bias b implies that the disfavored one receives a bias $-b$.

As mentioned above, the principal aims to induce the high efforts at the minimum cost, which amounts to minimizing the prize $T^a(b)$ by choosing b ($b \geq 0$). Obviously, the principal only needs to provide the same incentive costs for both agents if there were no favoritism, that is, $T^a_d(0) = T^a_f(0)$. However, whether introducing favoritism could reduce the incentive cost depends on the properties of the thresholds $T^a_f(b)$ and $T^a_d(b)$. Differentiating $T^a_d(b)$ with respect to b, it is then straightforward to see that the incentive cost for the disfavored agent increases in the degree of favoritism:

$$\frac{dT^a_d(b)}{db} = \frac{g(b) - g(h+b)}{(G(h+b) - G(b))^2} \geq 0,$$

as $g(x)$ decreases with x for all $x \geq 0$ (since $g(\cdot)$ inherits the properties of $f(\cdot)$).[20] That is, the more bias against the disfavored agent, the higher incentive cost needs to compensate him for taking the high effort.

The intuition is indeed very simple. Suppose the principal offers $t = T^a_d(b)$, which makes the disfavored agent indifferent between taking the high and low efforts (and we assume that he will take the high effort in this case): $G(-b)\, T^a_d(b) - c = G(-b-h)T^a_d(b)$. Increasing now the bias b to b' reduces the expected payoff of the disfavored agent for taking the high effort as well as the low effort (both $G(-b)\,t$ and $G(-b-h)t$ decrease with b), however, the first effect dominates the second one as $G(-b)$ decreases quicker than $G(-b-h)$ under the

[19] This is the well-known result in the literature of tournaments. See Prendergast (1999) for an excellent survey.

[20] It is strictly increasing in b for any non-uniform distribution function.

assumption of the density function. As a result, $G(-b')\,T_d^a(b) - c < G(-b'-h)T_d^a(b)$, and the principal must then increase the incentive prize from $t = T_d^a(b)$ to $t' = T_d^a(b')$ such that $G(-b')\,T_d^a(b') - c = G(-b'-h)T_d^a(b')$. In other words, granting more bias against the disfavored agent tightens the incentive compatibility constraint and, as a result, the principal must provide higher compensation to induce the high effort.

On the other hand, one would expect that introducing favoritism could provide strong incentives for the favored agent and thus reduce the principal's incentive cost. This is, however, not true. Recall that the favored agent will take the high effort if $G(b)\,t - c \geq G(b-h)t$, and both $G(b)$ and $G(b-h)$ increase with b. Thus, whether offering more favoritism could relax the favored agent's incentive constraint depends on whether $G(b)$ increases faster than $G(b-h)$, or equivalently, whether the derivative $g(b)$ is greater than the derivative $g(b-h)$. Notice that $g(b) > g(b-h)\,(= g(h-b))$ if and only if $b < h/2$ (i.e., $b < h - b$).[21] This implies that $G(b)$ increases faster than $G(b-h)$ if and only if the degree of favoritism is less than half of the effort gap (i.e., $b < h/2$), in which case increasing bias could provide stronger incentives for the favored agent to take the high effort and thus the principal needs to pay less incentive cost ($T_f^a(b)$ decreases with b). Whenever $b > h/2$, however, granting more favoritism to the favored agent would instead provide stronger incentives to take the low effort, which makes the principal even worse off (see Figure 1 for illustration).

Therefore, excessive favoritism (i.e., $b > h/2$) mitigates the incentives for taking the high efforts, for both the favored and disfavored agents. While non-excessive favoritism ($b < h/2$) could reduce the incentive cost for the favored one, it also increases the compensation for the disfavored one. As the principal must encourage both types of agents to take the high effort, and since $G(b) - G(b-h) \geq G(h+b) - G(b)$, which implies $T_d^a(b) \geq T_f^a(b)$ and thus $T^a(b) = T_d^a(b)$ (see Appendix B for the proof), the cost-increasing effect always dominates the cost-decreasing effect. It follows that the incentive cost $T^a(b) = T_d^a(b)$ is minimized at $b = 0$,[22] and the equilibrium prize is given by

$$T^a(0) = \frac{c}{G(h) - G(0)}.$$

[21] See Appendix B for detailed proof.

[22] When the distribution function is non-uniform, then $T_d^a(b)$ is strictly increasing in b, in which case favoritism is strictly dominated by non-favoritism. When the distribution function is uniform such that $G(b) - G(b-h) = G(b+h) - G(b)$ for any $b \geq 0$, then $T_d^a(b) = T_l^a(b)$ and both incentive costs are independent of b, in which case, favoritism does not benefit the principal.

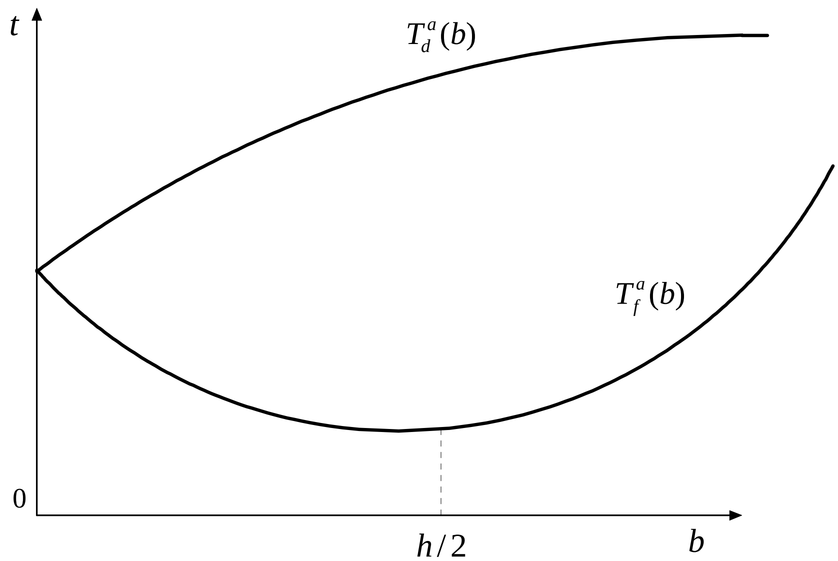

Figure 1

Summarizing the above analysis leads to our first result:

Proposition 1 *Favoritism does not benefit the principal when the agents are unable to collude; instead it makes the principal strictly worse-off when the density function $g(\cdot)$ is strictly decreasing for some b close to zero.*

Proof. See Appendix B. ■

4 Collusion and Favoritism

Tournaments are vulnerable to collusion. Since the probabilities of winning are determined by the difference of effort levels (plus the bias), i.e., $p_f(e;b) = G(e_f - e_d + b)$ and $p_d(e;b) = G(e_d - e_f - b)$, cutting efforts collectively with the same amount does not affect the expected gain of each agent, however each agent benefits from saving the effort cost.

Collusion among subordinates in the workplaces are often sustained by non-judicial mechanisms such as reputation, social norms or reciprocity in long-term relationship.[23] We are not

[23] For instance, Miller (1992) describes a so-called "binging" game played between workers when discussing the compensation scheme of the bank wiring room in the Hawthorne plant of Western Electric; this game is played

motivated to study the collusion-enforcement mechanisms in this paper, and will thus simply take the formation of coalition as given. For the simplicity of analysis and moreover following the methodology of the literature of collusion-proof mechanism design pioneered by Tirole (1986, 1992), we assume that collusion among agents is enforced by a mediator. The mediator will design the collusive agreement for the agents, and then enforce the agreement when it is approved by both parties.

A typical collusive agreement must specify the effort that each agent will take, and the side payment transferred from the winner to the loser. Since the agents' efforts are not observed by other parties, the side payment plays an essential role here in this one-shot game, which can mitigate the incentives for each agent to deviate from the collusive agreement and take the effort different from the agreed level unilaterally.

To see how the mechanism works, consider a simple example where there is no favoritism. The principal offers the prize $t \geq T^a(0)$, which is sufficient to induce the high efforts in the absence of collusion from the above analysis. Suppose now the mediator proposes that both agents take the low efforts, without imposing any side payments, then each agent has incentives to deviate unilaterally. Given that the other party taking low effort, the deviating agent would earn an expected payoff $G(h)\,t - c$ by taking the high effort, which exceeds the payoff when taking the low effort $G(0)\,t$ as $t \geq T^a(0)$, thus collusion cannot be sustained without side payments. To ensure that no one has incentives to take high efforts under collusion, the mediator must use side payments to reduce the payoff gap between the winner and loser. Consider a side-payment transfer scheme that makes the winner and loser to share the prize equally. Under such scheme, the agents are fully insured regardless whoever wins or loses, and each one is granted a fixed payoff equal to half of the prize. Thus, no one has incentives to take the high effort.

In reality, the collusive agreements are often enforced by non-judicial mechanisms. While in practice the "technologies" for side transfers are diversified largely, from monetary bribes to friendly relationships, in most organizations personal monetary transactions between employees are prohibited. In general, the cost of side transfer for the donor differs from the value for the recipient, and thus side transfers often incur a dead-weight loss due to the inefficiency of enforcement and/or the restrictions of the legal environments. To highlight such efficiency loss in the enforcement of collusive agreement, we follow Tirole (1992) and assume that a side payment

to punish the workers who produce too much, which is indeed a collusion enforcement device to prevent workers from exerting high effort levels. Moreover, collusive agreements among employees that aim to reduce their efforts collectively are not prohibited by law and can even be enforced by mediators like labor unions.

s from the donor is worth of ks to the recipient, where $k \in (0,1)$ is a parameter measuring the efficiency of collusion and its value is common knowledge. That is, there is a deadweight loss of $(1-k)\,s$ due to the inefficiency of the collusion enforcement mechanism. One may also think the loss of $(1-k)\,s$ as the enforcement fee charged by the mediator.

Remark: Enforceable versus self-enforcing side contracts. The theory of side contracting can be built in two ways. The first approach assumes that side contracts are enforceable and the second approach traces the foundations of enforceability to repeated interaction and reputation. Tirole (1992, p.156) argued that "The enforceability approach seems more innocuous when collusion is enforced by word of honor. When enforcement is ensured by repeated interaction and reputation, enforceable side contracts at best depict a polar case in which reputation mechanisms work well to enforce collusion (the word-of-honor paradigm can be viewed as an extreme case of a reputation model in which the prior probability of being trustworthy is equal to one)." Moreover, "These remarks may reflect some amount of cognitive dissonance since the literature has embraced the enforceability approach, and self-enforceability seems important in practice. The reason for this methodological choice is that the enforceability approach allows the use of classical contract theory by describing the organization as a nexus of contracts.... I believe that the enforceability approach may offer a realistic description of side contracting, and that it still yields precious insights when it does not."

The game is simplified thanks to this modelling approach, and the timing of game can be illustrated as follows:

Stage 1. The principal proposes a tournament contract, each agent then decides to accept or not; if no one rejects then:

Stage 2: A mediator proposes a side contract and each agent then decides to accept or reject; if no one vetoes then:

Stage 3: Each agent takes the effort simultaneously;

Stage 4: Outputs are realized and contracts (the tournament contract and the side contract) are enforced.

A tournament contract consists of a prize $t \geq T^a(b)$ for the winner and an assessment rule with bias $b \geq 0$ for the selection of the winner. We assume that this contract is publicly observed by contracting parties. A side contract specifies the effort level and the side payment for each agent. Absent favoritism, the agents are treated identical in the tournament as well as under collusion, thus the side payments should be the same regardless of who is the winner. When the principal introduces favoritism and makes it commonly known the identity of the favored agent (and thus the identity of the disfavored one), the side payments could be set differently according to the identity of the agent. Let $s_f \geq 0$ (resp. $s_d \geq 0$) denote the side payment from the winner to the loser when the winner is the favored agent (resp. disfavored agent).[24]

We solve for the equilibrium by backward induction. First of all, given the tournament contract, we analyze the incentive constraints of collusion. We then consider the design of the tournament contract which is immune to collusion.

Side Contract

Suppose the mediator proposes the agents to take the low efforts and moreover specifies the side payments s_f and s_d respectively for the favored and disfavored agents. First of all, the side contract must ensure that both types of agents are willing to participate collusion, that is, they must be strictly better-off under collusion. Recall that the favored agent can earn an expected payoff $G(b)t - c$ absent collusion, whereas he can receive an expected payoff $G(b)(t - s_f) + (1 - G(b))\, ks_d$ by joining the coalition, where $t - s_f$ is the net payment received as a winner and ks_d is the net payment earned as a loser (who then receives the payment from the winner, the disfavored agent, which is worth ks_d). Therefore, the participation constraint facing the favored agent requires

$$G(b)(t - s_f) + (1 - G(b))\, ks_d > G(b)t - c,$$

which amounts to (using the relation $G(b) = 1 - G(-b)$)

$$G(b)s_f - G(-b)ks_d < c. \quad (CIR_f) \tag{1}$$

That is, the gain of joining collusion, c, must be greater than the expected loss, $G(b)s_f - G(-b)ks_d$, which is the difference of the expected side payments between winning and losing. Similarly, the side payments must ensure a higher expected payoff for the disfavored agent to participate collusion, which requires that the following participation constraint be satisfied:

$$G(-b)s_d - G(b)ks_f < c. \quad (CIR_d)$$

Moreover, to induce low efforts in equilibrium, the mediator must mitigate the agents' incentives for unilateral deviation to the high effort. For the favored one, such deviation increases the probability of winning from $G(b)$ to $G(h+b)$ and yields an expected gain $(G(h + b) - G(b))\,(t - s_f - ks_d)$ (as now the "virtual" incentive power, the payoff gap between winning and losing, is reduced to $t - s_f - ks_d$ due to side payments), thus the favored agent is discouraged from deviating

[24] The side payments cannot be negative since the loser of the tournament receives no payment.

unilaterally to the high effort if this expected gain is less than the saving of effort cost c under collusion, i.e.,

$$(G(h+b)-G(b))\,(t-s_f-ks_d)<c.$$

This requires that the gap of payoffs between the states of winning (i.e., $t-s_f$) and losing (i.e., ks_d) be lower enough such that

$$t-s_f-ks_d<\frac{c}{G(h+b)-G(b)}=T_d^a(b). \quad (CIC_f)$$

Thus, the side contract provides exactly the "countervailing incentives" by reducing the payoff gap under collusion (from t to $t-s_f-ks_d$), which mitigates the incentives for taking the high effort. Recall that $G(h+b)-G(b)=G(-b)-G(-h-b)$, that is, when the favored agent deviates from the low effort to the high effort unilaterally, the increase of his winning probability is equal to the increase of the losing probability of the disfavored one, and the latter is equal to the decrease of the winning probability as if the disfavored one deviates from the high effort to the low effort unilaterally.

By analogy, deviating from the collusive agreement unilaterally (i.e., taking instead the high effort) yields the disfavored agent an extra payoff equal to $(G(h-b)-G(-b))\,(t-s_d-ks_f)$, as his probability of winning is increased from $G(-b)$ to $G(h-b)$, and the disfavored one is discouraged from deviation if this extra gain is strictly less than the cost of effort c. This implies the following incentive compatibility constraint for the disfavored agent

$$t-s_d-ks_f<\frac{c}{G(h-b)-G(-b)}=T_f^a(b). \quad (CIC_d)$$

Hence, to sustain collusion, the side contract must satisfy the two participation constraints (CIR_f) and (CIR_d), and the two incentive compatibility constraints (CIC_f) and (CIC_d). The set of all incentive feasible side payments (s_f,s_d) that satisfy the four constraints is denoted by $\Gamma\left(\underline{e}\right)$, where $\underline{e}$ stands for the pair of low efforts, and is depicted as the shaded region in Figure 2 (where we use locus such as CIR_f standing for the binding constraint of (CIR_f), with slightly abuse of notation).

Insert Figure 2 here.

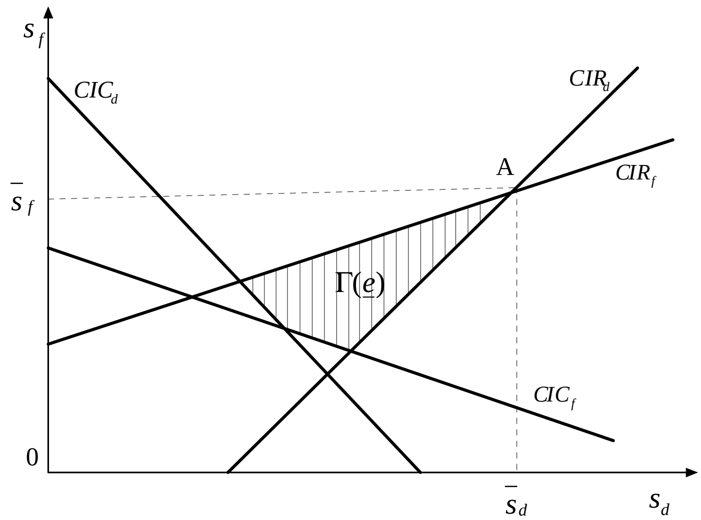

Figure 2

Preventing Collusion

We now analyze how the set $\Gamma(\underline{e})$ is affected by the incentive prize t. Notice that increasing t has no impact on the participation constraints (CIR_f) and (CIR_d), but tightens the constraints (CIC_f) and (CIC_d) as both lines move upwards to the direction of north-east in Figure 2. Thus, when one of the lines representing the constraints (CIC_f) and (CIC_d) goes through the intersection point of the loci for the constraints (CIR_f) and (CIR_d) (point A in Figure 2), the set $\Gamma(\underline{e})$ turns to be empty, in which case no side payments can satisfy the four constraints. Thus collusion on low efforts is not sustainable when the incentive prize t is sufficiently large.

Hence, there exists a lowerbound of the incentive prize t such that the set $\Gamma(\underline{e})$ is empty if and only if t exceeds this bound, and the lowerbound can be easily derived by solving the incentive constraints. Rearranging (CIR_d) yields

$$s_d < \frac{G(b)}{G(-b)} k s_f + \frac{c}{G(-b)},$$

which requires that the side payment of s_d be bounded above in order to keep the disfavored agent in collusion. Rearranging the constraint (CIR_f) in a similar way and moreover substituting the above relation into (CIR_f), we obtain

$$\begin{aligned} s_f &< \frac{G(-b)}{G(b)} k s_d + \frac{c}{G(b)} \\ &< \frac{G(-b)}{G(b)} k \left(\frac{G(b)}{G(-b)} k s_f + \frac{c}{G(-b)} \right) + \frac{c}{G(b)} \\ &= k^2 s_f + \frac{(1+k)\, c}{G(b)}, \end{aligned}$$

which further implies

$$(1-k)\, G(b) s_f < c,$$

or equivalently

$$s_f < \bar{s}_f \equiv \frac{c}{G(b)\,(1-k)}. \tag{2}$$

Recall that $(1-k)\, G(b) s_f$ is the expected deadweight loss of the collusive stakes due to the inefficiency of side transfer from the favored agent, thus a necessary condition for sustaining collusion is that such deadweight loss must be lower than the gain of collusion, i.e., the saving of effort cost c.

By analogy, the same restriction on the side payment applies to participation constraint of the disfavored agent, which requires

$$(1-k)\, G(-b) s_d < c,$$

or equivalently

$$s_d < \bar{s}_d \equiv \frac{c}{G(-b)\,(1-k)}. \tag{3}$$

The upper bounds $\bar{s}_f$ and $\bar{s}_d$ can be also derived from the condition where the loci of the binding constraints (CIR_f) and (CIR_d) coincide, which is depicted as point A in Figure 2.

On the other hand, raising the prize t provides stronger incentives for the agent to take the high effort unilaterally, as we have noted above, which tightens the incentive compatibility constraints (CIC_f) and (CIC_d). Using the two upper bounds for the side payments and rearranging the constraints, we obtain from (CIC_f)

$$\begin{aligned} t &< s_f + k s_d + T_d^a(b) < \bar{s}_f + k \bar{s}_d + T_d^a(b) \\ &= \frac{c\,(kG(b) + G(-b))}{G(-b) G(b)\,(1-k)} + T_d^a(b) \equiv T_f^c(b;k), \end{aligned}$$

and from (CIC_d)

$$\begin{aligned} t &< s_d + k s_f + T_f^a(b) < \bar{s}_d + k \bar{s}_f + T_f^a(b) \\ &= \frac{c\,(G(b) + kG(-b))}{G(-b) G(b)\,(1-k)} + T_f^a(b) \equiv T_d^c(b;k). \end{aligned}$$

That is, the favored (resp. disfavored) agent has no incentives to take the high effort unilaterally only if the incentive prize is lower than $T_f^c(b;k)$ (resp. $T_d^c(b;k)$).

Summarizing the above analysis, we can conclude that set of the incentive feasible side payments $\Gamma(\underline{e})$ is empty if

$$t \geq T^c(b;k) \equiv \min\{T_f^c(b;k), T_d^c(b;k)\},$$

which constitutes a sufficient condition for preventing collusion on low efforts. In Appendix C, we show that this condition is also necessary for collusion-proofness, thus collusion on low efforts can be prevented if and only if $t \geq T^c(b;k)$.

The intuition can be further demonstrated. Side payments from the winner to the loser reduce the payoff gap between winning and losing and thus suppresses the incentives for high efforts. On the other hand, side payments also entail the dead-weight loss which reduces the stake of collusion, and the loss is proportional to the amount of the side payments. Thus, agents are willing to collude only if the dead-weight loss is less than the gain of collusion, which gives the upper bounds for the side payments, $\bar{s}_f$, $\bar{s}_d$, as characterized by (2) and (3) respectively. This implies that there exist the minimum payoff gaps between the states of winning and losing, which are given by $t - \bar{s}_f - k\bar{s}_d$ for the favored agent and $t - \bar{s}_d - k\bar{s}_f$ for the disfavored one. In other words, collusion reduces the incentive power (i.e., the payoff gap) by up to $\bar{s}_f + k\bar{s}_d$ for the favored agent and $\bar{s}_d + k\bar{s}_f$ for the disfavored one.

To induce the favored agent to take the high effort in the presence of collusion, the principal must then offer extra incentives to restore the payoff gap such that the "virtual payoff gap", $t - \bar{s}_f - k\bar{s}_d$, provides sufficiently high incentives for the favored agent to deviate, i.e., $t - \bar{s}_f - k\bar{s}_d \geq T_d^a(b)$, which implies $t \geq T_f^c(b;k)$. Similarly, the disfavored agent will deviate to the high effort only if $t - \bar{s}_d - k\bar{s}_f \geq T_f^a(b)$, which implies $t \geq T_d^c(b;k)$. For collusion-proofness, it is sufficient to induce one agent to deviate unilaterally, therefore it suffices to offer $t \geq \min\{T_f^c(b;k), T_d^c(b;k)\} = T^c(b;k)$.

Remark: Cost of Collusion-Proofness. Preventing collusion is costly. We can decompose the threshold $T_f^c(b;k)$ as $T_f^c(b;k) = \Delta_f(b;k) + T_d^a(b)$, where

$$\Delta_f(b;k) \equiv \frac{c\,(kG(b) + G(-b))}{G(-b)G(b)\,(1-k)} = \bar{s}_f + k\bar{s}_d,$$

is the maximum reduction of incentive power for the favored agent that can be made under collusion, since the payoff gap between winning and losing is reduced from t to $t - s_f - ks_d$, with a magnitude of $s_f + ks_d$, and it is indeed the extra incentive cost paid for the favored agent to take the high effort as now the principal must ensure that the "virtual" incentives, $t - \Delta_f(b;k)$, exceed the threshold $T_d^a(b)$. Similarly, we can rewrite the threshold $T_d^c(b;k)$ as $T_d^c(b;k) = \Delta_d(b;k) + T_f^a(b)$, where $T_f^a(b)$ is the incentive cost that is paid for the disfavored agent to deviate and take the high effort unilaterally, while

$$\Delta_d(b;k) \equiv \frac{c\left(G(b) + kG(-b)\right)}{G(-b)G(b)\left(1-k\right)} = \bar{s}_d + k\bar{s}_f$$

is indeed the extra incentive cost incurred due to the reduction of incentive power by side payments.

Recall that, in the absence of collusion, it is sufficient to offer $t = T^a(b) = \max\{T_f^a(b), T_d^a(b)\} = T_d^a(b)$ for inducing the high efforts, whereas the principal must grant a prize at least equal to $T^c(b;k)$ to prevent collusion (and implement the high efforts). It is straightforward to see that $T_f^c(b;k) = \Delta_f(b;k) + T_d^a(b) > T_d^a(b)$, and we show also in Appendix D that $T_d^c(b;k) > T_d^a(b)$, thus $T^c(b;k) > T^a(b)$. **Q.E.D.**

The Properties of the thresholds

It is straightforward to see that, with no favoritism (i.e., $b = 0$), both agents face the same incentive constraint, that is, $T_f^c(0;k) = T_d^c(0;k)$. Favoritism then creates different incentive effects for the favored and disfavored agents to take the high efforts. To see the impacts on the incentive constraints, we need to characterize further the properties of the thresholds $T_f^c(b;k)$ and $T_d^c(b;k)$.

While the threshold $T_d^a(b)$ increases with b, the extra incentives $\Delta_f(b;k)$ may not be monotonic for general distribution function $G(\cdot)$, and thus $T_f^c(b;k)$ may not be monotonically increasing in b. To see this, differentiating $T_f^c(b;k)$ with respect to b, we obtain

$$\begin{aligned}\frac{\partial T_f^c(b;k)}{\partial b} &= \frac{\partial \Delta_f(b;k)}{\partial b} + \frac{\partial T_d^a(b)}{\partial b} \\ &= \frac{cg(b)}{(1-k)}\left(\frac{kG^2(b) - G^2(-b)}{G^2(-b)G^2(b)}\right) + \frac{g(b) - g(h+b)}{\left(G(h+b) - G(b)\right)^2}c.\end{aligned} \tag{4}$$

Notice that the second term, the derivative of $\frac{\partial T_d^a(b)}{\partial b}$, is always positive while the first term, which is the derivative of $\Delta_f(b;k)$, becomes positive when $kG^2(b) > G^2(-b)$, which amounts to

$$b > b^0 \equiv G^{-1}\left(\frac{1}{1+\sqrt{k}}\right).$$

Hence $T_f^c(b;k)$ must be increasing in b for all $b \geq b^0$. Meanwhile, differentiating $T_d^c(b;k)$ with respect to b yields

$$\begin{aligned}\frac{\partial T_d^c(b;k)}{\partial b} &= \frac{\partial \Delta_d(b;k)}{\partial b}+\frac{\partial T_f^a(b)}{\partial b} \\ &= \frac{cg(b)}{(1-k)}\left(\frac{1}{G^2(-b)}-\frac{k}{G^2(b)}\right)+\frac{g(b-h)-g(b)}{\left(G(b)-G(b-h)\right)^2}c.\end{aligned} \tag{5}$$

The first term in the above equation is the derivative of $\Delta_d(b;k)$, which is always positive; while the second term, which is the derivative of $T_f^a(b)$, is positive for $b \geq 0.5h$. Thus $T_d^c(b;k)$ must be increasing for all $b \geq 0.5h$. Summarizing the above analysis, there must exist some $\bar{b} \leq \max\{b^0, h/2\}$ such that $T^c(b;k) = \min\{T_f^c(b;k), T_d^c(b;k)\}$ increases in b for all $b \geq \bar{b}$.

By contrast, we can show that $T^c(b;k)$ decreases in b for sufficiently small b. Notice that the two thresholds are indeed "symmetric" in the sense that $T_f^c(b;k) = T_d^c(-b;k)$, thus, the thresholds $T_f^c(b;k)$ and $T_d^c(b;k)$ coincides at $b=0$: $T_f^c(0;k) = T_d^c(0;k)$. Moreover, since

$$\frac{\partial T_f^c(b;k)}{\partial b} = -\frac{\partial T_d^c(-b;k)}{\partial b},$$

it follows that

$$\frac{\partial T_f^c(0;k)}{\partial b} = -\frac{g(0)c}{G^2(0)}+\frac{\left(g(0)-g(h)\right)c}{\left(G(h)-G(0)\right)^2} = -\frac{\partial T_d^c(0;k)}{\partial b},$$

that is, these two thresholds move towards the opposite directions when some degree of favoritism is introduced (for b increasing slightly from 0). This ensures that the threshold $T^c(b;k) = \min\{T_f^c(b;k), T_d^c(b;k)\}$ decreases for sufficiently small b.

Summarizing the above analysis, we know that the threshold $T^c(b;k)$ decreases in b first and increases in b finally, as characterized in the following lemma:

Lemma 2 *There exist thresholds $\underline{b}$ and $\bar{b}$ satisfying $0 < \underline{b} \leq \bar{b}$ such that $T^c(b;k)$ decreases in $b \leq \underline{b}$ and increases in $b \geq \bar{b}$.*

Proof. See Appendix C. ■

We have shown that collusion on low efforts can be prevented when the principal offers sufficiently large incentive prize such that $t \geq T^c(b;k)$. However, the agents may seek to collude on other effort levels instead of low efforts for both of them. For instance, the agents may reach the collusive agreement which specifies that the favored agent takes the low effort while the disfavored one takes the high effort, or vise versa. To implement the high efforts, the principal must ensure that collusion on any kinds of effort levels other than the high efforts is not sustainable. Indeed, as we show in Appendix D, the condition $t \geq T^c(b;k)$ is sufficient for this purpose, that is, no collusion on effort levels other than the high efforts can be sustained when $t \geq T^c(b;k)$. On the other hand, since $T^c(b;k) > T^a(b)$, the incentive prize is sufficiently high to induce high efforts in the absence of collusion.

Since the principal's utility is strictly decreasing in t, the optimal tournament prize must satisfy

$$t^*(b;k) = T^c(b;k),$$

and the principal then chooses the degree of favoritism b to minimize $t^*(b;k)$; denote by b^* the solution of $\min_b T^c(b;k)$.

Lemma 2 indicates that $T^c(b;k)$ decreases with b for $b \leq \underline{b}$ and increases with b for $b \geq \bar{b}$, it thus follows that there must exist some $b^* \in [\underline{b}, \bar{b}]$ such that b^* minimizes $T^c(b;k)$. Summarizing the analysis above leads to the following main result:

Proposition 2 *Collusion between agents can be deterred and the high efforts can be induced whenever the principal offers the sufficiently high incentive prize such that $t \geq T^c(b;k)$. It is desirable to introduce some degree of favoritism (i.e., $b > 0$) when agents are able to collude, but excessive favoritism is not desirable; the optimal degree of favoritism must be between $\underline{b}$ and $\bar{b}$.*

Proof. See Appendix D. ■

Proposition 2 validates two facts. First of all, it validates that, when agents are able to collude, introducing some degree of favoritism reduces the incentive cost for implementing the high efforts and makes the principal strictly better off. This is because that favoritism creates different impacts on the incentive constraints for the favored and disfavored agents and thus differentiates their incentive constraints.

To illustrate the intuition, suppose that $\partial T_d^c(0;k)/\partial b > 0$,[25] and consider the "marginal" case where the side payments are such that $s_f = \bar{s}_f$ and $s_d = \bar{s}_d$, and thus $s_d + ks_f = \bar{s}_d + k\bar{s}_f = \Delta_d(b;k)$. We start from the case with slight favoritism such that b is very close to zero, and suppose the principal offers $t = T_d^c(b;k)$ such that $(G(b) - G(b-h))(T_d^c(b;k) - \bar{s}_d - k\bar{s}_f) = c$, which makes the disfavored agent indifferent between respecting the collusive agreement (i.e., taking the low effort) and deviating unilaterally (i.e. taking the high effort). Since the side payments have reached the upper bounds, the mediator cannot increase further the side payments such that the agents strictly prefer collusion. Consider now that the principal increases b slightly to b'. This causes two opposite effects. First, since $\Delta_d(b;k) = \bar{s}_d + k\bar{s}_f$ increases in

[25] This holds when $\frac{g(0)}{G^2(0)} > \frac{g(0)-g(h)}{(G(h)-G(0))^2}$. The some logic applies to the case with $\partial T_f^c(0;k)/\partial b > 0$, in which the relevant threshold for preventing collusion is $T_f^c(b;k)$.

b, this reduces the the incentive power and makes the disfavored agent better off under collusion. Second, the difference of the winning probabilities, $G(b) - G(b-h)$, also increases in b for $b < h/2$, which provides higher incentives for deviation. However, as we show in Appendix C, the first effect dominates the second one and thus the overall effect of such slight increase in b results to $(G(b') - G(b'-h))(T_d^c(b;k) - \Delta_d(b';k)) < c$, that is, the expected gain from deviation is strictly less than the benefit from collusion, and thus the disfavored agent strictly prefers collusion. Therefore providing higher degree of favoritism suppresses the disfavored agent's incentives to deviate from collusion, and this in turn calls for higher incentive prize of preventing collusion $T_d^c(b';k) > T_d^c(b;k)$.

By contrast, increasing b causes different net effect on the threshold $T_f^c(b;k)$. Suppose the principal offers $t = T_f^c(b;k)$ such that $(G(h+b) - G(b))\left(T_f^c(b;k) - \bar{s}_f - k\bar{s}_d\right) = c$, which makes the favored agent indifferent between respecting the collusive agreement and deviating unilaterally. Increasing b slightly to b' reduces $\Delta_f(b;k) = \bar{s}_f + k\bar{s}_d$ as well as the gap of winning probabilities $G(h+b) - G(b)$, however the first effect dominates the second one (see Appendix C) for sufficiently small b and the overall effect leads to $(G(h+b') - G(b'))\left(T_f^c(b;k) - \Delta_f(b';k)\right) > c$. Thus, granting more favoritism to the favored agent provides higher incentives to deviate from collusion, which in turn calls for lower threshold of deterring collusion: $T_f^c(b';k) < T_f^c(b;k)$.

To prevent collusion, it is sufficient to attract one agent (say the favored one) to deviate unilaterally, thus the threshold of deterring collusion, $T^c(b;k)$, is the minimum of the two thresholds $T_f^c(b;k)$ and $T_d^c(b;k)$. Since both thresholds coincide at $b = 0$, and $T_d^c(b;k)$ increases in b but $T_f^c(b;k)$ decreases in b for sufficiently small b, it follows that $T^c(b;k) = T_f^c(b;k)$, which decreases in b. Thus, giving some degree of favoritism makes the principal strictly better off.

Secondly, the proposition also validates that granting excessive favoritism indeed increases the incentive cost and makes the principal strictly worse off. The optimal degree of favoritism must be bounded above by $\bar{b}$, which is less than $\max\{b^0, \frac{h}{2}\}$ as shown in Appendix C. The intuition is quite simple. Consider the "marginal" case where the side payments are such that $s_f = \bar{s}_f$ and $s_d = \bar{s}_d$, and thus $s_f + ks_d = \bar{s}_f + k\bar{s}_d = \Delta_f(b;k)$. Suppose the principal grants excessive high favoritism with $b > \bar{b}$, and offers $t = T_f^c(b;k)$ such that $(G(h+b) - G(b))\left(T_f^c(b;k) - \bar{s}_f - k\bar{s}_d\right) = c$. Decreasing b slightly to b' increases the gap of the winning probabilities, $G(h+b) - G(b)$ and moreover decreases $\Delta_f(b;k)$ ($\Delta_f(b;k)$ increases for $b > \bar{b}$), which leads to

$$\left(G\left(h+b'\right) - G\left(b'\right)\right)\left(T_f^c(b;k) - \Delta_f(b';k)\right) > c.$$

By analogy, suppose the principal offers $t = T_d^c(b;k)$ such that $(G(b) - G(b-h))(T_d^c(b;k) - \bar{s}_d - k\bar{s}_f) =$

c. Then decreasing b slightly to b' increases the gap of the winning probabilities $G(b) - G(b-h)$ ($G(b) - G(b-h)$ decreases in b for $b > \bar{b} > h/2$) and moreover decreases $\Delta_d(b;k)$ (which increases in b), which results to

$$\left(G\left(b'\right) - G\left(b'-h\right)\right)\left(T_d^c(b;k) - \Delta_d(b';k)\right) > c.$$

This shows that, under excessive favoritism, decreasing b slightly provides higher incentives for both agents to deviate, which reduces the incentive cost for preventing collusion.

5 Optimal Favoritism

Proposition 2 indicates that the optimal degree of favoritism, b^*, which minimizes $T^c(b;k)$, must be greater than $\underline{b}$ and lower than $\bar{b}$. Since $T^c(b;k)$ is continuous in b, such optimum always exists. While it is very difficult to solve the optimum explicitly for general distribution function, the following analysis shows that, under some plausible conditions, the thresholds $T_d^c(b;k)$ and $T_f^c(b;k)$ display the property of convexity, which contributes to characterizing the optimal degree of favoritism.

Recall that

$$\frac{\partial T_f^c(b;k)}{\partial b} = \frac{\partial \Delta_f(b;k)}{\partial b} + \frac{\partial T_d^a(b)}{\partial b},$$

then, differentiating both sides with respect to b, we obtain

$$\frac{\partial^2 T_f^c(b;k)}{\partial b^2} = \frac{\partial^2 \Delta_f(b;k)}{\partial b^2} + \frac{\partial^2 T_d^a(b)}{\partial b^2}.$$

It is easy to check that $\frac{\partial^2 \Delta_f(b;k)}{\partial b^2} > 0$ for all $b \leq b^0$, and moreover $\frac{\partial^2 T_d^a(b)}{\partial b^2} > 0$ if $g(\cdot)$ is weakly concave. Thus $T_f^c(b;k)$ is convex for $b \leq b^0$ under the condition of the weak concavity of the density function.

Similarly,

$$\frac{\partial^2 T_d^c(b;k)}{\partial b^2} = \frac{\partial^2 \Delta_d(b;k)}{\partial b^2} + \frac{\partial^2 T_f^a(b)}{\partial b^2},$$

and we can show that $\frac{\partial^2 T_f^a(b)}{\partial b^2} > 0$ when $b < h/2$ and moreover $\frac{\partial^2 \Delta_d(b;k)}{\partial b^2} > 0$ if $g(\cdot)$ is weakly concave and also satisfies

$$g'(\frac{h}{2})G(-\frac{h}{2}) + 2g^2\left(\frac{h}{2}\right) \geq 0, \tag{6}$$

which requires that the slope of the density function is bounded at $b = h/2$. Therefore $T_d^c(b;k)$ is convex for $b \leq h/2$ under these two conditions.

The above analysis is summarized in the following lemma:

Lemma 3 *Suppose the density function $g(\cdot)$ is weakly concave, then the threshold $T_f^c(b;k)$ is convex for $b \leq b^0$; moreover, the threshold $T_d^c(b;k)$ is convex for $b \leq h/2$ if the condition* (6) *holds.*

Proof. See Appendix E. ■

Thanks to this lemma, we can now characterize the optimal degree of favoritism. Notice that the two thresholds $T_d^c(b;k)$ and $T_f^c(b;k)$ move in the opposite directions as b increases from 0. Assuming the weak concavity of the density function and moreover the condition (6) holds, we consider two cases.

Case A: Suppose

$$\frac{g(0)}{G^2(0)} \geq \frac{g(0)-g(h)}{(G(h)-G(0))^2}, \tag{7}$$

which implies that $\frac{\partial T_d^c(0;k)}{\partial b} \geq 0$ and $\frac{\partial T_f^c(0;k)}{\partial b} \leq 0$. In this case, starting from 0, $T_d^c(b;k)$ increases in b for $b \leq h/2$ by lemma 3, and keeps increasing for $b > h/2$ (see the analysis before lemma 1). Since $T_d^c(0;k) = T_f^c(0;k)$ and moreover $T_d^c(b;k)$ increases for all b, the optimal b must minimize $T_f^c(b;k)$. Suppose there exists some $b_f > 0$ minimizes $T^c(b;k) = T_d^c(b;k)$, which implies $T_d^c(b_f;k) \leq T_f^c(b_f;k)$, but then decreasing b_f by ε would reduce $T^c(b;k)$ as $T_d^c(b;k)$ increases in b, thus b_f is not the optimum. Since $T_f^c(b;k)$ decreases from 0 and then increases, and since $T_f^c(b;k)$ is convex for $0 < b \leq b^0$, it follows that the optimal degree of favoritism b^* must satisfy

$$\frac{\partial T_f^c(b^*;k)}{\partial b} = 0,$$

and this condition is sufficient and necessary. This condition further implies that

$$-\frac{\partial \Delta_f(b;k)}{\partial b} = \frac{\partial T_d^a(b)}{\partial b}. \tag{8}$$

To highlight the intuition behind, notice that $\Delta_f(b;k) = \bar{s}_f + k\bar{s}_d$ is the mitigation of incentive power for the favored agent that can be secured under collusion, which measures the extra cost of preventing collusion, while $T_d^a(b)$ is the incentive cost for inducing the favored agent to deviate from the low effort to the high effort unilaterally absent side payments. Starting from the case of no favoritism, increasing b engenders two opposite effects. It reduces $\Delta_f(b;k)$, but increases the threshold $T_d^a(b)$ by reducing the the gap of winning probabilities $G(h+b) - G(b)$, and the first effect dominates the second one when b is quite small. However, the second effect is enhanced when b keeps increasing, which offsets exactly the first effect at $b = b^*$, as indicated by equation (8). The analysis is also demonstrated by Figure 3.

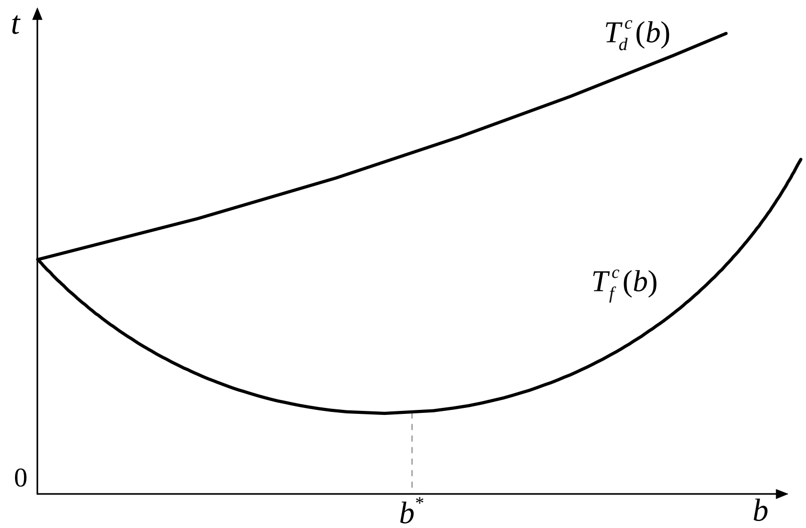

Figure 3

Case B: Suppose instead

$$\frac{g(0)}{G^2(0)} < \frac{g(0) - g(h)}{(G(h) - G(0))^2},$$

which implies $\frac{\partial T_f^c(0;k)}{\partial b} > 0$ and $\frac{\partial T_d^c(0;k)}{\partial b} < 0$. Then $T_f^c(b;k)$ increases from $b = 0$ and keeps increasing for $b \leq b^0$ by lemma 3, and for $b > b^0$ by lemma 2. In this case, the optimal b must minimize $T_d^c(b;k)$. As $T_d^c(b;k)$ decreases first and then increases, and moreover since $T_d^c(b;k)$ is convex for $0 < b \leq h/2$, the optimal degree of favoritism b^* must satisfy

$$\frac{\partial T_d^c(b^*;k)}{\partial b} = 0,$$

and this condition is sufficient and necessary. This condition further implies that

$$-\frac{\partial \Delta_d(b;k)}{\partial b} = \frac{\partial T_f^a(b)}{\partial b},$$

that is, the increasing effect of $\Delta_d(b;k)$ is exactly offset by the decreasing effect of $T_f^a(b)$.

Summarizing the above analysis leads to:

Proposition 3 *Suppose the density function $g(\cdot)$ is weakly concave and moreover the condition (6) holds, then the optimal favoritism minimizes $T_f^c(b;k)$ (resp.$T_d^c(b;k)$) and is determined by the first-order condition if $T_f^c(b;k)$ (resp. $T_d^c(b;k)$) decreases from beginning.*

Example: Uniform Distribution

A simple example is useful for demonstration. Notice that, the results of Proposition 2 hold for any distribution function that satisfy the assumption in section 2, including the simple case of uniform distribution. Suppose now the random shock ε is uniformly distributed in the region $[-l, l]$ with the density function $f(x) = \frac{1}{2l}$ for $x \in [-l, l]$ and $f(x) = 0$ otherwise, which implies $F(x) = (x + l)/2l$. Then

$$G(x) = \int_{-l}^{l} F(x+\varepsilon) f(\varepsilon)\, d\varepsilon = \frac{x+l}{2l},$$

and

$$g(x) = \frac{1}{2l}.$$

In this case, the incentive costs for inducing the high efforts in the absence of collusion are

$$T_f^a(b) = T_d^a(b) = \frac{2lc}{h}.$$

When agents are able to collude, the incentive cost for encouraging the favored agent to deviate from collusion is[26]

$$T_f^c(b;k) = \frac{c\,(kG(b) + G(-b))}{G(-b)G(b)\,(1-k)} + T_d^a(b) = \frac{2lc\,(k\,(l+b) + (l-b))}{(l+b)(l-b)\,(1-k)} + \frac{2lc}{h},$$

and that for inducing the disfavored agent to deviate is

$$T_d^c(b;k) = \frac{c\,(G(b) + kG(-b))}{G(-b)G(b)\,(1-k)} + T_l^a(b) = \frac{2lc\,(k\,(l-b) + (l+b))}{(l+b)(l-b)\,(1-k)} + \frac{2lc}{h}.$$

Since

$$\begin{aligned} T_d^c(b;k) - T_f^c(b;k) &= \frac{c\,(G(b) - G(-b))}{G(-b)G(b)} + T_f^a(b) - T_d^a(b) \\ &= \frac{2bc}{(l+b)(l-b)} > 0, \end{aligned}$$

it is always less costly to induce the favored agent to deviating from collusion, and thus $T^c(b;k) = T_f^c(b;k)$.

Moreover, the threshold $T_d^c(b;k)$ is strictly increasing in b

$$\begin{aligned} \frac{\partial T_d^c(b;k)}{\partial b} &= \frac{cg(b)}{(1-k)}\left(\frac{1}{G^2(-b)} - \frac{k}{G^2(b)}\right) + \frac{g(b-h) - g(b)}{(G(b) - G(b-h))^2}c \\ &= \frac{2lc}{(1-k)}\left(\frac{1}{(l-b)^2} - \frac{k}{(l+b)^2}\right) > 0, \end{aligned}$$

[26] We focus on the case with $b < l$.

which implies that the disfavored agent has stronger incentives to collude under favoritism than absent favoritism. On the other hand, differentiating $T_f^c(b;k)$ with respect to b, we obtain

$$\begin{aligned}\frac{\partial T_f^c(b;k)}{\partial b} &= \frac{c}{(1-k)}\left(\frac{kg(-b)}{G^2(-b)} - \frac{g(b)}{G^2(b)}\right) - \frac{(g(h+b)-g(b))\,c}{(G(h+b)-G(b))^2} \\ &= \frac{2lc}{(1-k)}\left(\frac{k(l+b)^2-(l-b)^2}{(l-b)^2(l+b)^2}\right),\end{aligned}$$

thus $\frac{\partial T_f^c(b;k)}{\partial b} \geq 0$ if and only if

$$\Upsilon(b;k) = k(l+b)^2 - (l-b)^2 \geq 0.$$

Notice that, $\Upsilon(0;k) < 0$, $\Upsilon(l;k) > 0$, and $\Upsilon'(b;k) = 2k(l+b) + 2(l-b) > 0$ for $b < l$, thus there exists a unique $b^*(k) \in (0,l)$ such that $\Upsilon(b^*;k) = 0$ and $\Upsilon(b;k) > 0$ if and only if $b > b^*$. Therefore, b^* minimizes $T_f^c(b;k) = T^c(b;k)$, which is given by

$$b^*(k) = \frac{l\left(1-\sqrt{k}\right)}{1+\sqrt{k}}.$$

6 Conclusion

Favoritism prevails in organizations that rely on subjective assessments of employee performance, and its harmful impact on the efficiency is widely recognized. This paper shows that favoritism could benefit the employer when collusion among employees becomes a serious threat in organizations. Favoritism differentiates the incentive constraints for the agents, and adequate favoritism reduces the cost for preventing collusion but excessive favoritism increases the incentive cost.

We explore the main insights in a stylized setting of tournament with one principal and two homogeneous agents. To highlight the efficiency-enhancing effect of playing favoritism, we assume away the altruistic motivation for favoritism. We use the discrete choice model for the efforts, and moreover focus on the incentive issues of implementing the socially efficient (high) efforts. We have assumed in the basic model that the agents could choose only two effort levels, namely, high and low, but the analysis can be easily extended to the case of more than two effort levels, subject the principal needs to offer sufficiently high incentive prize to discourage the agents deviating to effort levels other than the highest one.

We have also assumed away the difference of productivity (or talents) between agents, thus the differentiation of incentive constraints steps only from the bias of subjective assessment.

Incorporating the heterogeneity of talents in modelling the agent productivity does not affect the basic insights if an agent's talent is at least known by himself and substitutes the effort in the production function (i.e., the production function takes the form of $y = e + \alpha + \varepsilon$, where α is the agent's talent). In this case, the total productivity, which is the sum of the effort and talent, determines the the expected output, and simply by replacing the efforts with the total productivity the analysis goes through.

Extending the basic model to the case with more than two agents involves more complex model like Green and Stocky (1983), and the probability of winning, which depends on the relative efforts among agents, cannot be characterized explicitly as the function of the effort difference in general. However, we believe that the basic insights that favoritism differentiates the incentive constraints between the favored agent and the disfavored agent and that increasing bias in subjective assessments engenders opposite effects of incentives for different types of agents would be still validated in the setting with more than two agents, which is left to the future research agenda.

Of course, there might exist other non-altruistic motivations for playing favoritism in organizations. When the employer (supervisor) is not the residual claimant of the organization, he may trade favoritism for the bribe from the employee, and may also favor some agent in promotion to secure his private benefit in the future. When instead the employer is the residual claimant, favoritism leads to different incentive effects on different types of agents, and how this differentiation of incentive constraints affects the incentives for taking the high efforts is not well explored in the literature. Finally, while we restrict the analysis of favoritism in a static model, in reality favoritism often exists in a long time, thus the study of the dynamic effect of favoritism is important to disclose the more profound insights.

Appendix A: Proof of Lemma 1

If $t \geq T^a(b)$, then both types of agents are discouraged from unilateral deviation to the low efforts, thus taking high efforts forms a Nash equilibrium. By contrast, if $t < T^a(b)$, say $t < T_d^a(b)$, then the disfavored agent can benefit from unilateral deviation to the low effort.

We show now taking low efforts for both agents cannot form a Nash equilibrium when $t \geq T^a(b)$. Suppose both agents take the low efforts, in which case the favored agent can earn an expected payoff equal to $G(b)t$. Whereas, by deviating to the high effort unilaterally, the favored agent can earn an extra payoff $(G(b+h) - G(b))\,t$, which overweighs the cost of effort c, since

$$\begin{aligned}(G(b+h) - G(b))\,t &\geq (G(b+h) - G(b))\,T^a(b) \\ &\geq (G(b+h) - G(b))\,T_d^a(b) = c.\end{aligned}$$

Thus, taking low efforts for both agents cannot be a Nash equilibrium.

Finally, we show that there does not exist any Nash equilibrium that involves one agent taking the high effort and another taking the low effort. To see this, consider a candidate equilibrium where the favored agent takes the high effort while the disfavored one takes the low effort, in which case the disfavored agent earns an expected payoff $G(-b-h)t$. However, by deviating to the high effort unilaterally, the disfavored agent can obtain the extra payoff $(G(-b) - G(-b-h))\,t = (G(b+h) - G(b))\,t$, which exceeds the effort cost c. By analogy, if instead the favored agent takes the low effort while the disfavored agent takes the high effort in the candidate equilibrium, then the favored one can benefit from deviating to the high effort unilaterally. It follows that taking asymmetric efforts cannot form a Nash equilibrium when $t \geq T^a(b)$. **Q.E.D.**

Appendix B: Proof of Proposition 1

Recall that

$$\frac{dT_d^a(b)}{db} = \frac{g(b) - g(h+b)}{(G(h+b) - G(b))^2} \geq 0,$$

thus the incentive cost for the disfavored agent, $T_d^a(b)$, increases with the degree of favoritism b. For the favored agent, notice that

$$\frac{dT_f^a(b)}{db} = \frac{g(b-h) - g(b)}{(G(b) - G(b-h))^2} = \frac{g(h-b) - g(b)}{(G(b) - G(b-h))^2}.$$

Suppose $b < h/2$, then $b \leq h - b$ and thus $g(b) \geq g(h-b)$. If instead $h/2 < b \leq h$, then $b > h - b \geq 0$ and thus $g(b) \leq g(h-b)$. Finally, if $b > h$, then $b > b - h > 0$ and thus $g(b) \leq g(b-h)$. Therefore, $g(b) > g(h-b)$ if and only if $b < h/2$, that is, the incentive cost for the favored agent decreases in b for $b < h/2$ and then increases in b for $b > h/2$.

We show now $T_d^a(b) \geq T_f^a(b)$. Consider two cases:

Case A: Suppose $b \leq h/2$, then $T_f^a(b)$ decreases with b, thus $T_d^a(b) \geq T_d^a(0) = T_f^a(0) \geq T_f^a(b)$.

Case B: Suppose instead $b > h/2$. If $h/2 < b \leq h$, then $b > h - b \geq 0$; if $b > h$, then $b > b - h > 0$. In both cases we have $g(b-h) = g(h-b) \geq g(b) \geq g(b+h)$, thus

$$\begin{aligned} & (G(b) - G(b-h)) - (G(h+b) - G(b)) \\ = \; & \int_{b-h}^{b} g(d)dd - \int_{b}^{b+h} g(d)dd \\ \geq \; & g(b)h - \int_{b}^{b+h} g(d)dd \\ \geq \; & g(b)h - g(b)h = 0. \end{aligned}$$

It follows that $G(b) - G(b-h) \geq G(h+b) - G(b)$, which implies $T_d^a(b) \geq T_f^a(b)$.

If the density $g(\cdot)$ is constant everywhere, which represents the case of uniform distribution, then we have $T_d^a(b) = T_f^a(b) = T^a(b)$, and $\frac{dT_d^a(b)}{db} = \frac{dT_f^a(b)}{db} = 0$. Therefore the incentive cost is independent of the degree of favoritism and introducing favoritism does not reduce the incentive cost.

Whenever the density $g(\cdot)$ is strictly decreasing for some b close to zero, there must exist a neighborhood $[0, \varepsilon]$ such that $g(\cdot)$ is strictly decreasing in this neighborhood. Then $G(b) - G(b-h) > G(h+b) - G(b)$ for some small $b \in [0, \varepsilon]$, and moreover $T_d^a(b)$ is strictly increasing in this neighborhood, thus $T^a(b) = T_d^a(b) > T_d^a(0)$. It follows that introducing favoritism makes the principal strictly worse off. **Q.E.D.**

Appendix C: Proof of Lemma 2

We show first that the sufficient and necessary condition to prevent collusion on low efforts is $t \geq T^c(b;k)$. Suppose $t \geq T^c(b;k) \geq T_f^c(b;k)$, which implies that

$$t \geq \frac{c\,(kG(b) + G(-b))}{G(-b)G(b)\,(1-k)} + T_d^a(b),$$

then there are no incentive feasible side payments s_f and s_d that satisfy the four constraints. Suppose there exists a pair of (s_f, s_d) satisfy the participation constraints (CIR_f) and (CIR_d), then it must be such that

$$s_f < \frac{c}{G(b)\,(1-k)}$$

and

$$s_d < \frac{c}{G(-b)\,(1-k)}.$$

Substituting these relations into the left hand side of the constraint (CIC_f), it must be that

$$\begin{aligned}
& t - s_f - ks_d \\
\geq\; & \frac{c\,(kG(b)+G(-b))}{G(-b)G(b)\,(1-k)} + T_d^a(b) - s_f - ks_d \\
>\; & \frac{c\,(kG(b)+G(-b))}{G(-b)G(b)\,(1-k)} + T_d^a(b) - \frac{c}{G(b)\,(1-k)} - \frac{ck}{G(-b)\,(1-k)} \\
=\; & T_d^a(b),
\end{aligned}$$

thus the incentive compatibility constraint for the favored agent is violated.

On the other hand, if $t = T^c(b;k) - \varepsilon < T^c(b;k)$, then there exists a pair of side payments (s_f, s_d) that satisfy the four constraints. To see this, suppose $T^c(b;k) = T_f^c(b;k) \leq T_d^c(b;k)$ and thus $t = T_f^c(b;k) - \varepsilon$. Let

$$\begin{aligned}
s_f &= \frac{c}{G(b)\,(1-k)} - \frac{\varepsilon}{2} \\
s_d &= \frac{c}{G(-b)\,(1-k)} - \frac{\varepsilon}{2}.
\end{aligned}$$

Then it is straightforward to see that the constraints (CIR_f) and (CIR_d) hold. Moreover, substituting them into the constraints (CIC_f) and (CIC_d), we obtain

$$\begin{aligned}
t - s_f - ks_d &= T_f^c(b;k) - \varepsilon - \frac{c}{G(b)\,(1-k)} - \frac{kc}{G(-b)\,(1-k)} + \frac{(1+k)\varepsilon}{2} \\
&= T_d^a(b) - \frac{(1-k)\,\varepsilon}{2} < T_d^a(b),
\end{aligned}$$

and

$$\begin{aligned}
t - s_d - ks_f &\leq T_d^c(b;k) - \varepsilon - \frac{kc}{G(b)\,(1-k)} - \frac{c}{G(-b)\,(1-k)} + \frac{(1+k)\varepsilon}{2} \\
&= T_f^a(b) - \frac{(1-k)\,\varepsilon}{2} < T_f^a(b),
\end{aligned}$$

thus the constraints (CIC_f) and (CIC_d) are also satisfied.

We now characterize the properties of $T^c(b;k)$. Notice that $T_f^c(b;k) = \Delta_f(b;k) + T_d^a(b)$, where

$$\Delta_f(b;k) \equiv \frac{c\,(kG(b)+G(-b))}{G(-b)G(b)\,(1-k)}$$

is the extra incentive that the principal must pay for deterring collusion. Differentiating $T_f^c(b;k)$ with respect to b, we obtain

$$\begin{aligned}\frac{\partial T_f^c(b;k)}{\partial b} &= \frac{\partial \Delta_f(b;k)}{\partial b} + \frac{\partial T_d^a(b)}{\partial b} \\ &= \frac{cg(b)}{(1-k)}\left(\frac{kG^2(b)-G^2(-b)}{G^2(-b)G^2(b)}\right) + \frac{g(b)-g(h+b)}{\left(G(h+b)-G(b)\right)^2}c,\end{aligned}$$

where

$$\frac{\partial \Delta_f(b;k)}{\partial b} = \frac{cg(b)}{(1-k)}\left(\frac{kG^2(b)-G^2(-b)}{G^2(-b)G^2(b)}\right)$$

is positive if and only if $\Phi(b;k) \equiv kG^2(b) - G^2(-b) > 0$. Notice that, $\Phi(0;k) = (k-1)\,G^2(0) < 0$ at $b=0$, $\Phi(\infty;k) = k > 0$ for b tends to infinity, and $\Phi'(b;k) = 2kG(b)g\,(b) + 2G(-b)g(-b) > 0$. It follows that there must exist a unique b^0 satisfying $\Phi(b^0;k) = 0$ and such that $\Phi(b;k) > 0$ if and only if $b > b^0$, that is,

$$b^0 = G^{-1}\left(\frac{1}{1+\sqrt{k}}\right),$$

where $G^{-1}(\cdot)$ is the inverse function of $G(\cdot)$. Thus, $\frac{\partial \Delta_f(b;k)}{\partial b} > 0$ if and only if $b > b^0$. This property, together with the fact that $\frac{\partial T_d^a(b)}{\partial b} \geq 0$, imply that $\frac{\partial T_f^c(b;k)}{\partial b} > 0$ for all $b > b^0$. To be more precise, let $\Sigma \equiv \{0 \leq b \leq b^0 | \frac{\partial T_f^c(b;k)}{\partial b} = 0\}$ be the set of all extreme points of $T_f^c(b;k)$. If Σ is not empty, then denote by $\bar{b}_f \equiv \sup \Sigma$ its upper bound, otherwise, let $\bar{b}_f = 0$. Thus, $\frac{\partial T_f^c(b;k)}{\partial b} > 0$ for all $b > \bar{b}_f$.

Similarly, note that $T_d^c(b;k) = \Delta_d(b;k) + T_f^a\,(b)$, where

$$\Delta_d(b;k) \equiv \frac{c\left(G(b)+kG(-b)\right)}{G(-b)G(b)\left(1-k\right)}.$$

Differentiating $T_d^c(b;k)$ with respect to b yields

$$\begin{aligned}\frac{\partial T_d^c(b;k)}{\partial b} &= \frac{\partial \Delta_d(b;k)}{\partial b} + \frac{\partial T_f^a\,(b)}{\partial b} \\ &= \frac{cg(b)}{(1-k)}\left(\frac{1}{G^2(-b)} - \frac{k}{G^2(b)}\right) + \frac{g(b-h)-g(b)}{\left(G(b)-G(b-h)\right)^2}c.\end{aligned}$$

Notice that the second term, which is the derivative of $T_f^a(b)$, is negative for $b < 0.5h$ and positive for $b \geq 0.5h$, while the first term, the derivative of $\Delta_d(b;k)$, is always positive. This implies that $\frac{\partial T_d^c(b;k)}{\partial b} > 0$ for all $b > 0.5h$. To characterize more precisely, let $\Psi \equiv \{0 \leq b < 0.5h | \frac{\partial T_d^c(b;k)}{\partial b} = 0\}$ denote the set of its extreme points, and denote by $\bar{b}_d \equiv \sup \Psi$ its upper bound (let $\bar{b}_d = 0$ if Ψ is empty), then $\frac{\partial T_d^c(b;k)}{\partial b} > 0$ for $b > \bar{b}_d$. Finally, let $\bar{b} \equiv \max\{\bar{b}_f, \bar{b}_d\}$, then $T^c(b;k) = \min\{T_d^c(b;k), T_f^c(b;k)\}$ increases in b for all $b \geq \bar{b}$. Notice that $\bar{b} \leq \max\{G^{-1}\left(\frac{1}{1+\sqrt{k}}\right), \frac{h}{2}\}$ since $\bar{b}_f \leq b^0$ and $\bar{b}_d \leq h/2$.

On the other hand, let $\underline{b}_f \equiv \inf \Sigma$ if Σ is not empty (otherwise let $\underline{b}_f = +\infty$), and let $\underline{b}_d \equiv \inf \Psi$ if Ψ is not empty (otherwise let $\underline{b}_d = +\infty$). Denote by $\underline{b} \equiv \min\{\underline{b}_f, \underline{b}_d\}$, we show that $T^c(b;k)$ decreases in b for $b < \underline{b}$. To see this, evaluating the derivative of $T_f^c(b;k)$ at $b = 0$, we obtain

$$\begin{aligned}
\frac{\partial T_f^c(0;k)}{\partial b} &= \frac{\partial \Delta_f(0;k)}{\partial b} + \frac{\partial T_d^a(0)}{\partial b} \\
&= \frac{cg(0)}{(1-k)}\left(\frac{k}{G^2(0)} - \frac{1}{G^2(0)}\right) + \frac{(g(0)-g(h))\,c}{(G(h)-G(0))^2} \\
&= -\frac{g(0)c}{G^2(0)} + \frac{(g(0)-g(h))\,c}{(G(h)-G(0))^2},
\end{aligned}$$

while evaluating $\frac{\partial T_d^c(b;k)}{\partial b}$ at $b = 0$ yields

$$\frac{\partial T_d^c(0;k)}{\partial b} = \frac{cg(0)}{G^2(0)} + \frac{(g(h)-g(0))\,c}{(G(h)-G(0))^2} = -\frac{\partial T_f^c(b;k)}{\partial b}|_{b=0}.$$

Consider two cases:

Case (1): If $\frac{\partial T_f^c(0;k)}{\partial b} < 0$, then $T_d^c(b;k)$ decreases for b sufficiently close to zero and increases for $b > \bar{b}_d$. Thus there must exist some $b \in (0, \bar{b}_d]$ such that $T_d^c(b;k) = 0$, that is, the set Ψ is not empty. In this case, $\underline{b}_d$ must be a local minimum of $T_d^c(b;k)$, and $\underline{b}_d \le \bar{b}_d$. On the other hand, since $\frac{\partial T_f^c(0;k)}{\partial b} > 0$, then $T_f^c(b;k)$ increases for b sufficiently close to zero and also increases for $b \ge \bar{b}_f$. If $T_f^c(b;k)$ increases for all $b > 0$, then $\Sigma = \emptyset$ and $\underline{b}_f = +\infty$, thus, $\underline{b} = \underline{b}_d$; otherwise, Σ is not empty and $\underline{b} \equiv \min\{\underline{b}_f, \underline{b}_d\}$.

Case (2): If $\frac{\partial T_f^c(0;k)}{\partial b} < 0$ (and $\frac{\partial T_d^c(0;k)}{\partial b} > 0$), then by analogy $T_f^c(b;k)$ decreases for b sufficiently close to zero and increases for $b \ge \bar{b}_f$, and thus $\underline{b}_f \le \bar{b}_f$ is the local minimum. On the other hand, since $\frac{\partial T_d^c(0;k)}{\partial b} > 0$, $T_d^c(b;k)$ increases for b sufficiently close to zero and also increases for $b \ge \bar{b}_d$. If $T_d^c(b;k)$ increases for all $b > 0$, then $\Psi = \emptyset$ and $\underline{b}_f = +\infty$, then $\underline{b} = \underline{b}_d$; otherwise Ψ is not empty and $\underline{b} \equiv \min\{\underline{b}_f, \underline{b}_d\}$.

Therefore there always exists some $\underline{b}$ such that $0 < \underline{b} = \min\{\underline{b}_f, \underline{b}_d\} \le \bar{b}$. Consider the neighborhood of 0, $N(0) = [0, \underline{b}]$, where both $T_f^c(b;k)$ and $T_d^c(b;k)$ are monotonic in the neighborhood. If $T_f^c(b;k)$ is increasing in this neighborhood, then $T_d^c(b;k)$ is decreasing. Note that $T_d^c(0;k) = T_f^c(0;k)$, then $T_f^c(b;k) \ge T_f^c(0;k) = T_d^c(0;k) \ge T_d^c(b;k)$, and thus $T^c(b;k) = T_d^c(b;k)$ is decreasing in $N(0)$. If instead $T_f^c(b;k)$ is decreasing in b and $T_d^c(b;k)$ is increasing, then $T_d^c(b;k) \ge T_d^c(0;k) = T_f^c(0;k) \ge T_f^c(b;k)$, and $T^c(b;k) = T_f^c(b;k)$ is also decreasing in $N(0)$.

Appendix D: Proof of Proposition 2

We have shown that collusion on the low efforts can be deterred if and only if $t \geq T^c(b;k)$. However, the agents could also collude on other effort levels when collusion on low efforts is not sustainable. In this simple model, since each agent can choose only two effort levels, there are three possible combinations of efforts other than the pair of high efforts, and we need then consider two other cases.

Case A: Suppose agents agree that the favored agent takes the high effort while the disfavored one takes the low effort, and we denote by $\hat{e} = (e_f, e_d) = (h, 0)$ the pair of efforts in this case. Using the same approach we adopted in the case of collusion of low efforts, we can derive the conditions for preventing such collusion.

The favored agent is willing to participate collusion if the expected payoff, $G(h+b)(t-s_f) + (1 - G(h+b))\, ks_d - c$, is higher than that absent collusion, $G(b)t - c$, which implies

$$G(h+b)s_f - G(-h-b)ks_d < (G(h+b) - G(b))\, t. \qquad (CIR_f\,(\hat{e}))$$

Likewise, the disfavored one is willing to participate if $G(-h-b)(t-s_d) + (1 - G(-h-b))\, ks_f > G(-b)t - c$, which implies

$$G(-h-b)s_d - G(h+b)ks_f < c - (G(h+b) - G(b)))\, t, \qquad (CIR_d\,(\hat{e}))$$

where we have used the relation $G(-b) - G(-h-b) = G(h+b) - G(b)$.

Moreover, the favored agent may deviate from the collusive agreement and instead take the low effort, in which case it could receive an expected payoff equal to $G(b)(t-s_f) + (1 - G(b))\, ks_d$. Such deviation is not profitable if

$$G(h+b)(t-s_f) + (1 - G(h+b))\, ks_d - c > G(b)(t-s_f) + (1 - G(b))\, ks_d,$$

which implies

$$t - s_f - ks_d > \frac{c}{G(h+b) - G(b)} = T^a_d(b). \quad (CIC_f\,(\hat{e}))$$

By contrast, the disfavored agent may deviate and instead take the low effort, by which it could earn the expected payoff $G(-b)(t-s_d) + (1 - G(-b))\, ks_f - c$. The disfavored agent does not benefit from such deviation if

$$G(-h-b)(t-s_d) + (1 - G(-h-b))\, ks_f > G(-b)(t-s_d) + (1 - G(-b))\, ks_f - c,$$

which amounts to

$$t - s_d - ks_f < \frac{c}{G(-b) - G(-h-b)} = \frac{c}{G(h+b) - G(b)}. \quad (CIC_d\,(\hat{e}))$$

To collude on the effort pair $\hat{e}$, the above four constraints must be satisfied. Denote by $\Gamma(\hat{e})$ the set of all incentive feasible side payments (s_f, s_d) that satisfy the above four constraints. We show now $\Gamma(\hat{e}) = \emptyset$ when the prize t is sufficiently high. From $(CIR_d(\hat{e}))$, we obtain

$$s_d < \frac{G(h+b)ks_f}{G(-h-b)} + \frac{c - (G(h+b) - G(b))\,t}{G(-h-b)}.$$

By rearranging $(CIR_f(\hat{e}))$ and using the above relation, we obtain

$$\begin{aligned} s_f &< \frac{G(-h-b)ks_d}{G(h+b)} + \frac{(G(h+b) - G(b))\,t}{G(h+b)} \\ &< \frac{G(-h-b)k}{G(h+b)}\left(\frac{G(h+b)ks_f}{G(-h-b)} + \frac{c - (G(h+b) - G(b))\,t}{G(-h-b)}\right) + \frac{(G(h+b) - G(b))\,t}{G(h+b)} \\ &= k^2 s_f + \frac{kc}{G(h+b)} + \frac{(G(h+b) - G(b))\,(1-k)\,t}{G(h+b)}, \end{aligned}$$

which implies

$$s_f < \frac{kc}{G(h+b)\,(1-k^2)} + \frac{(G(h+b) - G(b))\,t}{G(h+b)\,(1+k)}.$$

By analogy, the side payment s_d must be bounded above

$$s_d < \frac{c}{G(-h-b)\,(1-k^2)} - \frac{(G(h+b) - G(b))\,t}{G(-h-b)\,(1+k)}. \tag{9}$$

On the other hand, from $(CIC_d(\hat{e}))$ we obtain

$$s_f < t - T_d^a(b) - ks_d,$$

and $(CIC_d(\hat{e}))$ we have

$$s_d > t - T_d^a(b) - ks_f.$$

Combining both constraints yields

$$\begin{aligned} s_d &> t - T_d^a(b) - ks_f \\ &> t - T_d^a(b) - k\,(t - T_d^a(b) - ks_d) \\ &= (1-k)\,(t - T_d^a(b)) + k^2 s_d, \end{aligned}$$

which implies that s_d must be bounded below by

$$s_d > \frac{t - T_d^a(b)}{1+k}. \tag{10}$$

Thus, to ensure the existence of incentive feasible side payments (s_f, s_d), the lower bound of s_d must be less than its upper bound, which requires (combining the equations (9) and (10)):

$$\frac{t - T_d^a(b)}{1+k} < \frac{c}{G(-h-b)\,(1-k^2)} - \frac{(G(h+b) - G(b))\,t}{G(-h-b)\,(1+k)}.$$

Solving for t implies the necessary condition for the existence of incentive feasible side payments (s_f, s_d):

$$\begin{aligned} t &< \hat{T}(b;k) \equiv \frac{c}{G(-b)(1-k)} + \frac{G(-h-b)c}{(G(h+b)-G(b))\,G(-b)} \\ &= \frac{c}{G(-b)(1-k)} + \frac{G(-h-b)}{G(-b)} T_d^a(b). \end{aligned}$$

It thus follows that $\Gamma(\hat{e}) = \emptyset$ if $t \geq \hat{T}(b;k)$.

Case B: Alternatively, the agents could collude on the effort pair that involves the favored one taking the low effort while the disfavored one take the high effort, which is denoted by $\tilde{e} = (0, h)$. By analogy, we can derive the four constraints for the incentive feasible side payments. The favored agent is willing to participate if the expected payoff under collusion, $G(b-h)(t-s_f) + (1-G(b-h))ks_d$, exceeds the payoff absent collusion, $G(b)t - c$, which further requires

$$G(b-h)\,s_f - G(h-b)\,ks_d < c - (G(b) - G(b-h))\,t. \quad (CIR_f(\tilde{e}))$$

Similarly, the disfavored agent is willing to participate collusion if

$$G(h-b)\,s_d - G(b-h)\,ks_f < (G(h-b) - G(-b))\,t. \quad (CIR_d(\tilde{e}))$$

Moreover, the favored agent may deviate and instead take the high effort, which yields an expected payoff $G(b)(t-s_f) + (1-G(b))ks_d - c$, and such deviation does not benefit the favored agent if

$$t - s_f - ks_d < \frac{c}{G(b) - G(b-h)} = T_f^a(b). \;\; (CIC_f(\tilde{e}))$$

Likewise, deviation from collusion is not profitable for the disfavored agent if

$$t - s_d - ks_f > \frac{c}{G(h-b) - G(-b)} = T_f^a(b). \;\; (CIC_d(\tilde{e}))$$

Let $\Gamma(\tilde{e})$ denote the set of all incentive feasible side contracts that satisfy all the above constraints. We show now $\Gamma(\tilde{e}) = \emptyset$ if t is large enough.

To see this, by rearranging $(CIR_f(\tilde{e}))$ we obtain

$$s_f < \frac{G(h-b))ks_d}{G(b-h)} + \frac{c}{G(b-h)} - \frac{(G(b) - G(b-h))\,t}{G(b-h)},$$

and rearranging $(CIR_d(\tilde{e}))$ leads to

$$s_d < \frac{G(b-h)\,ks_f}{G(h-b)} + \frac{(G(h-b) - G(-b))\,t}{G(h-b)}.$$

Combing these two constraints, we have

$$\begin{aligned}
s_f &< \frac{G(h-b))ks_d}{G(b-h)} + \frac{c}{G(b-h)} - \frac{(G(b)-G(b-h))\,t}{G(b-h)} \\
&< \frac{G(h-b))k}{G(b-h)}\left(\frac{G(b-h)\,ks_f}{G(h-b)} + \frac{(G(h-b)-G(-b))\,t}{G(h-b)}\right) + \frac{c}{G(b-h)} - \frac{(G(b)-G(b-h))\,t}{G(b-h)} \\
&= k^2 s_f + \frac{c}{G(b-h)} - \frac{(G(b)-G(b-h))\,(1-k)\,t}{G(b-h)},
\end{aligned}$$

which implies the upper bound of s_f:

$$s_f < \frac{c}{G(b-h)\,(1-k^2)} - \frac{(G(b)-G(b-h))\,t}{G(b-h)\,(1+k)}.$$

By analogy, we obtain

$$\begin{aligned}
s_d &< \frac{G(b-h)\,ks_f}{G(h-b)} + \frac{(G(h-b)-G(-b))\,t}{G(h-b)} \\
&< \frac{G(b-h)\,k}{G(h-b)}\left(\frac{G(h-b))ks_d}{G(b-h)} + \frac{c}{G(b-h)} - \frac{(G(b)-G(b-h))\,t}{G(b-h)}\right) + \frac{(G(h-b)-G(-b))\,t}{G(h-b)} \\
&= k^2 s_d + \frac{kc}{G(h-b)} + \frac{(G(h-b)-G(-b))\,(1-k)\,t}{G(h-b)},
\end{aligned}$$

which further implies the upper bound of s_d:

$$s_d < \frac{kc}{G(h-b)\,(1-k^2)} + \frac{(G(b)-G(b-h))\,t}{G(h-b)\,(1+k)}.$$

On the other hand, we can rewrite $(CIC_f(\tilde{e}))$ as

$$s_f > t - T_f^a(b) - ks_d,$$

and also rewrite $(CIC_d(\tilde{e}))$ as

$$s_d < t - T_f^a(b) - ks_f.$$

Therefore, combining the above two constraints leads to

$$\begin{aligned}
s_f &> t - T_f^a(b) - ks_d \\
&> t - T_f^a(b) - k\left(t - T_f^a(b) - ks_f\right) \\
&= (1-k)\left(t - T_f^a(b)\right) + k^2 s_f,
\end{aligned}$$

which implies the lower bound for s_f:

$$s_f > \frac{t - T_f^a(b)}{1+k}.$$

To ensure the existence of incentive feasible side payments, the lower bound of s_f must be lower than its upper bound, which requires

$$\frac{t-T_f^a(b)}{1+k}<\frac{c}{G\left(b-h\right)\left(1-k^2\right)}-\frac{\left(G(b)-G\left(b-h\right)\right)t}{G\left(b-h\right)\left(1+k\right)}.$$

Solving for t, we obtain the necessary condition for the existence of incentive feasible side payments

$$t<\tilde{T}\left(b;k\right)\equiv\frac{c}{G\left(b\right)\left(1-k\right)}+\frac{G\left(b-h\right)c}{\left(G(b)-G\left(b-h\right)\right)G(b)}.$$

It follows that $\Gamma\left(\tilde{e}\right)=\emptyset$ if $t\geq\tilde{T}\left(b;k\right)$.

We show now $T^c(b;k)>\{\hat{T}\left(b;k\right),\tilde{T}\left(b;k\right)\}$, that is, preventing collusion on low efforts is more costly than preventing collusion on other effort levels. For this purpose, we need the following lemma:

Lemma 4 *The assumption that the inverse hazard rate $H\left(x\right)=G\left(x\right)/g\left(x\right)$ is (strictly) increasing implies that $G\left(x\right)^2>G\left(x-z\right)G\left(x+z\right)$ for any x and any positive z.*

Proof: Let $\Omega(z)\equiv G\left(x-z\right)G\left(x+z\right)$. Then

$$\begin{aligned}\Omega'(z) &= G\left(x-z\right)g\left(x+z\right)-g\left(x-z\right)G\left(x+z\right)\\ &= g\left(x+z\right)g\left(x-z\right)\left(\frac{G\left(x-z\right)}{g\left(x-z\right)}-\frac{G\left(x+z\right)}{g\left(x+z\right)}\right)<0,\end{aligned}$$

as the function $G\left(\cdot\right)/g\left(\cdot\right)$ is increasing. It follows that $\Omega(z)<\Omega(0)=G\left(x\right)^2$. **Q.E.D.**

It is easy to see that $T_f^c(b;k)>\hat{T}\left(b;k\right)$:

$$\begin{aligned}& T_f^c(b;k)-\hat{T}\left(b;k\right)\\ =\ & \frac{c\left(kG(b)+G(-b)\right)}{G(-b)G(b)\left(1-k\right)}+\frac{c}{\left(G(h+b)-G(b)\right)}-\frac{c}{G(-b)\left(1-k\right)}-\frac{G(-h-b)c}{\left(G(h+b)-G(b)\right)G(-b)}\\ =\ & \frac{c\left(G(-b)-\left(1-k\right)G(b)\right)}{G(-b)G(b)\left(1-k\right)}+\frac{c}{G(-b)}\\ =\ & \frac{c}{G(b)\left(1-k\right)}>0.\end{aligned}$$

Moreover, the above lemma implies that $G(-b)^2>G(h-b)G(-h-b)$, and thus

$$\begin{aligned}& T_d^c(b;k)-\hat{T}\left(b;k\right)\\ =\ & \frac{c\left(G(b)+kG(-b)\right)}{G(-b)G(b)\left(1-k\right)}+\frac{c}{G(h-b)-G(-b)}-\frac{c}{G(-b)\left(1-k\right)}-\frac{G(-h-b)c}{\left(G(h+b)-G(b)\right)G(-b)}\\ =\ & \frac{ck}{G(b)\left(1-k\right)}+\frac{c}{G(h-b)-G(-b)}-\frac{G(-h-b)c}{\left(G(h+b)-G(b)\right)G(-b)}\\ =\ & \frac{ck}{G(b)\left(1-k\right)}+\frac{G(-b)^2-G(h-b)G(-h-b)}{\left(G(h-b)-G(-b)\right)\left(G(h+b)-G(b)\right)G(-b)}c\\ >\ & 0.\end{aligned}$$

Hence, $T^c(b;k) = \min\{T^c_f(b;k), T^c_d(b;k)\} > \hat{T}(b;k)$.

Next, we show that $T^c(b;k) > \tilde{T}(b;k)$. It is easy to check that

$$\begin{aligned}
& T^c_d(b;k) - \tilde{T}(b;k) \\
= \; & \frac{c\,(G(b)+kG(-b))}{G(-b)G(b)\,(1-k)} + \frac{c}{(G(b)-G(b-h))} - \frac{c}{G(b)\,(1-k)} - \frac{G(b-h)\,c}{(G(b)-G(b-h))\,G(b)} \\
= \; & \frac{c}{G(-b)\,(1-k)} + \frac{kc}{G(b)\,(1-k)} - \frac{c}{G(b)\,(1-k)} + \frac{c}{G(b)} \\
= \; & \frac{c}{G(-b)\,(1-k)} > 0.
\end{aligned}$$

On the other hand, comparing $T^c_f(b;k)$ and $\tilde{T}(b;k)$, we have

$$\begin{aligned}
& T^c_f(b;k) - \tilde{T}(b;k) \\
= \; & \frac{c\,(kG(b)+G(-b))}{G(-b)G(b)\,(1-k)} + \frac{c}{(G(h+b)-G(b))} - \frac{c}{G(b)\,(1-k)} - \frac{G(b-h)\,c}{(G(b)-G(b-h))\,G(b)} \\
= \; & \frac{ck}{G(-b)\,(1-k)} + \frac{c}{(G(h+b)-G(b))} - \frac{G(b-h)\,c}{(G(b)-G(b-h))\,G(b)} \\
= \; & \frac{ck}{G(-b)\,(1-k)} + \frac{G(b)^2 - G(h+b)G(b-h)}{(G(h+b)-G(b))\,(G(b)-G(b-h))\,G(b)}c \\
> \; & 0,
\end{aligned}$$

where the last line comes from relation $G(b)^2 > G(b+h)G(b-h)$ by Lemma 2. It thus follows that $T^c(b;k) = \min\{T^c_f(b;k), T^c_d(b;k)\} > \tilde{T}(b;k)$.

Finally, we show that $T^c(b;k) > T^a(b) = T^a_d(b)$. Since $T^c_f(b;k) = \Delta_f(b;k) + T^a_d(b) > T^a_d(b)$, it suffices to show $T^c_d(b;k) > T^a_d(b)$. To see this, notice that

$$\begin{aligned}
& T^c_d(b;k) - T^a_d(b) \\
= \; & \frac{ck}{G(b)\,(1-k)} + \frac{c}{G(-b)\,(1-k)} + \frac{c}{(G(b)-G(b-h))} - \frac{c}{(G(h+b)-G(b))} \\
> \; & \frac{c}{G(-b)} + \frac{c}{(G(b)-G(b-h))} - \frac{c}{(G(h+b)-G(b))} \\
= \; & \frac{G(-b)G(h+b) - G(-b)G(b) - (1-G(h+b))\,G(b) + (1-G(h+b))\,G(b-h)}{G(-b)\,(G(h+b)-G(b))\,(G(b)-G(b-h))} \\
= \; & \frac{G(h+b)G(h-b) + G(-b)G(-b) - G(h-b)}{G(-b)\,(G(h+b)-G(b))\,(G(b)-G(b-h))} \\
> \; & \frac{G(h+b)G(h-b) + G(h-b)\,G(-h-b) - G(h-b)}{G(-b)\,(G(h+b)-G(b))\,(G(b)-G(b-h))} \\
= \; & 0,
\end{aligned}$$

where we have used the fact $k < 1$ to derive the first inequality, the relations $G(h+b) + G(-h-b) = 1$, $G(h-b) + G(b-h) = 1$, and $G(b) + G(-b) = 1$ in the third and fourth lines, and the relation $G(-b)^2 > G(h-b)\, G(-h-b)$ to derive the second inequality. Thus, $T^c(b;k) > T_d^a(b) = T^a(b)$.

To summarize, $t \geq T^c(b;k)$ implies $t > \{\tilde{T}(b;k), \hat{T}(b;k)\}$, thus collusion on all possible effort levels other than the high efforts is not sustainable if $t \geq T^c(b;k)$. Moreover, since $t \geq T^c(b;k) > T^a(b)$, the prize is sufficiently high to induce high efforts, which can be implemented through the side contract with $s_f = 0 = s_d$. The equilibrium incentive prize must be that $t^*(b) = T^c(b;k)$. By lemma 2, the optimal degree of favoritism must satisfy $\underline{b} \leq b^* \leq \bar{b}$. **Q.E.D.**

Appendix E: Proof of Lemma 3

Notice that

$$\begin{aligned}
\frac{\partial T_f^c(b;k)}{\partial b} &= \frac{\partial \Delta_f(b;k)}{\partial b} + \frac{\partial T_d^a(b;k)}{\partial b} \\
&= \frac{c}{(1-k)}\left(\frac{kg(b)}{G^2(-b)} - \frac{g(b)}{G^2(b)}\right) + \frac{g(b)-g(h+b)}{\left(G(h+b)-G(b)\right)^2}c.
\end{aligned}$$

Differentiating $\frac{\partial \Delta_f(b;k)}{\partial b}$ with respect to b, we obtain

$$\begin{aligned}
\frac{\partial^2 \Delta_f(b;k)}{\partial b^2} &= \frac{c}{(1-k)}\left(\frac{kg'(b)G(-b)+2kg^2(b)}{G^3(-b)} + \frac{2g^2(b)-g'(b)G(b)}{G^3(b)}\right) \\
&= \frac{c}{(1-k)}\left(\frac{kG^3(b)\left(g'(b)G(-b)+2g^2(b)\right)+G^3(-b)\left(2g^2(b)-g'(b)G(b)\right)}{G^3(-b)G^3(b)}\right).
\end{aligned}$$

We focus on the case where $G^2(-b) \geq kG^2(b)$ (i.e., $b \leq b^0$), then the numerator is positive since

$$\begin{aligned}
& kG^3(b)\left(g'(b)G(-b)+2g^2(b)\right)+G^3(-b)\left(2g^2(b)-g'(b)G(b)\right) \\
\geq\ & kG^2(b)\left(G(b)\left(g'(b)G(-b)+2g^2(b)\right)+G(-b)\left(2g^2(b)-g'(b)G(b)\right)\right) \\
=\ & kG^2(b)\left(2g^2(b)\left(G(-b)+G(b)\right)\right) \\
>\ & 0.
\end{aligned}$$

Since the denominator is also positive, it follows that $\frac{\partial^2 \Delta_f(b;k)}{\partial b^2} > 0$ for $b < b^0$. Moreover, differentiating $\frac{\partial T_d^a(b;k)}{\partial b}$ with respect to b, we obtain

$$\begin{aligned}
\frac{\partial^2 T_d^a(b;k)}{\partial b^2} &= \frac{\left(g'(b)-g'(h+b)\right)\left(G(h+b)-G(b)\right)+2\left(g(b)-g(h+b)\right)^2}{\left(G(h+b)-G(b)\right)^3} \\
&\geq \frac{2\left(g(b)-g(h+b)\right)^2}{\left(G(h+b)-G(b)\right)^3} \\
&\geq 0,
\end{aligned}$$

where the first inequality comes from the fact that $g'(b) \geq g'(h+b)$ by the assumption of weak concavity of $g(\cdot)$. Thus, for $b \leq b^0$,

$$\frac{\partial^2 T_f^c(b;k)}{\partial b^2} = \frac{\partial^2 \Delta_f(b;k)}{\partial b^2} + \frac{\partial^2 T_d^a(b;k)}{\partial b^2} > 0.$$

On the other hand, recall that

$$\begin{aligned}
\frac{\partial T_d^c(b;k)}{\partial b} &= \frac{\partial \Delta_d(b;k)}{\partial b} + \frac{\partial T_f^a(b)}{\partial b} \\
&= \frac{c}{(1-k)}\left(\frac{g(-b)}{G^2(-b)} - \frac{kg(b)}{G^2(b)}\right) + \frac{g(b-h)-g(b)}{(G(b)-G(b-h))^2}c.
\end{aligned}$$

Differentiating $\frac{\partial T_f^a(b)}{\partial b}$ with respect to b yields

$$\frac{\partial^2 T_f^a(b)}{\partial b^2} = \frac{(g'(b-h)-g'(b))(G(b)-G(b-h)) + 2(g(b-h)-g(b))^2}{(G(b)-G(b-h))^3}.$$

Since $g'(b-h) \geq 0$ and $g'(b) \leq 0$ for $b < h/2$, it follows that $\frac{\partial^2 T_f^a(b)}{\partial b^2} \geq 0$. Moreover, differentiating $\frac{\partial \Delta_d(b;k)}{\partial b}$ with respect to b leads to

$$\frac{\partial^2 \Delta_d(b;k)}{\partial b^2} = \frac{c}{(1-k)}\left(\frac{g'(b)G(-b)+2g^2(b)}{G^3(-b)} + \frac{2kg^2(b)-kg'(b)G(b)}{G^3(b)}\right).$$

Note that the second term in the blanket is (strictly) positive since $g'(b)$ is negative, thus $\frac{\partial^2 \Delta_d(b;k)}{\partial b^2} > 0$ if the first term is also positive. Denote by $\Pi(b) = g'(b)G(-b) + 2g^2(b)$, then

$$\Pi'(b) = g''(b)G(-b) + 3g(b)g'(b) \leq 0.$$

Since by assumption

$$\Pi(\frac{h}{2}) = g'(\frac{h}{2})G(-\frac{h}{2}) + 2g^2\left(\frac{h}{2}\right) \geq 0,$$

it follows that $\Pi(b) \geq 0$ for all $b \leq h/2$. Hence, $\frac{\partial^2 T_f^c(b;k)}{\partial b^2} > 0$ for all $b \leq h/2$. **Q.E.D.**

References

Albright, M. and C. Carr (1997) *Biggest Mistakes Managers Make: And How to Avoid Them*, Prentice Hall, Inc.

Berger, Johannes, Claus Herbertz, and Dirk Sliwka (2011), "Managerial Incentives and Favoritism in Promotion Decisions: Theory and Field Evidence", *IZA working paper* No. 5543.

Comerford, H (2002) "Findings Report Respectful Workplace Project", Government of Newfoundland and Labrador, mimeo.

Green, J and Stocky, N (1983) "A comparison of Tournament and Contracts" *Journal of Political Economy* 91:349-364.

Ishiguro, S (2004) "Collusion and Discrimination in Organizations" *Journal of Economic Theory* 116: 357-369.

Kwon, I (2006) "Endogenous Favoritism in Organizations" *Topics in Theoretical Economics* 6(1): 1-24.

Laffont, J-J and Martimort, D (2002) *The Theory of Incentives 1: the Principal-Agent Model* Princeton University Press.

Lazear, E and S. Rosen (1981) "Rank Order Tournaments as Optimal Labor Contracts" *Journal Political Economy* 89:841-864.

Miller, G (1992) *Managerial Dilemma: The Political Economy of Hierarchy,* Cambridge University Press, New York, NY.

Nalebuff, B and J.E. Stigilitz (1983) "Prizes and Incentives: Towards a General Theory of Compensation and Competition" *Bell Journal of Economics* 13: 21-43.

Prendergast, C (1999) "The Provision of Incentives in Firms" *Journal of Economic Literature* 37: 7-63.

Prendergast, C and R. Topel (1993) "Discretion and Bias in Performance Evaluation" *European Economic Review* 37: 355-365.

Prendergast, C and R. Topel (1996) "Favoritism in Organizations" *Journal of Political Economy* 104: 958-978.

Tirole, J (1986) "Hierarchies and Bureaucracies: On the Role of Collusion in Organizations" *Journal of Law, Economics and Organization*, 2: 181-214.

Tirole, J (1992) "Collusion and the Theory of Organizations" In *Advances in Economic Theory: Sixth World Congress*, vol 2, edited by Jean-Jacques Laffont.