Department of Economics Discussion Paper Series

Selecting the Best when Selection is Hard

Mikhail Drugov, Margaret Meyer and Marc Möller

Number 981
July, 2022

Department of Economics
Manor Road Building, Manor Road
Oxford, OX1 3UQ

ISSN 1471-0498

Selecting the Best when Selection is Hard*

Mikhail Drugov,† Margaret Meyer‡ and Marc Möller§

July 14, 2022

Abstract

In dynamic promotion contests, where performance measurement is noisy and constrained to be ordinal, selection of the most able agent can be improved by biasing later stages in favor of early performers. We show that even in the *worst-case scenario*, where external random factors swamp the difference in agents' abilities in determining their relative performance, optimal bias is (i) strictly positive and (ii) locally insensitive to changes in the ratio of heterogeneity to noise. To explain these, arguably surprising, limiting results, we demonstrate a close relationship in the limit between optimal bias under ordinal information and the expected optimal bias when bias can be conditioned on cardinal information about relative performance. As a consequence of these two limiting properties, the simple rule of setting bias as if in the worst-case scenario achieves most of the potential gains in selective efficiency from biasing dynamic rank-order contests.

Keywords: Dynamic Contests; Selective Efficiency; Bias; Learning; Promotions.
JEL *classification*: D21, D82, D83, M51.

1 Introduction

Assigning productive resources or decision-making authority to the most able individuals constitutes a source of economic gains, whose importance becomes most apparent in the face of dramatic success or failure. For instance, the passing of the UK premiership from Neville Chamberlain to Winston Churchill, rather than to Lord Halifax, during the early years of Word War II brought to an end the British policy of appeasement and has

*We thank Paul Klemperer, Dmitry Ryvkin and participants of the 2021 Transatlantic Theory Workshop, the 8th Annual Conference on "Contests: Theory and Evidence" and the 2022 Conference on Mechanism and Institution Design for valuable comments. All errors are our own.

†New Economic School and CEPR, mdrugov@nes.ru.

‡Nuffield College and Department of Economics, Oxford University, and CEPR, margaret.meyer@nuffield.ox.ac.uk.

§University of Bern, marc.moeller@vwi.unibe.ch.

been credited as a major contributor to the Allied victory (Roberts, 2018). Conversely, the infamous decline of Kodak has been attributed to the appointment of CEO Kay Whitmore, who was criticized for lacking the visionary foresight of his rival, Phil Samper, concerning the emergence of digital photography (Oren, 2019). Family businesses and monarchies are further examples where the allocation of power on the basis of bloodline rather than talent has frequently led to collapse.

In modern organizations and societies, selection of the best frequently takes the form of a multi-stage contest in which the winners of earlier stages may be given an advantage at a later stage. For example, consulting or law firms may assign their most effective mentors to the best-performing juniors, thereby improving these juniors' chances of becoming partners. Similarly, in academia, well-publishing assistant professors may be awarded research grants or teaching reductions, which may increase their future output and hence their probability of obtaining tenure. An important question in such a setting is how large an advantage, or *bias*, should be granted to early winners, when the goal is to maximize *selective efficiency*, i.e., the expected ability of those who are promoted.

The problem of choosing bias optimally for selection is complicated by the fact that performance measurement is often constrained to be ordinal rather than cardinal. Moreover, performance rankings typically provide only a noisy measure of ability, and the degree of randomness can be substantial due to the complexity of the underlying task.[1] Finally, skills that can make an important difference at the next higher level—such as leadership style or managerial vision—may have only little impact on current performance.[2] Consequently, in many practically relevant cases, *selection is hard*, in that external random factors (performance shocks or measurement errors) are of large magnitude relative to the differences in abilities that determine agents' relative performance.

Consider what we will term the *worst-case scenario*, where selection is hardest, because the ratio of the heterogeneity in agents' abilities to the scale of the noise tends to zero. In this scenario, intuition may suggest that organizations should refrain from the use of bias altogether, because bias then tends to be assigned entirely on the basis of luck rather than evidence of ability. Our starting point is the observation that this intuition is incomplete. It neglects the fact that the optimal size of bias depends not only on the informativeness of the (early) stages before bias is introduced, but also on how bias impacts the informativeness of the (later) stages in which bias is actually employed.

Section 2 presents a stylized model of a two-agent, two-stage selection contest, capable

[1] Lazear (2000) documents that for managers, piece rates are employed ten times less frequently than for operatives, and attributes this difference to the difficulty of measuring managerial performance.

[2] Hansen et al. (2021) report the increasing relevance of social skills in top managerial positions and emphasize the importance of designing mechanisms to facilitate the match between firms and executives.

of capturing the resulting trade-off. Individual performance, at each stage, is the sum of an agent's time-invariant, unobservable ability, multiplied by a stage-specific weight, and a transitory noise component, identically distributed across stages. The organization observes the *ordinal* ranking of performances at each stage and can assign an additive bias to the first-stage winner's performance in the second stage. The organization's problem is to choose the size of the bias to maximize the expected productivity of the winner of the final stage, where productivity is an arbitrary increasing function of an agent's ability.

Generally, optimal bias favors the first-stage winner and is such that, if the first-stage loser just managed to achieve a second-stage draw despite the bias against him, the organization would be indifferent as to which agent to select. Optimal bias thus strikes a balance between the informativeness of the ordinal first-stage ranking and the marginal second-stage outcome (a draw). Yet, as we illustrate in Section 3, the impact on optimal bias of a change in the ratio of the heterogeneity in agents' abilities to the scale of the noise can be complex. Despite this complexity, we demonstrate that, under mild assumptions, optimal bias satisfies two striking properties *in the limit* as the heterogeneity-to-noise ratio goes to zero: (1) optimal bias remains positive; and (2) optimal bias is locally insensitive to the ratio of heterogeneity to noise.

To provide further insight into these properties of optimal bias, Section 4 considers the counterfactual situation in which relative performance information was *cardinal* rather than ordinal. In this case, the second-stage bias could be conditioned on the first-stage margin of victory. If, for example, the stage-specific weights attached to abilities are equal, then the optimal bias would advantage the first-stage winner by exactly the first-stage margin of victory. We show that, in the worst-case scenario, the optimal bias under ordinal information (characterized in Section 3) equals an appropriately-defined average of the optimal cardinal biases, as the margin of victory varies. Because a positive cardinal bias would be necessary to offset *any* non-zero margin of victory, optimal ordinal bias must thus also be positive. Moreover, in the worst-case scenario, expected optimal cardinal bias, just like optimal ordinal bias, would be locally insensitive to the heterogeneity-to-noise ratio.

Section 5 focuses on the performance of optimal bias. The limiting results in Section 3 imply that the positive value of optimal bias in the worst-case scenario represents a second-order approximation for the optimal bias when the heterogeneity-to-noise ratio is small. We prove that as a consequence, the organization's payoff from setting bias as if in the worst-case scenario is a fifth-order approximation for its maximized payoff when noise is large relative to heterogeneity. This analytical result thus suggests that the simple heuristic of setting bias *as if selection is hardest* may perform well even when selection

is only moderately hard. We confirm this conjecture by showing that, for an important family of distributions, containing normally and uniformly distributed noise as special cases, a large fraction of the potential gains in selective efficiency, for a wide range of values of the heterogeneity-to-noise ratio, are captured by setting bias at the level that would be optimal in the worst-case scenario.

To highlight the informational role of bias in improving selective efficiency in dynamic contests, our main model suppresses agents' effort incentives. Section 6 shows that the properties of optimal bias for selection are robust to allowing agents to choose efforts strategically. We assume that performance is additive in effort and that each agent's only informational advantage relative to the organization is the private observation of his effort. In the unique equilibrium, in each stage agents choose equal efforts. Efforts thus have no effect on the informativeness of either stage about relative abilities, so all of our conclusions about the impact of bias on selective efficiency remain valid.

Related literature The use of biases—also called "handicaps" or, when additive, "head starts"—is a popular topic in the contest literature. However, the focus has been on the influence of bias on effort incentives, rather than on selection. Starting with Lazear and Rosen (1981) and O'Keeffe, Viscusi and Zeckhauser (1984), it has been recognized that biases are useful in "leveling the playing field" and creating "competitive balance" among heterogeneous competitors when the organizer aims to maximize the total effort of the participants (see also, e.g., Meyer, 1992; Schotter and Weigelt, 1992; Fain, 2009; Epstein, Mealem and Nitzan, 2011; Franke et al., 2013).[3] In dynamic settings, biases may mitigate the "momentum", or "dynamic discouragement", effect, which arises from an imbalance in accumulated wins or losses (Barbieri and Serena, 2022).

Despite the importance of contests as mechanisms to select the most-able or highest-valuing individuals, little attention has been paid to the role of bias in enhancing selective efficiency. Besides the early contribution of Meyer (1991), the few exceptions we are aware of include Kawamura and Moreno de Barreda (2014), who show that biasing an all-pay auction with ex ante identical bidders may increase the likelihood of awarding the object to the highest-valuing bidder, and Drugov and Ryvkin (2017), who provide several similar results for noisy contests. Our paper builds on Meyer (1991), who provides the basic insight that bias favoring early winners can improve selective efficiency. Our focus instead is on the properties and performance of optimal bias.

[3]More recent literature such as Drugov and Ryvkin (2017) and Fu and Wu (2020) highlights the limits of the "leveling the playing field" argument beyond the most popular contest models.

2 Model

We consider an organization consisting of a risk-neutral principal and two agents $i \in \{A, B\}$ with differing abilities. Agents' abilities a_i are assumed to be identically distributed on $\{\underline{a}, \underline{a} + h\}$.[4] The parameter $h > 0$ captures the degree of potential *heterogeneity* in agents' abilities. The principal needs to select one of the agents for promotion. The principal's choice is complicated by the fact that abilities are not observable and that performance information is noisy and only ordinal rather than cardinal.

To capture the dynamic nature of selection processes, we assume that the principal observes the agents' relative performance during two stages. In each stage $t \in \{1, 2\}$, the performance of agent i, $p_{i,t}$, is the sum of the agent's time-invariant ability a_i, multiplied by a stage-specific weight λ_t, and a time-varying random component $\epsilon_{i,t}$. That is,

$$p_{i,t} \equiv \lambda_t a_i + \epsilon_{i,t}.$$

Variation in λ_t over time can capture variation across stages in the impact of ability on performance.

The principal identifies the agent with the higher performance $p_{i,1}$ as the winner of the first stage, with ties broken randomly. In the second stage, the principal may assign a bias $\beta \in \Re$ to the winner of the first stage. Having won the first stage, agent i then becomes the winner of the second stage if $p_{i,2} + \beta > p_{j,2}$. If agent i wins the second stage, he is promoted, and the principal's payoff is given by $\Pi(a_i)$, where Π is an increasing function measuring the productivity of the promoted agent in the new job.

As stage outcomes depend only on the performance *differentials* between agents, the distribution of the *difference* in the noise terms, $\Delta\epsilon_t \equiv \epsilon_{A,t} - \epsilon_{B,t} \in \Re$, is a key primitive of our model. We assume that $\Delta\epsilon_t$ is i.i.d. across stages. Denote its support by $[-z, z]$ (where z may be infinite), its cdf by G, and its density by g.

Assumption 1 *(i) g is symmetric around 0; (ii) g is strictly log-concave, i.e., $\ln g$ is strictly concave; (iii) g is twice differentiable on $(-z, z)$; (iv) $\lim_{y \to z} L(y) = \infty$, where*

$$L(y) \equiv -\frac{g'(y)}{g(y)}.$$

The symmetry of g captures the idea that the only source of asymmetry between agents is their (initially unknown) difference in abilities and the explicit second-stage bias; it is a weaker assumption than agents' shocks being i.i.d. Log-concavity of g is equivalent to the

[4] All of our propositions about selection remain valid when $a_i = \underline{a} + h\alpha_i$, where the joint distribution of (α_i, α_j) is symmetric with respect to the two components but otherwise arbitrary.

monotone likelihood ratio property in this setting; it guarantees that, in either stage, the larger the performance-difference, between agents A and B, $p_{A,t} - p_{B,t}$, the higher is the likelihood that A's ability exceeds B's, relative to the likelihood that B's ability exceeds A's. The assumption that log-concavity is strict implies that L is strictly increasing. It is technical and ensures, together with the remaining two assumptions, that the principal's problem is well-behaved.

The principal's problem is to choose the size of the bias β to maximize the expected productivity of the promoted agent. Given our distributional assumptions, this objective is equivalent to maximizing *selective efficiency*, $S(\beta, h)$, defined as the probability of promoting the more able agent, conditional on agents' abilities being different.[5]

The key parameter $h > 0$, capturing the degree of potential heterogeneity in agents' abilities, has a broader interpretation as the *ratio* of agents' heterogeneity to the scale of noise. To see this, introduce a scaling transformation $\Delta\epsilon_t \to \sigma\Delta\epsilon_t$, with $\sigma > 1$, which makes the difference in the noise terms more dispersed: The cdf becomes $G(\frac{\Delta\epsilon_t}{\sigma})$, the pdf $\frac{1}{\sigma}g(\frac{\Delta\epsilon_t}{\sigma})$, and the support $[-\sigma z, \sigma z]$. If the underlying heterogeneity in abilities is H, then $G(\frac{\lambda_1 H}{\sigma})$ is the probability that the more able agent wins the first stage. It depends on H and σ only through the *heterogeneity-to-noise ratio* $h \equiv \frac{H}{\sigma}$. A large part of our analysis will focus on the limit as $h \to 0$, where the scale of noise becomes large *relative* to the agents' heterogeneity.

3 Optimal bias

The selective efficiency $S(\beta, h)$ of the dynamic contest is the probability with which, conditional on agents' abilities being different, the more able agent wins the final stage. Given that the first stage is unbiased, the probability that the more able agent wins the first stage is $G(\lambda_1 h)$. In contrast, given any non-zero value of bias β, the more able agent's probability of winning the second stage depends on the first-stage outcome. If the more able agent won the first stage, then his chance of winning the second stage is $G(\lambda_2 h + \beta)$, whereas if he lost his chance of winning is $G(\lambda_2 h - \beta)$. Overall, selective efficiency is

$$S(\beta, h) = G(\lambda_1 h)G(\lambda_2 h + \beta) + [1 - G(\lambda_1 h)]G(\lambda_2 h - \beta). \tag{1}$$

[5] One can show that a randomly assigned first-stage bias does not increase selective efficiency, as long as the contest is non-discriminatory, in that the size of the second-stage bias cannot condition on the first-stage winner's *identity*. We thus abstract from the possibility that bias is assigned in *both* stages.

Differentiating with respect to β and rearranging leads to the following first-order condition for the optimal bias:

$$\frac{G(\lambda_1 h)}{1 - G(\lambda_1 h)} = \frac{g(\lambda_2 h - \beta)}{g(\lambda_2 h + \beta)}. \tag{2}$$

The ratio on the left-hand side, which is larger than one, is the relative likelihood that the first-stage winner is the more able agent rather than the less able one. The higher this likelihood ratio, the more informative is the first-stage ranking about agents' relative abilities. The ratio on the right-hand side is also a likelihood ratio: It is the relative likelihood that a draw in the second stage ($p_{i,2} + \beta = p_{j,2}$) arises when the more able agent is disadvantaged by the bias compared to when the bias advantages this agent. The higher this likelihood ratio, the more informative a signal is a second-stage draw about the relative ability of the first-stage loser—who managed to achieve a draw despite being handicapped by the bias.

For $\beta = 0$, the right-hand side equals one, so a second-stage draw is uninformative. Moreover, given the strict log-concavity of g, as the size of the bias increases, a second-stage draw becomes a strictly stronger signal about the relative ability of the first-stage loser. It thus follows from Assumption 1 that the first-order condition (2) has a unique solution, $\beta^*(h)$, which maximizes selective efficiency. Optimal bias strikes a balance between the informativeness of the ordinal ranking from stage one and that of the marginal outcome (a draw) in stage two. More precisely, optimal bias is such that, if the principal were to observe a draw in stage two, she would be indifferent about which agent to promote. Accordingly, optimal bias is increasing (decreasing) in λ_1 (λ_2), the sensitivity of performance in stage one (stage two) to ability.

Though the logic behind the optimal level of bias is clear, the dependence of $\beta^*(h)$ on the heterogeneity-to-noise ratio h can be complex, because a fall in h reduces both sides of (2): It lowers both the informativeness of a first-stage win and, by log-concavity, the informativeness of a second-stage draw about the relative ability of the first-stage loser, for any given level of bias. The complex dependence of $\beta^*(h)$ on h is illustrated in Figure 1. The left-hand panel plots the density functions for the family of exponential power distributions with mean 0 and shape parameter $\alpha > 1$. These density functions are (see, e.g., Evans, Hastings and Peacock, 2003)

$$g(\Delta\epsilon_t; \alpha) = \frac{\alpha}{2\Gamma(\frac{1}{\alpha})} \exp(-|\Delta\epsilon_t|^\alpha), \tag{3}$$

and for all $\alpha > 1$, they satisfy Assumption 1. For $\alpha = 2$, $g(\Delta\epsilon_t; \alpha)$ is a normal distribution with variance $\frac{1}{2}$; as $\alpha \to \infty$, $g(\Delta\epsilon_t; \alpha)$ approaches a uniform distribution with support $[-1, 1]$; and as $\alpha \to 1$, $g(\Delta\epsilon_t; \alpha)$ approaches a Laplace distribution with scale parameter

1. The right-hand panel plots the optimal bias $\beta^*(h)$ as a function of h, for $\lambda_1 = \lambda_2 = 1$. Despite the myriad possibilities illustrated, two regularities are suggested by the plots. First, even as h gets small, optimal bias remains positive for all members of the family. Second, as h gets small, optimal bias appears to become locally insensitive to h.

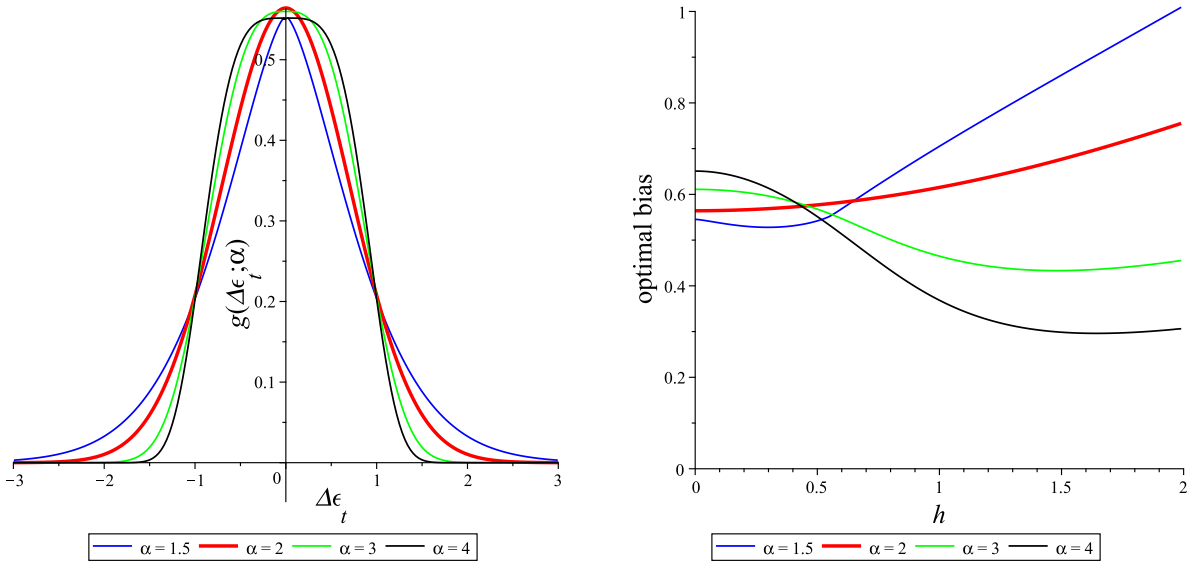

Figure 1: **Optimal bias for the family of exponential power distributions with mean 0 and shape parameter $\alpha > 1$.** The left-hand panel plots the density function and the right-hand panel plots the optimal bias, as h varies, for $\lambda_1 = \lambda_2 = 1$.

Proposition 1 shows that these regularities hold in general.

Proposition 1 *For all $\lambda_1, \lambda_2 > 0$, in the worst-case scenario where the heterogeneity-to-noise ratio h tends to zero, optimal bias satisfies the following two properties:*

(i) Optimal bias $\beta_0^* \equiv \lim_{h\to 0} \beta^*(h)$ is strictly positive and satisfies

$$L(\beta_0^*) = 2\frac{\lambda_1}{\lambda_2} g(0). \tag{4}$$

(ii) Optimal bias is locally insensitive to the heterogeneity-to-noise ratio:

$$\lim_{h\to 0} \frac{d\beta^*(h)}{dh} = 0.$$

At first sight, the fact that optimal bias remains strictly positive, even as the scale of the noise swamps the heterogeneity in abilities, may seem counterintuitive, because when h tends to 0, the first-stage ranking becomes uninformative about relative abilities. This argument resonates well with the frequently raised concern that dynamic selection

contests may be skewed in favor of the most lucky rather than the most able.[6] However, this reasoning neglects the fact that, as h tends to 0, a second-stage draw also becomes uninformative about the relative ability of the first-stage loser, for any level of bias. Formally, as h tends to 0, both sides of (2) approach 1. Part (i) of Proposition 1 characterizes optimal bias as h tends to 0 by equating the *rates* at which the informativeness of the two stages tend to zero as h gets small. Since L is a strictly increasing function, $L(0) = 0$, and the right-hand side of (4) is positive, the limiting value of optimal bias must be positive.

Part (ii) of Proposition 1 shows that, as h tends to 0, changes in h have no first-order effect on the size of optimal bias. In Section 5, we prove that as a consequence, the organization's payoff from setting $\beta = \beta_0^*$ is a fifth-order approximation for its maximized payoff when h is small. We also show numerically that the simple heuristic of setting bias as if in the worst-case scenario captures a large fraction of the potential gains from bias, for a wide range of values of the heterogeneity-to-noise ratio.

4 A counterfactual: cardinal information

To provide further intuition for Proposition 1, in this section we consider the counterfactual situation where the organization has access to stage-one relative performance information that is *cardinal*, rather than ordinal.[7] Second-stage bias can then condition on the first-stage *margin of victory* $k \equiv |p_{A,1} - p_{B,1}|$. The optimal bias based only on ordinal information, $\beta^*(h)$, can be thought of, loosely, as a form of average of the optimal cardinal biases $\beta^{card}(k, h)$ as the margin of victory k varies. Proposition 2 below, which mirrors Proposition 1, makes this intuition precise for the worst-case scenario where the heterogeneity-to-noise ratio h tends to 0.

The properties of optimal bias given stage-one cardinal relative performance information are particularly transparent when performance in the two stages is equally sensitive to ability, that is, $\lambda_1 = \lambda_2$. Here, it is optimal to select the agent with the higher *total* performance over the two stages. This optimal selection rule can be implemented by biasing the second-stage contest in favor of the stage-one winner by exactly k, the stage-one margin of victory. Hence, the optimal cardinal bias is

$$\beta^{card}(k, h) = k, \quad \forall k, h. \tag{5}$$

[6] For example, Deaner, Lowen and Cobley (2013) provide evidence for the fact that, due to repeated selection within cohorts of matching age, professional hockey players with dates of birth during the first half of the year are over-represented in the NHL.

[7] Choosing bias optimally in the second stage allows the organization to use ordinal second-stage information as efficiently as if it could observe cardinal relative-performance information in this stage. The same is true in the setting of Section 3.

For this case, it is easy to understand why as the heterogeneity-to-noise ratio h goes to 0, expected optimal cardinal bias satisfies the same two properties satisfied by optimal ordinal bias. First, given (5), expected optimal cardinal bias is just the expected stage-one margin of victory, which, as $h \to 0$, approaches $\mathbb{E}[|\Delta\epsilon_1|] > 0$. Second, an increase in h raises the expectation of k when the better agent wins but lowers it when the better agent loses, and as $h \to 0$, these two effects cancel out, thus making $\mathbb{E}[\beta^{card}(k,h)]$, just like $\beta^*(h)$, locally insensitive to h in this limit.

Proposition 2 shows that this parallel between the properties of optimal ordinal and cardinal bias remains intact even when $\lambda_1 \neq \lambda_2$:

Proposition 2 *For all $\lambda_1, \lambda_2 > 0$, when cardinal information is available, in the worst-case scenario where the heterogeneity-to-noise ratio h tends to zero,*

(i) Expected optimal cardinal bias is strictly positive:

$$\lim_{h\to 0} \mathbb{E}[\beta^{card}(k,h)] > 0.$$

(ii) Expected optimal cardinal bias is locally insensitive to the heterogeneity-to-noise ratio:

$$\lim_{h\to 0} \frac{d\mathbb{E}[\beta^{card}(k,h)]}{dh} = 0.$$

(iii) Optimal biases under ordinal and cardinal information are related according to:

$$L(\beta_0^*) = \mathbb{E}[L(\beta_0^{card}(k))],$$

where $\beta_0^{card}(k) \equiv \lim_{h\to 0} \beta^{card}(k,h)$.

Part (iii) characterizes the precise relationship between ordinal and cardinal biases in the worst-case scenario. In each case, the principal sets the bias so that she would be indifferent whom to promote after a tie, which pins down the likelihood ratio L at the optimal bias. Under ordinal information, the principal must consider all possible margins of victory and set the bias in an "average" way—making expected likelihood ratios at optimal ordinal and cardinal biases equal. When the noise distribution is normal, so L is linear, expected biases are equal: $\beta_0^* = \mathbb{E}[\beta_0^{card}(k)]$. Part (iii) thus provides further perspective on why optimal ordinal bias remains positive even as the noise swamps the heterogeneity in abilities.

5 A simple heuristic for setting bias

Proposition 1 implies that the positive value of optimal bias in the worst-case scenario, β_0^*, is a second-order approximation for the optimal bias, $\beta^*(h)$, when the heterogeneity-to-noise ratio is small. A striking consequence is that, for small h, selective efficiency when $\beta = \beta_0^*$ and selective efficiency when $\beta = \beta^*(h)$ are equal up to order *four*:

Proposition 3 *For all* $\lambda_1, \lambda_2 > 0$,

$$S(\beta^*(h), h) = S(\beta_0^*, h) + O(h^5).$$

Proposition 3 strongly suggests that the simple heuristic of setting bias *as if selection is hardest* may perform well even when selection is only moderately hard. We substantiate this conjecture by showing that, for the family of exponential power distributions introduced in Section 3, a large fraction of the potential gains in selective efficiency, for a wide range of values of the heterogeneity-to-noise ratio, are captured by setting bias at the level that would be optimal in the worst-case scenario.

If second-stage bias is unavailable, then when $\lambda_1 = \lambda_2$, selective efficiency in the two-stage contest is the same as with only one stage—since if the two stages are won by different agents, it is *an* optimal rule to select the winner of the first stage. More generally, i.e., when $\lambda_1 \neq \lambda_2$, the optimal selection rule without bias promotes the winner of the more informative stage, making selective efficiency equal to the maximum of $G(\lambda_1 h)$ and $G(\lambda_2 h)$. Define the *gain in selective efficiency* from using bias β, relative to using no bias, as

$$\Delta(\beta, h) \equiv S(\beta, h) - \max\{G(\lambda_1 h), G(\lambda_2 h)\}. \tag{6}$$

Note that $\Delta(\beta, h)$ is bounded above by $\frac{1}{2}$. The gain in selective efficiency when bias is optimally calibrated to h is $\Delta(\beta^*(h), h)$, while the gain under the simple rule of setting bias equal to β_0^* is $\Delta(\beta_0^*, h)$.

The maximum potential gain from using bias, $\Delta(\beta^*(h), h)$, is largest when performance in the two stages is equally sensitive to ability ($\lambda_1 = \lambda_2$). Raising $\min\{\lambda_1, \lambda_2\}$, holding $\max\{\lambda_1, \lambda_2\}$ fixed, raises $\Delta(\beta^*(h), h)$ because two-stage selective efficiency, $S(\beta^*(h), h)$, increases while one-stage selective efficiency, $\max\{G(\lambda_1 h), G(\lambda_2 h)\}$, is unchanged. Moreover, raising $\max\{\lambda_1, \lambda_2\}$, holding $\min\{\lambda_1, \lambda_2\}$ fixed, lowers $\Delta(\beta^*(h), h)$ because it increases $\max\{G(\lambda_1)h, G(\lambda_2 h)\}$ more than $S(\beta^*(h), h)$. Since the maximum potential gain from bias is greatest when $\lambda_1 = \lambda_2$, it is natural to focus, at least initially, on this case.

For the exponential power family of densities (3) with shape parameter $\alpha > 1$, optimal bias in the worst-case scenario is readily calculated and given by $\beta_0^* = \{\lambda_1/[\lambda_2 \Gamma(\frac{1}{\alpha})]\}^{\frac{1}{\alpha-1}}$.

Figure 2 plots, for various values of α, λ_1 and λ_2, $\Delta(\beta^*(h), h)$ (solid curves) and $\Delta(\beta_0^*, h)$ (dashed curves) as h varies.

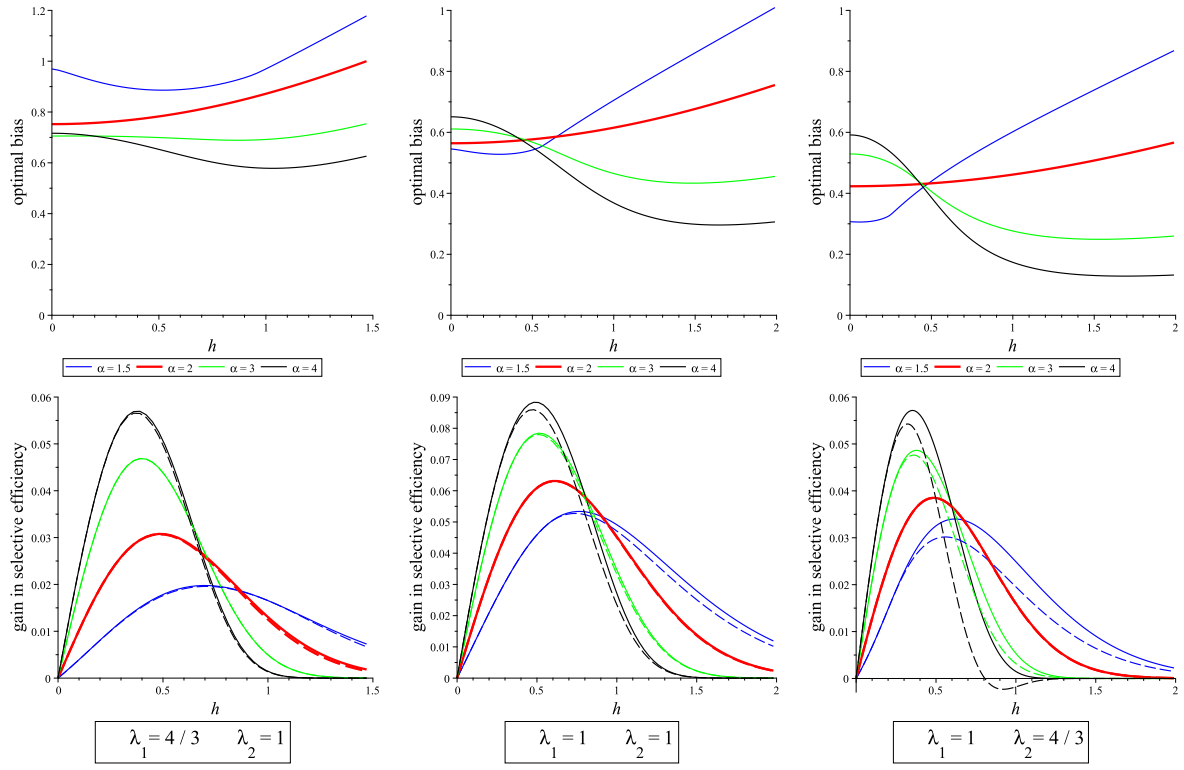

Figure 2: **Gain in selective efficiency due to bias.** The top panels plot the optimal bias and the bottom panels plot $\Delta(\beta^*(h), h)$ (solid curves) and $\Delta(\beta_0^*, h)$ (dashed curves), defined in (6), as h varies, for various values of the shape parameter α, λ_1 and λ_2.

Figure 2 substantiates the conjecture that setting bias equal to its value β_0^* in the worst-case scenario captures a very large fraction of the potential gains in selective efficiency. Focusing first on the middle column ($\lambda_1 = \lambda_2 = 1$), this is true independently of the shape parameter α and holds for a wide range of values of the heterogeneity-to-noise ratio h. Although optimal bias $\beta^*(h)$ can be increasing, decreasing, or non-monotone in h, Proposition 1 guarantees that there exists a range of h over which $\beta^*(h)$ exhibits rather limited variation, with the consequence (Proposition 3) that over this range, $\Delta(\beta_0^*, h)$ very closely approximates $\Delta(\beta^*(h), h)$. The extremely good overall performance of the heuristic of setting bias equal to β_0^* arises because this range of h, broadly speaking, contains the range of h values for which bias matters: Once h becomes very large, even $\Delta(\beta^*(h), h)$ declines rapidly, because a single-stage unbiased contest is already very effective at identifying the more able agent.

As explained above, the importance of choosing bias wisely declines as the difference

in informativeness of the two contest stages grows. Figure 2(left) shows that for $\lambda_1 = \frac{4}{3}$, $\lambda_2 = 1$, the heuristic of setting bias equal to β_0^* continues to perform extremely well; this reflects the even more pronounced insensitivity of $\beta^*(h)$ with respect to h. Figure 2(right) shows that for $\lambda_1 = 1$, $\lambda_2 = \frac{4}{3}$, this heuristic also performs well, except for densities g approaching the uniform distribution (large α). The weaker performance of the heuristic for this one case reflects the relatively rapid decrease in $\beta^*(h)$ with h, which results in β_0^* being significantly too large once h exceeds 1. Strikingly, Figure 2, both for equal and unequal values of λ_1 and λ_2, shows that with normally distributed noise ($\alpha = 2$), the optimal bias is particularly insensitive to h. Consequently, with normally distributed noise, the heuristic of setting bias as if in the worst-case scenario performs exceptionally well over the whole range of heterogeneity-to-noise ratios.

6 Efforts

To highlight the role of bias in improving selective efficiency, we have so far suppressed agents' effort incentives. We now show that the properties of optimal bias for selection are robust to allowing agents to choose efforts strategically in each stage.

Let the performance of agent i in stage t be given by

$$p_{i,t} = \lambda_t a_i + e_{i,t} + \epsilon_{i,t},$$

where $e_{i,t}$ is agent i's privately chosen effort in stage t. Consistent with our assumption that all heterogeneity between agents is captured by the parameter h, we assume that agents attach the same value (normalized to 1) to receiving the promotion after stage two and that, within each stage, agents have identical, strictly convex costs of effort, $c_t(e_{i,t})$.[8] We assume that agents enter the dynamic contest no better informed about their abilities than the organization and receive only ordinal information about their performance. The organization chooses the stage-two bias β in favor of the stage-one winner to maximize the contest selective efficiency, taking into account the induced equilibrium effort choices.

Proposition 4 *When performance is additive in effort and each agent's only informational advantage relative to the organization is his private choice of efforts,*

(i) There is a unique pure-strategy equilibrium. In this equilibrium, both in the first and in the second stage, agents choose identical efforts.

[8]It is well known that sufficient convexity of effort cost functions ensures that second-order conditions for utility maximization are satisfied.

(ii) The bias that maximizes selective efficiency is the same as in the model without efforts, i.e., it is given by (2) and satisfies the properties in Proposition 1.

In the second stage, the agents are no longer symmetric: The first-stage winner is both more likely to be more able and is advantaged by the bias. Yet, since efforts affect the contest outcome only through the difference between them, the marginal return to effort is the same for the two agents, so the second-stage efforts of the first-stage winner and loser are equal. In the first stage, since the agents are ex ante symmetric, there exists a pair of identical efforts that are best responses to each other. Furthermore, our proof shows that unequal stage-one efforts could not be best responses. Since equilibrium effort choices by the agents are identical within each stage, efforts have no effect on the informativeness of either stage of the contest about relative abilities. The bias that maximizes selective efficiency is thus the same as in the model without efforts.

7 Conclusion

Our characterization of optimal bias in dynamic contests where selection is hard is of both theoretical and practical value. It shows why bias should continue to be used, even as the role of luck relative to ability in determining whom bias should favor grows very large, and it illuminates the forces that determine optimal bias in this limit. Moreover, the limiting properties of optimal bias explain why, even in environments where selection is only moderately hard, the simple heuristic of setting bias as if selection is hardest performs strikingly well.

While we provided conditions under which our results are robust with respect to the introduction of efforts, an open issue is how selective efficiency and optimal bias are influenced by incentives when agents enter the contest better informed than the organization about their abilities. More able agents might then be inclined to exert higher or lower first-stage effort because they attach a larger or a smaller value to the prospect of being advantaged by bias in the second stage. Bias could thus affect selective efficiency not only directly, via the informational channel highlighted by our analysis, but also indirectly, via its influence on incentives.

Appendix

Proof of Proposition 1

The optimal second-stage bias maximizes selective efficiency $S(\beta, h)$ in (1). Subindices denote partial derivatives.

Part (i) For any $h > 0$, Assumption 1 ensures that the first-order condition $S_\beta(\beta, h) = 0$ uniquely determines the optimal bias $\beta^*(h)$:

$$S_\beta(\beta^*(h), h) = G(\lambda_1 h)\, g(\lambda_2 h + \beta^*(h)) - (1 - G(\lambda_1 h))\, g(\lambda_2 h - \beta^*(h)) = 0.$$

However, $\lim_{h\to 0} S_\beta(\beta, h) = 0\ \forall \beta$. Characterizing $\beta_0^* \equiv \lim_{h\to 0} \beta^*(h)$ requires totally differentiating $S_\beta(\beta^*(h), h)$ with respect to h, setting it equal to 0, and letting $h \to 0$. Total differentiation yields

$$S_{\beta h}(\beta^*(h), h) + S_{\beta\beta}(\beta^*(h), h)\frac{\partial \beta^*(h)}{\partial h}, \tag{7}$$

where $\lim_{h\to 0} S_{\beta\beta}(\beta, h) = 0\ \forall \beta$ (since $\lim_{h\to 0} S_\beta(\beta, h) = 0\ \forall \beta$). Hence, (7) and Assumption 1(i) imply that β_0^* solves

$$\lim_{h\to 0} S_{\beta h}(\beta^*(h), h) = S_{\beta h}(\beta_0^*, 0) = 2\lambda_1 g(0) g(\beta_0^*) + \lambda_2 g'(\beta_0^*) = 0, \tag{8}$$

which gives (4). Since Assumptions 1(i) and 1(iii) guarantee $L(0) = 0$, $\beta_0^* > 0$.

Part (ii) Evaluating $\lim_{h\to 0} \frac{\partial \beta^*(h)}{\partial h}$ requires totally differentiating (7) with respect to h, setting it equal to 0, and letting $h \to 0$. Total differentiation of (7) yields

$$S_{\beta hh}(\beta^*(h), h) + 2S_{\beta\beta h}(\beta^*(h), h)\frac{\partial \beta^*(h)}{\partial h} + S_{\beta\beta\beta}\left(\frac{\partial \beta^*(h)}{\partial h}\right)^2 + S_{\beta\beta}(\beta^*(h), h)\frac{\partial^2 \beta^*(h)}{\partial h^2}, \tag{9}$$

where $\lim_{h\to 0} S_{\beta\beta\beta}(\beta, h) = 0\ \forall \beta$ (since $\lim_{h\to 0} S_{\beta\beta}(\beta, h) = 0\ \forall \beta$). Assumption 1 implies that $\lim_{h\to 0} S_{\beta hh}(\beta, h) = 2\lambda_1^2 g'(0) g(\beta) = 0\ \forall \beta$, and that $\lim_{h\to 0} S_{\beta\beta h}(\beta^*(h), h) = 2\lambda_1 g(0) g'(\beta_0^*) + \lambda_2 g''(\beta_0^*)$, which is negative by (4). Hence, (9) implies that $\lim_{h\to 0} \frac{\partial \beta^*(h)}{\partial h} = 0$. ■

Proof of Proposition 2

Part (i) Given the observed first-stage margin of victory $k \geq 0$, the principal chooses β to maximize $S^{card}(\beta, k, h)$, the probability of promoting the more able agent, conditional on agents' abilities being different. Conditional on abilities being different, the probability of margin of victory k is $g(k - \lambda_1 h)$ when the stronger agent wins and $g(k + \lambda_1 h)$ when the weaker agent wins. Hence

$$S^{card}(\beta, k, h) = g(k - \lambda_1 h) G(\lambda_2 h + \beta) + g(k + \lambda_1 h) G(\lambda_2 h - \beta).$$

The first-order condition is

$$S_\beta^{card}(\beta, k, h) = g(k - \lambda_1 h)g(\lambda_2 h + \beta) - g(k + \lambda_1 h)g(\lambda_2 h - \beta) = 0 \tag{10}$$

which, by Assumption 1, uniquely determines the optimal cardinal bias $\beta^{card}(k, h)$ as a strictly increasing function of k, equal to 0 for $k = 0$. Since $\lim_{h\to 0} S_\beta(\beta, k, h) = 0 \ \forall \beta, k$, characterizing $\beta_0^{card}(k) \equiv \lim_{h\to 0} \beta^{card}(k, h)$ requires totally differentiating $S_\beta^{card}(\beta^{card}(k, h), k, h)$ with respect to h, setting it equal to 0, and letting $h \to 0$. Steps paralleling the proof of Proposition 1(i) show that $\beta_0^{card}(k)$ solves $\lim_{h\to 0} S_{\beta h}^{card}(\beta, k, h) = 0$, which yields

$$L(\beta_0^{card}(k)) = \frac{\lambda_1}{\lambda_2} L(k). \tag{11}$$

By Assumption 1, $L(0) = 0$ and $L(k) > 0 \ \forall k > 0$. Hence, $\beta_0^{card}(k) > 0 \ \forall k > 0$.

To compute $\mathbb{E}[\beta^{card}(k, h)]$, denote by $q_{\Delta a}^0$ the prior probability that $a_A - a_B = \Delta a \in \{-h, 0, h\}$. Since a_A and a_B are identically distributed,

$$q_{-h}^0 = q_h^0. \tag{12}$$

The unconditional density of k on its support $[0, z + \lambda_1 h]$ is

$$2q_h^0 g(k - \lambda_1 h) + 2q_h^0 g(k + \lambda_1 h) + 2q_0^0 g(k). \tag{13}$$

Hence

$$\begin{aligned}
\mathbb{E}[\beta^{card}(k, h)] &= 2 \int_0^{z+\lambda_1 h} \beta^{card}(k, h) \left[q_h^0 g(k - \lambda_1 h) + q_h^0 g(k + \lambda_1 h) + q_0^0 g(k) \right] dk \\
&= 2q_h^0 \int_{-\lambda_1 h}^{z} \beta^{card}(v + \lambda_1 h, h) g(v) dv + 2q_h^0 \int_{\lambda_1 h}^{z+2\lambda_1 h} \beta^{card}(v - \lambda_1 h, h) g(v) dv \\
&\quad + 2q_0^0 \int_0^z \beta^{card}(k, h) g(k) dk,
\end{aligned} \tag{14}$$

Since $2q_h^0 + q_0^0 = 1$,

$$\lim_{h\to 0} \mathbb{E}[\beta^{card}(k, h)] = 2 \int_0^z \beta_0^{card}(v) g(v) dv.$$

Since $\beta_0^{card}(k) > 0 \ \forall k > 0$, $\lim_{h\to 0} \mathbf{E}[\beta^{card}(k, h)] > 0$.

Part (ii) Totally differentiating (14) with respect to h and letting $h \to 0$ yields

$$\lim_{h\to 0} \frac{d\mathbb{E}[\beta^{card}(k, h)]}{dh} = 2 \int_0^z \frac{\partial \beta^{card}(v, 0)}{\partial h} g(v) dv, \tag{15}$$

since the two integrals involving the derivative of β^{card} with respect to its first argument cancel out when $h \to 0$. Steps paralleling the proof of Proposition 1(ii) then show that $\forall k$, $\lim_{h\to 0} \frac{\partial \beta^{card}(k,h)}{\partial h} = 0$, so (15) equals 0.

Part (iii) Given (4) and (11), we need only show that $\mathbb{E}[L(k)] = 2g(0)$. As $h \to 0$, (13) converges to $2g(k)$ on support $[0, z]$. Hence

$$\mathbb{E}[L(k)] = \int_0^z L(k)2g(k)dk = -2\int_0^z g'(k)dk = 2g(0),$$

using $g(z) = 0$, which is implied by Assumption 1(iii). ■

Proof of Proposition 3

The proposition claims that

$$\lim_{h\to 0} \frac{d^n}{dh^n} \left[S(\beta^*(h), h) - S(\beta_0^*, h)\right] = 0 \qquad \text{for } n = 1, 2, 3, 4.$$

Under Assumption 1(i), as $h \to 0$, the following partial derivatives of $S(\beta^*(h), h)$ go to 0 $\forall \beta$: S_β, $S_{\beta\beta}$, $S_{\beta\beta\beta}$, $S_{\beta\beta\beta\beta}$, S_{hh}, and $S_{\beta hh}$. Also, $\lim_{h\to 0} S_{\beta h}(\beta^*(h), h) = 0$ and $\lim_{h\to 0} \beta^{*\prime}(h) = 0$, by Proposition 1(i) and 1(ii). The claim is verified by totally differentiating $S(\beta^*(h), h) - S(\beta_0^*, h)$ n times, for $n = 1, 2, 3, 4$, and taking $h \to 0$. ■

Proof of Proposition 4

Part (i) Denote by superscripts W and L the cases when agent A won and lost, respectively, the first stage. Define $\Delta e_1 = e_{A,1} - e_{B,1}$, $\Delta e_2^W = e_{A,2}^W - e_{B,2}^W$, and $\Delta e_2^L = e_{A,2}^L - e_{B,2}^L$. Define $q_{\Delta a}^W(\Delta e_1)$ (respectively, $q_{\Delta a}^L(\Delta e_1)$) as the posterior probability that $a_A - a_B = \Delta a$, given A won (respectively, lost) the first stage and given Δe_1.

Step 1 We first show that in stage two, the agents exert the same effort. In case W, A's and B's second-stage utilities are, respectively,

$$q_h^W(\Delta e_1)G\left(h + \beta + \Delta e_2^W\right) + q_0^W(\Delta e_1)G\left(\beta + \Delta e_2^W\right) + q_{-h}^W(\Delta e_1)G\left(-h + \beta + \Delta e_2^W\right) - c_2\left(e_{A,2}^W\right)$$
$$q_h^W(\Delta e_1)G\left(-h - \beta - \Delta e_2^W\right) + q_0^W(\Delta e_1)G\left(-\beta - \Delta e_2^W\right) + q_{-h}^W(\Delta e_1)G\left(h - \beta - \Delta e_2^W\right) - c_2\left(e_{B,2}^W\right).$$

By Assumption 1(i), the marginal return to effort is the same for A and B, so $e_{A,2}^{*W} = e_{B,2}^{*W}$. An analogous argument shows $e_{A,2}^{*L} = e_{B,2}^{*L}$.

Step 2 By Assumption 1(i) and condition (12), if $\Delta e_1 = 0$, then $e_{A,2}^{*W} = e_{A,2}^{*L}$. We now show that if $e_{A,1} > e_{B,1}$, then $e_{A,2}^{*W} > e_{A,2}^{*L}$.

Given $e^{*W}_{A,2} = e^{*W}_{B,2}$ and $e^{*L}_{A,2} = e^{*L}_{B,2}$, $e^{*W}_{A,2}$ and $e^{*L}_{A,2}$ satisfy, respectively,

$$\begin{aligned} q^W_h(\Delta e_1)g(h+\beta) + q^W_0(\Delta e_1)g(\beta) + q^W_{-h}(\Delta e_1)g(-h+\beta) &= c'(e^{*W}_{A,2}) \\ q^L_h(\Delta e_1)g(h-\beta) + q^L_0(\Delta e_1)g(-\beta) + q^L_{-h}(\Delta e_1)g(-h-\beta) &= c'(e^{*L}_{A,2}). \end{aligned} \tag{16}$$

Given Assumption 1(i), the difference between the left-hand sides in (16) is

$$\begin{aligned} &[q^W_h(\Delta e_1) - q^L_{-h}(\Delta e_1)]g(h+\beta) + [q^W_{-h}(\Delta e_1) - q^L_h(\Delta e_1)]g(-h+\beta) \\ &\quad + [q^W_0(\Delta e_1) - q^L_0(\Delta e_1)]g(\beta). \end{aligned} \tag{17}$$

To complete Step 2, we show that (17) is strictly positive which, combined with (16), implies that $e^{*W}_{A,2} > e^{*L}_{A,2}$.

Assumption 1(ii) implies that, for $\Delta e_1 > 0$,

$$q^W_h(\Delta e_1) - q^L_{-h}(\Delta e_1) < 0 \qquad \text{and} \qquad q^W_{-h}(\Delta e_1) - q^L_h(\Delta e_1) > 0. \tag{18}$$

We now show that for $\Delta e_1 > 0$, $q^W_0(\Delta e_1) - q^L_0(\Delta e_1) > 0$. Assumption 1(i) and condition (12) give

$$\begin{aligned} & q^W_0(\Delta e_1) > q^L_0(\Delta e_1) \iff q^W_0(\Delta e_1) > q^W_0(-\Delta e_1) \\ \Leftrightarrow \frac{G(\Delta e)}{q^0_h[(G(h+\Delta e_1) + G(-h+\Delta e_1)] + q^0_0 G(\Delta e_1)} &> \frac{G(-\Delta e)}{q^0_h[(G(h-\Delta e_1) + G(-h-\Delta e_1)] + q^0_0 G(-\Delta e_1)} \\ &\iff 2G(\Delta e_1) > G(h+\Delta e_1) + G(-h+\Delta e_1). \end{aligned} \tag{19}$$

Assumptions 1(i) and 1(ii) imply (a) strict convexity of G on $[-z, 0]$ and (b) strict concavity on $[0, z]$. If $-h + \Delta e_1 \geq 0$, (19) follows from (b). Otherwise, (a) and (b) together imply

$$\begin{aligned} G(h+\Delta e_1) + G(-h+\Delta e_1) &< G(h+\Delta e_1) + \left(\frac{2\Delta e_1}{h+\Delta e_1}\right) G(0) + \left(\frac{h-\Delta e_1}{h+\Delta e_1}\right) G(-h-\Delta e_1) \\ &= 2\left(\frac{h}{h+\Delta e_1}\right) G(0) + 2\left(\frac{\Delta e_1}{h+\Delta e_1}\right) G(h+\Delta e_1) \\ &< 2G(\Delta e_1). \end{aligned}$$

Returning to (17), Assumption 1(i) and 1(ii) imply that $g(h+\beta) < g(\beta)$ and $g(h+\beta) < g(-h+\beta)$. Also, the three differences in posteriors in square brackets sum to 0. Hence the inequality $q^W_0(\Delta e_1) - q^L_0(\Delta e_1) > 0$, combined with those in (18), implies that (17) is strictly positive.

Step 3 The overall utility of agent A is

$$
\begin{aligned}
&q_h^0[G\left(h+\Delta e_1\right)\left(G\left(h+\beta+e_{A,2}^{*W}-e_{B,2}^{*W}\right)-c_2\left(e_{A,2}^{*W}\right)\right)\\
&\quad+\left(1-G\left(h+\Delta e_1\right)\right)\left(G\left(h-\beta+e_{A,2}^{*L}-e_{B,2}^{*L}\right)-c_2\left(e_{A,2}^{*L}\right)\right)]\\
&q_0^0[G\left(\Delta e_1\right)\left(G\left(\beta+e_{A,2}^{*W}-e_{B,2}^{*W}\right)-c_2\left(e_{A,2}^{*W}\right)\right)\\
&\quad+\left(1-G\left(\Delta e_1\right)\right)\left(G\left(-\beta+e_{A,2}^{*L}-e_{B,2}^{*L}\right)-c_2\left(e_{A,2}^{*L}\right)\right)]\\
&q_{-h}^0[G\left(-h+\Delta e_1\right)\left(G\left(-h+\beta+e_{A,2}^{*W}-e_{B,2}^{*W}\right)-c_2\left(e_{A,2}^{*W}\right)\right)\\
&\quad+\left(1-G\left(-h+\Delta e_1\right)\right)\left(G\left(-h-\beta+e_{A,2}^{*L}-e_{B,2}^{*L}\right)-c_2\left(e_{A,2}^{*L}\right)\right)]-c_1\left(e_{A,1}\right)
\end{aligned}
$$

A change in $e_{A,1}$ does not affect $e_{B,2}^{*W}$, $e_{B,2}^{*L}$, or β, because it is unobservable, and the local effect via the induced changes in $e_{A,2}^{*W}$ and $e_{A,2}^{*L}$ is zero by the envelope theorem. Using $e_{A,2}^{*W}=e_{B,2}^{*W}$, $e_{A,2}^{*L}=e_{B,2}^{*L}$, Assumption 1(i), and condition (12), the marginal benefit of $e_{A,1}$ simplifies to

$$
\begin{aligned}
&q_h^0\left[g\left(h+\Delta e_1\right)+g\left(-h+\Delta e_1\right)\right]\left\{G\left(h+\beta\right)-c_2\left(e_{A,2}^{*W}\right)-\left[G\left(h-\beta\right)-c_2\left(e_{A,2}^{*L}\right)\right]\right\}\\
&\qquad+q_0^0 g(\Delta e_1)\{G(\beta)-c_2\left(e_{A,2}^{*W}\right)-\left[G\left(-\beta\right)-c_2\left(e_{A,2}^{*L}\right)\right]\}. \qquad (20)
\end{aligned}
$$

Analogously, for agent B the marginal benefit of $e_{B,1}$ becomes

$$
\begin{aligned}
&q_h^0\left[g\left(h-\Delta e_1\right)+g\left(-h-\Delta e_1\right)\right]\left\{G\left(h+\beta\right)-c_2\left(e_{B,2}^{*L}\right)-\left[G\left(h-\beta\right)-c_2\left(e_{B,2}^{*W}\right)\right]\right\}\\
&\qquad+q_0^0 g(\Delta e_1)\{G(\beta)-c_2\left(e_{B,2}^{*L}\right)-\left[G\left(-\beta\right)-c_2\left(e_{B,2}^{*W}\right)\right]\}. \qquad (21)
\end{aligned}
$$

By Assumption 1(i) and Step 1, the difference between (20) and (21) has the sign of $c_2(e_{A,2}^{*L})-c_2(e_{A,2}^{*W})$, which by Step 2 is negative when $e_{A,1}-e_{B,1}>0$. But $e_{A,1}-e_{B,1}>0$ implies $c_1'(e_{A,1})-c_1'(e_{B,1})>0$, so such efforts cannot be optimal for both agents. Analogously, $e_{A,1}<e_{B,1}$ would also yield a contradiction. Hence, equilibrium requires equal first-stage efforts: $e_{A,1}^*=e_{B,1}^*$. These are unique since with $\Delta e_1=0$, (20) and (21) are independent of the common level of e_1.

Part (ii) Since agents' equilibrium efforts are identical within each stage, efforts have no effect on the informativeness of either stage about relative abilities. The bias that maximizes selective efficiency is thus the same as that characterized in Section 3. ■

References

Barbieri, Stefano, and Marco Serena. 2022. "Biasing Dynamic Contests Between Ex-Ante Symmetric Players." https://sites.google.com/site/marcoserenaphd/.

Deaner, Robert O., Aaron Lowen, and Stephen Cobley. 2013. "Born at the Wrong Time: Selection Bias in the NHL Draft." *PLoS ONE*, 8(2).

Drugov, Mikhail, and Dmitry Ryvkin. 2017. "Biased contests for symmetric players." *Games and Economic Behavior*, 103: 116–144.

Epstein, Gil S., Yosef Mealem, and Shmuel Nitzan. 2011. "Political culture and discrimination in contests." *Journal of Public Economics*, 95(1-2): 88–93.

Evans, Merran, Nicholas Hastings, and Brian Peacock. 2003. *Statistical Distributions.* John Wiley & Sons. 3rd edition.

Fain, James R. 2009. "Affirmative Action Can Increase Effort." *Journal of Labor Research*, 30(2): 168–175.

Franke, Jörg, Christian Kanzow, Wolfgang Leininger, and Alexandra Schwartz. 2013. "Effort maximization in asymmetric contest games with heterogeneous contestants." *Economic Theory*, 52(2): 589–630.

Fu, Qiang, and Zenan Wu. 2020. "On the optimal design of biased contests." *Theoretical Economics*, 15(4): 1435–1470.

Hansen, Stephen, Tejas Ramdas, Raffaella Sadun, and Joseph Fuller. 2021. "The Demand for Executive Skills." *CEPR Discussion Paper No. 16280.*

Kawamura, Kohei, and Inés Moreno de Barreda. 2014. "Biasing selection contests with ex-ante identical agents." *Economics Letters*, 123(2): 240–243.

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

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

Meyer, Margaret. 1991. "Learning from coarse information: Biased contests and career profiles." *Review of Economic Studies*, 58(1): 15–41.

Meyer, Margaret A. 1992. "Biased Contests and Moral Hazard: Implications for Career Profiles." *Annals of Economics and Statistics / Annales d'Économie et de Statistique*, 25/26: 165–187.

O'Keeffe, Mary, W. Kip Viscusi, and Richard J. Zeckhauser. 1984. "Economic Contests: Comparative Reward Schemes." *Journal of Labor Economics*, 2(1): 27–56.

Oren, Ravit. 2019. "How mediocre managers ruined Kodak." *HR Magazine*. https://www.hrmagazine.co.uk/content/features/how-mediocre-managers-ruined-kodak.

Roberts, Andrew. 2018. *Churchill - Walking with Destiny.* New York:Viking.

Schotter, Andrew, and Keith Weigelt. 1992. "Asymmetric Tournaments, Equal Opportunity Laws, and Affirmative Action: Some Experimental Results." *Quarterly Journal of Economics*, 107(2): 511–539.