

# Selecting the Best: The Persistent Effects of Luck\*

Mikhail Drugov,<sup>†</sup> Margaret Meyer<sup>‡</sup> and Marc Möller<sup>§</sup>

July 25, 2026

## Abstract

This paper contributes to the debate about the role of luck for individual success, a debate that shapes the perception of economic inequality and preferences over redistributive policies. We consider an organization aiming to select the more able of two competing agents. Selection is based on coarsely-measured performance reflecting unobservable ability, privately-chosen effort, and noise. We show that, even when noise swamps ability differences, optimally designed talent-selection processes favor early strong performers, thus making early luck persistent and, under mild conditions on noise, making early luck more important for selection than late luck. Organizational selection also affects the persistence of luck associated with agents' observable identities: we show that identity-dependent selection processes, such as gender-specific mentoring, create incentives that make selection not only more efficient but also more equitable. Our theory offers testable predictions about the effects of job characteristics on the persistence of both types of luck.

*Keywords:* Organizational learning, Incentives, Selective efficiency, Inequality

*JEL classification:* D21, D82, D83, J70, M51

---

\*For valuable comments, we thank Ghazala Azmat, Luis Garicano, Ángel Hernando-Veciana, Paul Klemperer, Igor Letina, Simon Loertscher, Matthew Mitchell, Inés Moreno de Barreda, Michele Pellizzari, Dmitry Ryvkin, Larry Samuelson, Marta Troya Martinez, and participants in various seminars, workshops, and conferences. All errors are our own.

<sup>†</sup>Universitat Autònoma de Barcelona and Barcelona School of Economics (BSE), New Economic School, and CEPR, mikhail.drugov@uab.cat.

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

<sup>§</sup>University of Bern, marc.moeller@unibe.ch.

## 1. INTRODUCTION

Fortune brings in some boats that are not steered.

---

*William Shakespeare “Cymbeline”*

Sometimes an individual’s success is explained, or even discredited, as resulting from an initial stroke of good luck. [Frank \(2016\)](#) documents a multitude of careers of over-achievers, ranging from the arts to business, that were kick-started by fortunate circumstances or events. Such narratives are extreme examples of a more fundamental debate about the role of luck for individual success, a debate that shapes individual preferences over redistributive policies and public perception of economic inequality more generally.

As differences in economic outcomes appear to be less tolerated when attributed to luck rather than merit ([Fong, 2001](#); [Cappelen et al., 2013](#)), stronger beliefs in the relevance of luck increase a country’s social spending ([Alesina and Angeletos, 2005](#)) and its citizens’ willingness to implement redistributive policies ([Almås et al., 2020](#)). Beliefs about the role of luck also affect political polarization by shaping the *social divide* between the “deserving” and the “undeserving” ([Sandel, 2020](#)). This is especially relevant when beliefs determine the choice between a low-redistribution “American” equilibrium emphasizing the role of merit and a high-redistribution “European” equilibrium acknowledging the role of luck ([Benabou and Tirole, 2006](#); [Alesina et al., 2018](#)).<sup>1</sup> Given the above debate, it is important to understand how economic institutions or organizational practices shape the role and the persistence of luck for individual success.

A common argument, across different social sciences, is that resources, training, mentoring or, more generally, *biases* granted to early strong performers increase the likelihood that an initial stroke of luck translates into a final economic advantage. For example, academic tracking in schools ([Gamoran and Mare, 1989](#)) and professional fast tracks in firms or public agencies ([Baker et al., 1994](#)) magnify the importance of early performance for final success. If initial success derives at least in part from differences in abilities, the use of such biases could be rationalized by *organizational learning*—an organization’s attempt to identify the most productive individuals ([Meyer, 1991](#)). However, a frequent counter-argument is that such biases merely *make luck persistent* by causing fortunate events or circumstances—such as graduating during a boom, being the oldest child in class, or possessing an “advantageous” *identity* (e.g. gender, race, or socioeconomic background)—to have long-lasting effects on labor market outcomes ([Oreopoulos et al., 2012](#)) and educa-

---

<sup>1</sup>Experimental studies on redistribution find that U.S. subjects implement Gini coefficients 0.2 points lower when incomes are based on luck than when incomes are based on merit, which would be sufficient to bring down U.S. inequality to European levels ([Lefgren et al., 2016](#)).

tional achievements (Bedard and Dhuey, 2006; Ciocca-Eller, 2023). Sociologists refer to this mechanism as the *Matthew effect* (Merton, 1968) or *cumulative advantage* (DiPrete and Eirich, 2006), and there is evidence that disadvantages faced by certain groups at different career-stages—e.g. women in school grading (Lavy and Megalokonomou, 2024) or in assessments of their management potential (Benson et al., 2026)—may accumulate over time (Blank, 2005).

Our analysis explains why making luck persistent is not an unwanted consequence of employing too much or the wrong kind of bias, but may be integral to an optimally designed organizational learning process. By rationalizing the persistence of luck as an inseparable feature of careers and illuminating its institutional determinants, we inform the public debate about the relevance of luck—as opposed to merit—for individual success.

In Section 2 we present a stylized but versatile model of a two-agent, two-stage selection process in which individual performance, at each stage, is the sum of an agent’s time-invariant unobservable ability, privately-chosen effort, and a transitory shock. Agents may differ in their “identity”—e.g. gender, race, socioeconomic background—and they share a common prior about their relative abilities. The organization observes the agents’ identities but cannot distinguish the “favorite”—the individual both agents believe is of higher ability—from the “underdog”. After the two stages, the organization attempts to select the more able agent. In each stage, it observes only information about the difference in agents’ performance, which filters out common shocks to both productivity and measurement (cf. Green and Stokey, 1983). Given the difficulty of quantifying performance in many settings, we assume that the organization observes only the ordinal ranking of performances at each stage.<sup>2</sup> Agents choose efforts to maximize the probability of being selected, minus their effort costs. The organization’s optimal selection rule augments the second-stage performance of the agent who performed better in the first stage with an additive bias and selects the agent who performs better in the second stage.

Our main focus is how organizational selection shapes the impact of luck, affecting an agent’s performance in the initial stage of his career, on the probability of being selected in the final stage. We distinguish two types of luck: (1) *early-career luck* that can potentially be obtained by either agent, independently of their identity, and (2) *societal luck* that only one agent obtains, the one with an “advantageous identity”. Conceptually, these two kinds of luck differ in that when luck is associated with an agent’s observable identity, the organization’s choice of bias, as well as the agents’ early actions, may condition on it.

We start our analysis in Section 3 by considering the case where agents are homo-

---

<sup>2</sup>In section 5.2, we analyze the impact of allowing measures of relative performance to be less coarse than ordinal.

geneous in terms of their identity. We first show that, if agents expect the organization to implement, in the second stage, a bias in favor of the first-stage winner, then in the first stage, the favorite will exert a larger effort than the underdog. In turn, the organization will find it optimal to favor the first-stage winner as long as it expects the favorite’s first-stage effort to exceed the underdog’s.

To understand how organizational talent selection affects the persistence of early-career luck, as distinct from the persistent effects of ability differences, we analyze the limiting case where the noise in the environment swamps the effect of ability differences. Even in this limit, the selection decision can still be important to the organization, because its post-selection productivity may be very sensitive to which agent is selected.<sup>3</sup> We show that in this limit, even though the first-stage outcome is almost entirely determined by random factors, *the organization continues to use a strictly positive bias in favor of the first-stage winner*. Moreover, although the favorite’s and the underdog’s efforts converge, limiting bias depends on the difference in their rates of convergence, underlining that bias constitutes the outcome of a strategic interaction.

Our results in Section 3 have two implications for the importance of early luck for individual success. First, a direct consequence of equilibrium bias remaining positive in the limit is that *early-career luck is made persistent*: first-stage winners are considerably more likely to be selected than first-stage losers, even when winning the first stage is almost entirely uninformative about ability. This persistence is thus an integral part of optimally-designed talent-selection processes. Second, under a mild (sufficient) condition on the distribution of noise, an even stronger result holds. Specifically, if the noise distribution is normal or thinner-tailed, then *early-career luck is made dominant*: an agent is more likely to be selected after being lucky in the first stage and unlucky in the second than when the ordering of the lucky and the unlucky spells is reversed, even when the noise distribution is identical across stages. This result is reminiscent of the observation, common among sociologists, that “first impressions matter most” (Asch, 1946). While typically rationalized as a consequence of bounded rationality (e.g. primacy effect, Murock, 1962) or behavioral biases (e.g. confirmation bias, Rabin and Schrag, 1999), our theory explains the emergence of this pattern in the context of well-designed hiring and promotion processes. Taken together, these results rationalize the significant impact of early-career luck on subsequent success that is illustrated by our motivating examples.

In Section 4, we analyze the case where agents not only differ in their ability but are heterogeneous in their identity. Societal luck endows one of the agents with an “advanta-

---

<sup>3</sup>Relatedly, skills that are highly valuable at the post-selection stage—such as a manager’s social skills (Hansen et al., 2021) or his ability to match subordinates with the right tasks (Minni, 2025)—might have only little impact on pre-selection performance.

geous” identity, which translates into a commonly known additive advantage augmenting his first-stage performance. Agents’ identities are assumed to be observable by the organization and independent of their abilities. Our focus here is on the implications of organizational learning for the *persistence of societal luck*, defined as the probability that the initially advantaged agent is ultimately selected. We first show that, irrespective of whether the organization’s second-stage bias favoring the first-stage winner can be identity-dependent (ID) or is required to be identity-independent (II), the disadvantaged agent reduces his exogenous disadvantage by exerting higher first-stage effort than his rival; nevertheless, the advantaged agent maintains a net advantage in the first stage. For any *given* net advantage, the organization’s optimal II-bias is larger (smaller) than the optimal ID-bias favoring the advantaged (disadvantaged) agent after a first-stage win, because winning the first stage with an advantage (disadvantage) is a weaker (stronger) signal about the agent’s ability.

One might be tempted to conclude from the above that allowing bias to be identity-dependent is *doubly beneficial*, as it not only improves selective efficiency, but also reduces the persistence of societal luck by “penalizing” agents for the advantages derived from their identity. However, such a conclusion would be premature as it ignores agents’ effort responses: A move from II bias to ID biases could have the opposite effect if it muted disadvantaged agents’ incentives to partially compensate for their disadvantage with higher efforts. A key finding is that the incentive effects of making biases identity-dependent actually *reinforce* the learning benefit of “penalizing” identity-advantaged agents, as long as agents’ heterogeneity is limited, so the shift to ID biases does in fact reduce the persistence of societal luck. We then show that this effort response reduces persistence by so much that it even reduces the utility difference between the advantaged and the disadvantaged agent, suggesting that implementing identity-dependent biases through gender-specific mentoring or grants accounting for socioeconomic background can be effective instruments for affirmative action.

Finally, in Section 5, we derive testable predictions by analyzing how the persistence of early-career luck and societal luck vary with workers’ information about their relative abilities, the coarseness of performance measurement, and the intensity of competition for selection. We show that both types of luck are more persistent in settings where agents are well informed about their relative abilities and where performance measurement is ordinal rather than cardinal. To the extent that team production improves workers’ information and that performance in managerial roles is frequently ranked rather than quantified (Lazear, 2000), luck should be especially persistent in jobs involving teamwork and management duties. Given the growing prevalence of such jobs (Deming, 2021), our theory

thus suggests that initial luck has become increasingly important for individual success. Increased competition for the top jobs and for positions in the “top” firms—brought about by the increasing prevalence of “superstar” features of the economy (Rosen, 1981; Autor et al., 2020)—further amplifies the persistence of early-career luck but dampens the persistence of identity-based advantages.

*Related literature.* The relevance of initial luck for long-term success has been documented in settings ranging from economics (Bedard and Dhuey, 2006; Du et al., 2012; Oreopoulos et al., 2012) to sociology (Cole et al., 1981; Ciocca-eller, 2023), sports (Deaner et al., 2013), law (Bagues and Esteve-Volart, 2010), and finance (Cong and Xiao, 2022; Nanda et al., 2020). We identify a mechanism rationalizing the persistence of luck as an institutional feature of a meritocratic society aiming to select the best.

To analyze how institutional design shapes the relevance of initial luck for long-run success, we combine elements from Meyer’s (1991) model of organizational learning and Lazear and Rosen’s (1981) model of promotion tournaments. Meyer (1991) showed that, to make optimal use of coarse performance information, organizations will bias promotion decisions in favor of early strong performers, and she emphasized that this practice increases the persistent effects of ability differences. Here, to focus on how organizational talent selection affects the persistence of early-career *luck*, as distinct from the persistent effects of ability differences, we ask whether, and if so, how much, bias should be used in a setting where early performance becomes *uninformative* about agents’ abilities. These questions were not posed in Meyer (1991). We show that it is always optimal to make early-career luck persistent, even when performance becomes uninformative about abilities, and under a mild additional assumption, early luck matters more for careers than later luck. A second significant difference relative to Meyer (1991) is that our analysis incorporates agents’ effort choices, using a dynamic version of Lazear and Rosen’s (1981) model of promotion incentives. Efforts are central because they affect different types of luck differently: promotion incentives amplify the persistence of early-career luck but attenuate that of societal luck, particularly when bias can depend on agents’ identities.

In focusing on an organization’s problem of identifying and promoting its most able agents, we contribute to the literature on organizational learning. The seminal studies by Farber and Gibbons (1996) and Gibbons and Waldman (1999) have argued that firms’ learning about workers’ productivity is key for explaining wage and promotion dynamics. A robust empirical finding is that early raises, either in wages or in position, increase the probability of later promotions. Whether this correlation is caused by workers’ inherent productivity differentials (as suggested by the seminal studies) or by a “fast-track effect”

(as suggested here) is controversial, with U.S. evidence in favor of the former (Belzil and Bognanno, 2008) and Japanese evidence pointing towards the latter (Ariga et al., 1999).

The special relevance of early performance for careers is underlined by Lange’s (2007) finding that “employers learn fast”, which lends empirical support to our result that “first impressions matter most”.<sup>4</sup> Pastorino (2024) supports this view by documenting firms’ tendency to assign newly employed managers to tasks that are particularly informative about their abilities. According to our theory, such task assignments increase the persistence of early-career luck even further because they increase the optimal bias.

Central to our theory is a mechanism—selection with the help of biases—that augments the relevance of initial performance for final success. Other mechanisms with similar effects exist in recent literature investigating the detailed process through which employers learn about workers’ productivity.<sup>5</sup> For instance, when organizational learning is viewed as a *bandit problem*, negative experiences can terminate an employer’s hiring from a group of potential employees or her task assignment to a specific worker. This can lead to long-run disadvantages for groups whose productivity is relatively undiscovered, e.g. minorities (Lepage, 2024), and to persistent discrimination for subgroups of workers in careers which are exposed to “bad news”, e.g. surgeons (Bardhi et al., 2025; Durandard, 2023). Alternatively, when employers rely on the evaluations a worker obtained from previous employers, beliefs about other employers’ information or preferences start to play a role. Focusing on this *social learning* component, Bohren et al. (2019) show that discrimination will be persistent, i.e. negative discrimination will never be reversed, unless it is belief-based and some employers have misspecified priors.<sup>6</sup>

Finally, other papers have also explored the implications of the observation that good performance in the face of a disadvantage is a particularly strong signal about an agent’s high ability—see the theoretical analysis of Fryer (2007) and the field experiment of Bohren et al. (2019). This observation also links our analysis to the focus in Sethi and Somanathan (2023) and Bohren et al. (2025) on systemic discrimination. In particular, Sethi and Somanathan (2023) argue that expanding the representation of disadvantaged groups in hiring, e.g. through exploration-prone algorithms (Li et al., 2025), can be *necessary* for selective efficiency. Unlike Sethi and Somanathan (2023) and Bohren et al. (2025), however, we highlight (in Sections 4 and 5) that the effort responses to identity-based policies amplify their beneficial effects with respect to both learning and persistence.

---

<sup>4</sup>Using Armed Forces Qualification Test scores as measures of unobserved ability, Lange (2007) finds it takes only 3 years for employers’ expectation error about workers’ productivity to fall by one-half.

<sup>5</sup>See Onuchic (2025) for an excellent survey of this literature.

<sup>6</sup>Gill and Prowse (2014) experimentally document a different mechanism by which initial performance influences final success, namely, the psychological impact of early wins and losses on subsequent effort choices. They find that this impact differs systematically between men and women.



## 2. MODEL

We consider an organization consisting of a risk-neutral principal and two agents  $i \in \{A, B\}$  with heterogeneous abilities  $a_i \in \mathfrak{R}$ . The difference in abilities or *heterogeneity* is given by  $h > 0$ , i.e.  $\Delta a \equiv a_A - a_B \in \{-h, h\}$ . The principal observes the agents' relative performance during two stages,  $t \in \{1, 2\}$ . After the second stage, the principal needs to select one of the agents for a higher-level task whose payoff to the principal is increasing in the selected agent's ability.

Agent  $i$ 's performance at stage  $t$ ,  $x_{i,t} \in \mathfrak{R}$ , is the sum of three elements: the agent's time-invariant ability  $a_i$ ; the agent's private choice of effort  $e_{i,t} \geq 0$ ; and a time-varying random component  $\epsilon_{i,t} \in \mathfrak{R}$ . That is,

$$x_{i,t} \equiv a_i + e_{i,t} + \epsilon_{i,t}.$$

*Information and choices.* The principal and the agents share a common prior belief on  $\Delta a$ ,  $q \equiv \mathbb{P}(\Delta a = h) \geq \frac{1}{2}$ , but for the principal, agents  $A$  and  $B$  are indistinguishable. If  $q = \frac{1}{2}$ , the agents are as uninformed as the principal, while if  $q > \frac{1}{2}$ , the agents have information about their relative abilities, with both agents believing that agent  $A$ , the *favorite*, is more likely to have higher ability than agent  $B$ , the *underdog*.<sup>7</sup>

The principal and the agents can observe only the *ranking* of the agents' performances after the first stage. In the second stage, the principal may costlessly and publicly assign a bias  $\beta \in \mathfrak{R}$  to the winner of the first stage. If  $\beta > 0$ , the bias increments the winner's second-stage performance, and we say that the bias "favors" the first-stage winner, whereas if  $\beta < 0$ , it reduces his second-stage performance. Having won the first stage, agent  $i$  is then identified as the winner of the second stage if  $x_{i,2} + \beta > x_{j,2}$ .

The principal chooses the size of bias  $\beta$  and the selection rule to maximize *selective efficiency*,  $S$ , defined as the probability that the more able agent is selected. The agents exert privately-observed efforts  $e_{i,t}$  in each stage to maximize the probability of being selected minus the effort costs.<sup>8</sup> The value of being selected is the same for both agents and normalized to 1. The cost-of-effort functions are quadratic  $C_t(e_{i,t}) = \frac{c_t}{2} e_{i,t}^2$  with

<sup>7</sup>Virtually all of the employer learning models reviewed in the Introduction analyze only the case where workers are ignorant about their own ability, and hence correspond to the case  $q = \frac{1}{2}$ .

<sup>8</sup>In our analysis, agents' efforts affect the principal's payoff (selective efficiency) only indirectly, through their impact on the informativeness of the outcomes in each stage. Suppose the principal did care directly about agents' efforts. In the absence of bias, agents would exert zero effort in the first stage. Moreover, we can show that for  $q = \frac{1}{2}$ , the bias that maximizes selective efficiency in the limit  $h \rightarrow 0$  ( $\beta_0^*$  characterized in Proposition 1) also maximizes aggregate effort. More generally, if the principal's objective were a weighted combination of selective efficiency and aggregate effort, our results would not be affected qualitatively, at least when the weight on effort was not too high.



$c_t > 0, t = 1, 2$ .<sup>9</sup> Thus effort costs may differ across stages but not across agents.

*Noise.* The distribution of the difference in the individual noise terms,  $\Delta\epsilon_t \equiv \epsilon_{A,t} - \epsilon_{B,t}$ , is a key primitive in our model because outcomes depend only on performance *differentials*. We assume that  $\Delta\epsilon_t$  are identically and independently distributed across stages and denote the corresponding support by  $[-z, z]$  (where  $z$  may be infinite), the cumulative distribution function by  $G$ , and its density by  $g$ . We make the following distributional assumptions:

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

$$L(y) \equiv -\frac{g'(y)}{g(y)}. \quad (1)$$

Symmetry of  $g$ , coupled with identical effort costs for agents, ensures that the only exogenous asymmetry between agents is the difference in abilities. Log-concavity of  $g$  is equivalent to the monotone likelihood ratio property here and guarantees that the larger the performance differential  $\Delta x_t \equiv x_{A,t} - x_{B,t}$ , the higher the likelihood that  $A$ 's ability exceeds  $B$ 's. It also implies that  $L(y)$  defined above is increasing. Strict log-concavity makes all the implications strict. Together with conditions (iii) and (iv), it ensures that the principal's problem is well-behaved.

*Identity.* Our analysis distinguishes between two settings. While Section 3 restricts attention to the setting where agents have the same identity (e.g. gender, ethnicity, socioeconomic background), in Section 4 we allow agents to differ in this dimension, assuming there that one agent—whose identity is common knowledge—has a relative advantage over the other during the first stage.<sup>10</sup> More specifically, in Section 4 we assume that the first-stage performance of the advantaged agent  $i$  is increased by a known amount  $\alpha \geq 0$  to  $x_{i,1} + \alpha$ . Because only relative performance matters, an equivalent interpretation is that agent  $j \neq i$  is disadvantaged in that his first-stage performance is reduced to  $x_{j,1} - \alpha$ . In this section we distinguish between the case where the principal can condition the bias on the winner's identity and the case where she must set an identity-independent bias.

*Timing.* In the first stage, the agents choose efforts. Then, first-stage noise is realized, and both the principal and the agents observe who has higher first-stage performance. At

---

<sup>9</sup> We assume throughout the paper the existence and uniqueness of a pure-strategy equilibrium, which can be ensured by letting  $c_1$  and  $c_2$  be high enough. Related issues arise even in static tournament models, as has been understood since Lazear and Rosen (1981).

<sup>10</sup> Given our interest in how organizational selection affects the persistence of initial (dis-)advantages, the assumption that identity has no *direct* effect on second-stage performance is most conservative.

the start of the second stage, the principal chooses the bias. Agents observe the bias and then choose efforts. Second-stage noise is realized, and both the principal and the agents observe who is the second-stage winner. Finally, the principal selects one of the agents.

*Equilibrium.* The solution concept is perfect Bayesian equilibrium (PBE). In a PBE, (i) the effort choice by each agent at each stage is optimal given his conjectures about the effort choices of the other agent and about the principal’s bias and selection rule; (ii) the bias and the selection rule are optimal for the principal given her conjectures about the agents’ efforts; and (iii) the conjectures of all players are correct. It is without loss of generality to assume the principal selects the second-stage winner, because the selection of the first-stage winner can be implemented by setting an infinite bias.

*Interpretation of heterogeneity.* The key parameter of our model,  $h > 0$ , captures the degree of agents’ heterogeneity in abilities. It also has a broader interpretation as the *ratio* of agents’ heterogeneity to the scale of noise.<sup>11</sup> To highlight the impact of early-career luck on final success, as distinct from the impact of differences in ability and effort, much of the analysis in Section 3 will focus on the setting in which  $h$  is very small: Here the scale of noise is large *relative* to the agents’ heterogeneity and, as we show, differences in agents’ efforts vanish.<sup>12</sup> Note that even in this environment, the selection decision may still be important to the principal, because the selected agent’s impact in the higher-level task may be highly relevant and very sensitive to ability.

### 3. EARLY-CAREER LUCK

Does early-career luck have a persistent effect on final success? And if so, how does the impact of early-career luck compare to that of luck at later stages of careers? In this section, we analyze the *interplay* between agents aiming to be selected and a principal trying to select the best (Section 3.1) to derive our main results about the persistence (Section 3.2) and the dominance (Section 3.3) of early-career luck. We abstract here from advantages or disadvantages agents derive from differences in their identities. To isolate

---

<sup>11</sup>To see this, introduce a scaling transformation  $\Delta\epsilon_t \rightarrow \sigma\Delta\epsilon_t$ , 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  $G(\frac{H}{\sigma})$  is the probability that, when the first-stage effort differential is zero, the more able agent wins the first stage. It depends on  $H$  and  $\sigma$  only through the *heterogeneity-to-noise ratio*  $h \equiv \frac{H}{\sigma}$ .

<sup>12</sup>When  $q = \frac{1}{2}$ , so the common prior belief about the difference in abilities is symmetric about 0, all of the main results in Section 3 (Proposition 1 and Corollaries 1 and 2) generalize to an arbitrary symmetric distribution for  $\Delta a$ ; for this generalization, we let the heterogeneity parameter  $h$  be a scaling factor for the distribution and focus on the limit as  $h \rightarrow 0$ , so the distribution becomes increasingly concentrated around 0.

the long-term effects of luck, we focus on the limiting case where noise swamps ability and effort differences in determining outcomes ( $h \rightarrow 0$ ).

### 3.1. Equilibrium

We begin by considering agents' effort choices. If agents anticipate that the principal will, in the second stage, employ a bias in favor of the first-stage winner ( $\beta > 0$ ), they will have incentives to exert effort to win the first stage. Because effort and ability are *substitutes* for agents' performance, the underdog might use effort to compensate for lack of ability. Yet, our first lemma shows the opposite.

**Lemma 1 (Effort differential)** *For any anticipated choice of bias, agents choose identical efforts in the second stage, but in the first stage, the favorite exerts a larger effort than the underdog, i.e.*

$$\Delta e_1^*(\beta; h) \equiv e_{A,1}^*(\beta; h) - e_{B,1}^*(\beta; h) \geq 0, \quad (2)$$

*with strict inequality if and only if  $\beta > 0$  and  $q > \frac{1}{2}$ .*

The agents work equally hard in the second stage, because the marginal benefit of effort is the same for both, despite the asymmetry arising from the bias and the agents' belief updating. This is because the agents' value of being selected, the marginal impact of their efforts on performance, and the pivotal realization of  $\Delta\epsilon_2$  deciding between winning and losing are all identical for  $A$  and  $B$  (cf. [Lazear and Rosen, 1981](#)).

To understand why, in the first stage, the favorite works harder than the underdog, note that because exactly one agent will be selected after the second stage, the “rewards” of winning the first stage arising from the increased probability of being selected are the same for the two agents. However, the level of second-stage effort that agents exert, and hence their effort cost, depends on the first-stage outcome. To see this most clearly, suppose  $q = 1$ , so agents know with certainty that agent  $A$  is more able. If agent  $A$  wins the first stage, bias will reinforce the agents' ability difference, and the pivotal noise realization  $\Delta\epsilon_2 = h + \beta$  will determine second-stage efforts via  $C_2'(e_{A,2}^*) = g(h + \beta) = C_2'(e_{B,2}^*)$ . If, instead, agent  $A$  loses the first stage, then bias will mitigate the agents' ability difference, so it is the pivotal realization  $\Delta\epsilon_2 = h - \beta$  that determines second-stage efforts via  $C_2'(e_{A,2}^*) = g(h - \beta) = C_2'(e_{B,2}^*)$ . Because  $g(h + \beta) < g(h - \beta)$  by unimodality of  $g$  (implied by log-concavity), agent  $A$  faces lower second-stage effort costs after winning the first stage than after losing, giving agent  $A$  a “cost-saving incentive” to win the first stage. For agent  $B$ , the argument is reversed, because bias mitigates agents' heterogeneity when  $B$  wins but reinforces it when  $B$  loses, thereby reducing  $B$ 's incentive to win the first stage.

Next, consider the principal's choice of bias. The principal chooses bias in the second stage, without knowing which agent was the favorite and without having observed first-stage efforts. Let  $\Delta e_1$  be the principal's conjecture about the first-stage effort differential. This effort differential matters to the principal because it influences the probability

$$\kappa(\Delta e_1; h) = qG(h + \Delta e_1) + (1 - q)G(h - \Delta e_1) \quad (3)$$

that the more able agent wins the first stage. Focusing on  $\Delta e_1 \geq 0$  (Lemma 1), note that the more able agent is strictly more likely to win the first stage, i.e.  $\kappa(\Delta e_1; h) > \frac{1}{2}$ .

The principal chooses bias  $\beta$  to maximize selective efficiency,  $S$ , defined as the probability that the more able agent is selected as the winner of the second stage. We have

$$S(\beta, \Delta e_1; h) = \kappa(\Delta e_1; h)G(h + \beta) + [1 - \kappa(\Delta e_1; h)]G(h - \beta). \quad (4)$$

Selective efficiency is thus the sum of two parts: the probability that the better agent wins the first stage and then wins the second with the bias in his favor, plus the probability that the better agent loses the first stage but then wins the second, despite being handicapped by the bias. From the FOC corresponding to (4) we obtain the following:

**Lemma 2 (Bias)** *For any conjectured first-stage effort differential  $\Delta e_1 \geq 0$ , the principal's optimal bias  $\beta^*(\Delta e_1; h)$  is strictly positive and given by the unique solution to*

$$\frac{\kappa(\Delta e_1; h)}{1 - \kappa(\Delta e_1; h)} = \frac{g(h - \beta)}{g(h + \beta)}. \quad (5)$$

The FOC (5) for optimal bias  $\beta^*(\Delta e_1; h)$  equates two likelihood ratios, one corresponding to each of the two stages. The ratio on the LHS measures the informativeness of a first-stage win by the relative likelihood that this win is achieved by the more able rather than the less able agent. The ratio on the RHS measures the informativeness of an agent's achieving a hypothetical second-stage draw, when disadvantaged by bias  $\beta$ , by the relative likelihood that such a draw is achieved by the more able rather than the less able agent ( $g(h + \beta) = g(-h - \beta)$  by symmetry of  $g$ ).

Lemma 2 shows that the principal's choice of bias strikes a balance between the informativeness of the ordinal first-stage ranking—an unbiased win—and the informativeness of the marginal second-stage outcome—a draw achieved despite being handicapped by bias. In other words, the principal's optimal bias is such that, if she were to observe a draw in the second stage, she would be indifferent about which agent to select.<sup>13</sup>

---

<sup>13</sup>When  $q = \frac{1}{2}$ , so agents are as uninformed about their relative abilities as the principal, Lemma 1

Lemma 2 is silent about the dependence of the optimal bias on  $h$ , the heterogeneity in agents' abilities. An increase in  $h$  raises *both* the informativeness of an unbiased first-stage win *and* the informativeness of a second-stage draw against any positive bias. As a consequence, the dependence of the optimal bias on  $h$  can be complex, even in the simplest case where  $q = \frac{1}{2}$ , so  $\Delta e_1 = 0$  (Lemma 1). This is illustrated in Figure 1 for the case where the noise distribution is in the exponential power family.

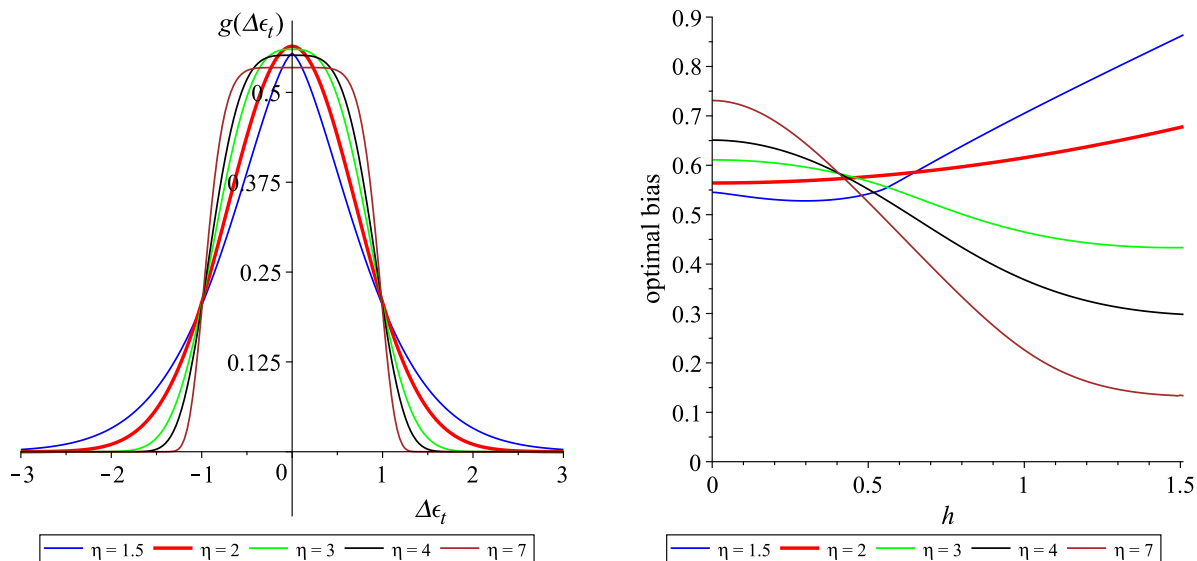


Figure 1: **Example distributions of noise and optimal bias.** The left panel depicts the density functions when noise follows an exponential power distribution  $g(\Delta\epsilon_t; \eta) = \frac{\eta}{2\Gamma(\frac{1}{\eta})} \exp(-|\Delta\epsilon_t|^\eta)$ , with mean zero and shape parameter  $\eta > 1$  satisfying Assumption 1. For  $\eta = 2$ ,  $g$  is normal, for  $\eta \rightarrow 1$ ,  $g$  approaches Laplace, and for  $\eta \rightarrow \infty$ ,  $g$  approaches uniform. The right panel plots the principal's choice of bias  $\beta^*(\Delta e_1; h)$  solving (5) for  $q = \frac{1}{2}$  and  $\Delta e_1 = 0$ .

Our main interest is the persistence of early-career luck. To study it, we combine Lemmas 1 and 2 to analyze the probability, in equilibrium, that the first-stage winner is ultimately selected. To focus on the persistent effects of luck, as distinct from those of ability differences, we examine the limit as noise swamps ability differences ( $h \rightarrow 0$ ).

In our setting, an equilibrium  $(\Delta e_1^*(h), \beta^*(h))$  consists of a first-stage effort differential and a bias that are mutual “best responses” in that  $\Delta e_1^*(h) = \Delta e_1^*(\beta^*(h); h)$  and  $\beta^*(h) = \beta^*(\Delta e_1^*(h); h)$ . Lemmas 1 and 2 together imply that for any  $h > 0$ , in any equilibrium,  $\Delta e_1^*(h) \geq 0$  and  $\beta^*(h) > 0$ . As the heterogeneity  $h$  goes to 0, the first-stage effort differential vanishes, for any  $q$  and any anticipated bias, and hence from (3) the

shows that  $\Delta e_1 = 0$ . In this case,  $\kappa(\Delta e_1; h)$  simplifies to  $G(h)$ , and (5) reduces to the FOC determining optimal bias in the pure learning model of Meyer (1991).

informativeness of a first-stage win approaches zero (the LHS of (5) converges to one). It might seem that, as  $h \rightarrow 0$ , the principal's choice of bias should converge to zero. Why bias the second stage in favor of an agent who is no more likely than his rival to be the better agent? Figure 1 (right panel) shows that this intuition is misleading. It ignores the fact that, as  $h \rightarrow 0$ , the informativeness of a second-stage draw, against any level of bias, *also* approaches zero (the RHS of (5) also converges to one). Hence, the limiting value of optimal bias must equate not just the *levels* of informativeness (the likelihood ratios in (5)), but the *rates* at which these likelihood ratios change with  $h$  as  $h \rightarrow 0$ .

In particular, since the informativeness of a first-stage win changes with  $h$  not only directly but also through the *change* in the agents' effort differential, the limiting value of optimal bias depends on the following limiting derivative, derived in the Appendix:

$$\frac{\partial \Delta e_1^*(\beta; 0)}{\partial h} \equiv \lim_{h \rightarrow 0} \frac{\partial \Delta e_1^*(\beta; h)}{\partial h} = -4g(0)g(\beta)g'(\beta)\frac{1}{c_1 c_2}(2q - 1) \geq 0. \quad (6)$$

The proposition below characterizes the limiting value of equilibrium bias as  $h \rightarrow 0$ .

**Proposition 1 (Equilibrium)** *In the limit as noise swamps ability differences ( $h \rightarrow 0$ ), the stage-1 effort differential vanishes ( $\Delta e_1^*(h) \rightarrow 0$ ), but equilibrium bias converges to a strictly positive value  $\beta_0^* \equiv \lim_{h \rightarrow 0} \beta^*(h)$  given by the unique solution to*

$$2g(0) \left[ 1 + (2q - 1) \frac{\partial \Delta e_1^*(\beta; 0)}{\partial h} \right] = L(\beta), \quad (7)$$

where  $L(\cdot)$  is defined in (1).

The fixed-point equation (7) determines equilibrium bias in the limit by equating the *rates* at which the informativeness of each stage tends to zero as  $h$  gets small, given the rate (6) at which agents' stage-one effort differential tends to zero in this limit.

Two observations are important. First, if bias were set at zero, then a second-stage draw would be uninformative about relative abilities *for any* value of  $h$ , so the RHS of (7), the rate at which second-stage informativeness rises with  $h$  as  $h$  rises from 0, would be  $L(\beta) = L(0) = 0$ . In contrast, since the informativeness of a first-stage win rises with  $h$  as  $h$  rises from 0, the LHS of (7) is strictly positive, even when (6) is 0. Hence, equilibrium bias must remain strictly positive as  $h \rightarrow 0$  because, unless first-stage losers are disadvantaged relative to first-stage winners even when ability differences are negligible, the informativeness of a second-stage draw cannot rise at the same rate as the informativeness of a first-stage win when ability differences start to matter.

Second, (7) and (6) together show that agents' competition to be selected amplifies equilibrium bias, as long as agents are better informed than the principal. This is because,

for all  $q > \frac{1}{2}$  and  $\beta > 0$ , the favorite agent's stage-1 effort rises faster than the underdog's when  $h$  increases from 0, i.e.  $\frac{\partial \Delta e_1^*(\beta; 0)}{\partial h} > 0$ . Consequently, the agents' strategic choice of efforts boosts the rate of increase in the informativeness of the first-stage outcome about ability (the LHS of (7)), thereby inducing the principal to choose a larger bias.

In summary, Proposition 1 reveals that bias favoring the stage-1 winner remains a feature of organizational selection in equilibrium, even in situations where ordinal rankings convey arbitrarily little information about agents' relative abilities.

### 3.2. Persistence

Our focus is the persistence of outcomes induced by the interaction between the principal's pursuit of selective efficiency and the agents' desire to be selected. We define the *persistence*  $P$  of the selection process as the probability that, in equilibrium  $(\Delta e_1^*(h), \beta^*(h))$ , the agent who performed better in the first stage is selected after the second stage:

$$P(\Delta e_1^*(h), \beta^*(h); h) = \kappa(\Delta e_1^*; h)G(h + \beta^*) + [1 - \kappa(\Delta e_1^*; h)][1 - G(h - \beta^*)]. \quad (8)$$

Persistence, like selective efficiency, is the sum of two parts: the probability that the better agent wins the first stage and then wins the second with the bias in his favor, plus the analogous probability for the worse agent. To isolate the persistent effects of early-career luck from those of ability differences, we focus on the limit as  $h \rightarrow 0$  in making the following definition:

**Definition 1** *Early-career luck has a persistent effect on final success iff*

$$P_0^* \equiv \lim_{h \rightarrow 0} P(\Delta e_1^*(h), \beta^*(h); h) > \frac{1}{2}. \quad (9)$$

Definition 1 says that early-career luck has a persistent effect if the first-stage winner is more likely to achieve final success than the first-stage loser, even when the first-stage outcome is determined almost entirely by random factors. The following result is an immediate consequence of the equilibrium characterization in Proposition 1:

**Corollary 1 (Persistence)** *Organizational selection induces early-career luck to have a persistent effect on final success, i.e.  $P_0^* = G(\beta_0^*) > \frac{1}{2}$ .*

Corollary 1 identifies the persistence of early-career luck as an inseparable consequence of the strategic interaction between agents aiming to be selected and a principal trying to select the best. Luck is made persistent because the principal optimally uses bias for selection even as noise swamps ability differences, i.e.  $P_0^* > \frac{1}{2} \Leftrightarrow \beta_0^* > 0$ . Moreover,



for  $q > \frac{1}{2}$ , the agents' effort responses amplify the limiting equilibrium bias and hence the persistence of early-career luck. The persistence of luck can thus be understood as a consequence of the principal's attempt to make optimal use of the little information contained in noisy performance rankings and to capitalize on the agents' information about their relative abilities.

### 3.3. Dominance

We have shown that organizational selection induces early-career luck to have a persistent effect on final success. In this subsection, we show when an even stronger conclusion holds, namely that organizational selection makes the impact of early-career luck *dominate* the impact of luck during the later stages of an agent's career. The dominance of early luck over late luck is reminiscent of the colloquial proverb that "first impressions matter most". We prove that a sufficient condition for the dominance of early-career luck is that realizations of extremely good or bad luck are (weakly) less likely than when noise is normally distributed.

To evaluate the *relative* importance of early-career versus late-career luck, we compare, in the limit as noise swamps heterogeneity, an agent's probability of being selected when a favorable noise realization is followed by an unfavorable noise realization with the corresponding probability when the sequence of these two events is reversed. Since  $\lim_{h \rightarrow 0} \Delta e_1^*(h) = 0$ , the ratio of these two selection probabilities is the same whether we are focusing on the favorite or on the underdog. Focusing for concreteness on the favorite (agent  $A$ ), and recalling that  $\Delta \epsilon_t \equiv \epsilon_{A,t} - \epsilon_{B,t}$ , we thus evaluate

$$D_0^* \equiv \lim_{h \rightarrow 0} \frac{\mathbb{P}(\text{select } A, \Delta x_1 > 0, \Delta x_2 < 0)}{\mathbb{P}(\text{select } A, \Delta x_1 < 0, \Delta x_2 > 0)} = \frac{\mathbb{P}(\text{select } A, \Delta \epsilon_1 > 0, \Delta \epsilon_2 < 0)}{\mathbb{P}(\text{select } A, \Delta \epsilon_1 < 0, \Delta \epsilon_2 > 0)}. \quad (10)$$

**Definition 2** *Early-career luck has a dominant effect on final success iff*

$$D_0^* > 1. \quad (11)$$

The size of  $D_0^*$  depends on the limiting equilibrium bias  $\beta_0^*$  characterized in Proposition 1. The numerator of the rightmost expression in (10) is the probability that the realizations of early- and late-career luck  $(\Delta \epsilon_1, \Delta \epsilon_2)$  fall into the red region in Figure 2 (left panel)—here the agent is selected after being lucky in the first stage ( $\Delta \epsilon_1 > 0$ ), and only mildly unlucky in the second stage ( $-\beta_0^* < \Delta \epsilon_2 < 0$ ) when favored by the bias. The denominator of that expression is the probability that  $(\Delta \epsilon_1, \Delta \epsilon_2)$  fall into the blue region in Figure 2 (left panel)—here the agent is selected after being unlucky in the first stage ( $\Delta \epsilon_1 < 0$ ),

and sufficiently lucky in the second stage ( $\Delta\epsilon_2 > \beta_0^*$ ) to overcome the bias against him.

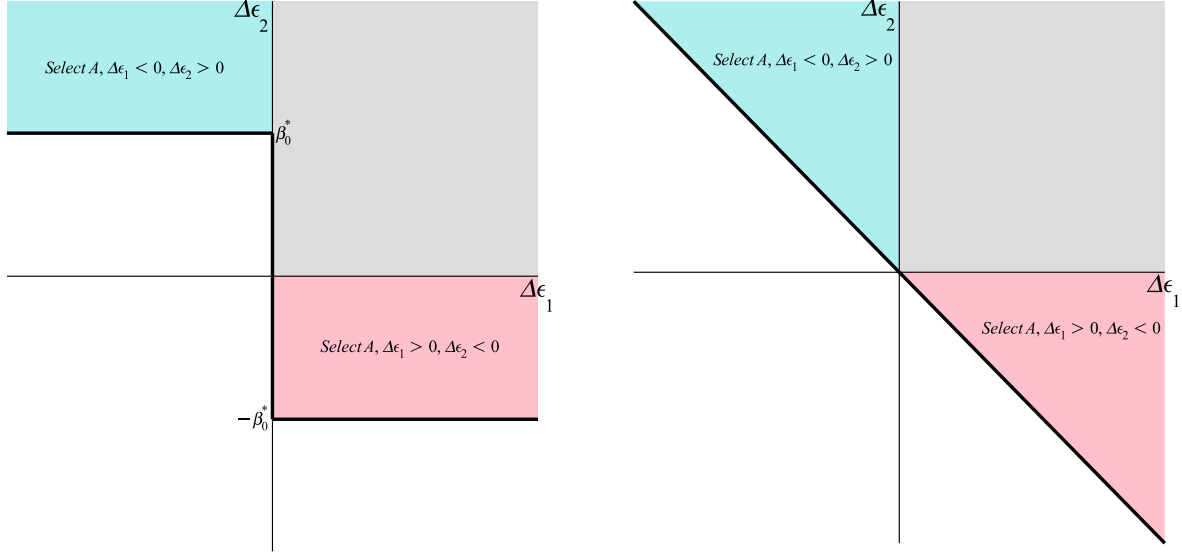


Figure 2: **Dominance.** Selection in the equilibrium described by Proposition 1 (left panel) and in the hypothetical selection process based on winning margins (right panel). The bold line separates the area where agent A is selected (colored) from the area where agent B is selected (white). Both processes select agent A after two strokes of good luck (grey). Selection differs between the processes when good luck is followed by bad luck (red) and when these events are reversed (blue).

Using Figure 2 (left panel), these probabilities are readily determined, and we get:

$$D_0^* = \frac{G(\beta_0^*) - \frac{1}{2}}{1 - G(\beta_0^*)}. \quad (12)$$

Given that  $G(\beta_0^*) = P_0^*$ , it follows that  $D_0^* > 1$  if and only if  $P_0^* > \frac{3}{4}$ . Clearly, therefore, the condition for the dominance of early-career luck is stronger than for the persistence of early-career luck, which Corollary 1 shows always holds.

**Corollary 2 (Dominance)** *Suppose that realizations of extremely good luck and extremely bad luck are relatively rare, in the sense that  $L(y)$  is convex for  $y \geq 0$ . Then organizational selection induces early-career luck to have a dominant effect on final success, i.e.  $D_0^* > 1$ .*

To understand why  $D_0^* > 1$  if and only if  $P_0^* > \frac{3}{4}$ , consider the right panel of Figure 2. It shows the limiting outcome, as  $h \rightarrow 0$ , of a *hypothetical* selection process which selects agents mechanically—on the basis of  $\Delta x_1 = \Delta a + \Delta\epsilon_1$  and  $\Delta x_2 = \Delta a + \Delta\epsilon_2$ —by selecting the agent with the larger “winning margin”  $|\Delta x_t|$  when  $\Delta x_1$  and  $\Delta x_2$  differ in sign, while selecting the agent who wins twice if  $\Delta x_1$  and  $\Delta x_2$  have the same sign. The figure shows

that, in contrast to our equilibrium, this hypothetical selection process forces early- and late-career luck to have a *symmetric* impact on selection. Given this symmetry, it follows that such a process selects the first-stage winner with probability  $P_0^{hyp} = \frac{3}{4}$ , because even after a second-stage loss, the first-stage winner is selected with probability  $\frac{1}{2}$ .

Now, since basing selection, in the limit as  $h \rightarrow 0$ , on the sign of  $\Delta\epsilon_1 + \Delta\epsilon_2$  is equivalent to using a bias of  $|\Delta\epsilon_1|$ , it follows, by analogy with  $P_0^* = G(\beta_0^*)$ , that  $P_0^{hyp} = \mathbb{E}[G(|\Delta\epsilon_1|)]$ . Moreover, since  $G(y)$  is strictly concave for  $y \geq 0$  (by Assumption 1),  $\mathbb{E}[G(|\Delta\epsilon_1|)] < G(\mathbb{E}[|\Delta\epsilon_1|])$ . Hence, a sufficient condition for the dominance of early-career luck is that equilibrium bias  $\beta_0^*$  is at least as large as the expected noise-margin,  $\mathbb{E}[|\Delta\epsilon_1|]$ .

The condition in Corollary 2 that the increasing function  $L(y) \equiv \frac{-g'(y)}{g(y)}$  is convex for  $y \geq 0$  ensures that  $\beta_0^* \geq \mathbb{E}[|\Delta\epsilon_1|]$ . This can be seen as follows. Since  $|\Delta\epsilon_1|$  has density  $2g$  on  $[0, z]$ , and  $g(z) = 0$  by Assumption 1,

$$L(\mathbb{E}[|\Delta\epsilon_1|]) \leq \mathbb{E}[L(|\Delta\epsilon_1|)] = \int_0^z L(|\Delta\epsilon_1|) 2g(|\Delta\epsilon_1|) d|\Delta\epsilon_1| = 2g(0) \leq L(\beta_0^*), \quad (13)$$

where the first inequality follows from the convexity of  $L$  and the second from (6) and (7). If the distribution of noise is normal, then  $L$  is linear, so the first inequality is an equality, and if  $q = \frac{1}{2}$ , then the second inequality is also an equality, so for this case we have  $\beta_0^* = \mathbb{E}[|\Delta\epsilon_1|]$ . Distributions for which  $L$  is convex are those which are *thinner-tailed* than the normal distribution, or, more precisely, those that are more log-concave in the sense that  $\ln g$  is a concave transform of  $\ln \tilde{g}$ , for  $\tilde{g}$  normal. The discussion in the previous paragraph makes clear that the convexity of  $L$ , though sufficient for the dominance of early-career luck, is not necessary. For the exponential power family of distributions depicted in Figure 1,  $L$  is convex if and only if  $\eta \geq 2$ , while even for  $q = \frac{1}{2}$ , early-career luck is dominant for all  $\eta > 1.38$ . Thus, early-career luck dominates late-career luck even for some distributions that are thicker-tailed than the normal, for which  $\eta = 2$ .

To summarize, noise distributions that are normal or thinner-tailed are sufficient for the dominance of early-career luck, because they induce a principal who is restricted to ordinal information to implement a bias larger than the average bias she would choose if she could base the bias on the margin of victory. The restriction to rank-order information then destroys the balanced treatment that early-career luck and late-career luck would be given if winning margins were observable, resulting in early-career luck having the dominant effect on ultimate success.

## 4. SOCIETAL LUCK

Section 3 highlights the relevance of early-career luck for an individual’s long-term success. Early-career luck derives from the inherent noisiness of individual performance. In this section, we introduce “societal luck”, reflecting the possession of an advantageous identity, for example, a particular socioeconomic background, gender, race, or ethnic origin.

There is evidence showing that individuals with certain identities obtain advantages early in their career. For instance, differences in educational achievement of students attending colleges of *equal* selectivity can be traced to heterogeneous socioeconomic backgrounds (Ciocca-eller, 2023), and exposure to elite-educated parents stimulates application to elite education even when exposure is only indirect through peers (Cattan et al., 2025). These findings are especially worrying when initial advantages translate into long-lasting differences in outcomes, such as the 8% earnings gap between first-generation and continuing-generation college graduates in OECD countries (Fabbri and Pellizzari, 2025).

In this section, we assume that exactly one agent  $i \in \{A, B\}$  possesses an “advantageous” identity, which translates into a known additive advantage of size  $\alpha > 0$  that augments only his first-stage performance. Agents’ identities are assumed to be independent of their abilities and common knowledge among all players. Conceptually, the luck of having an advantageous identity is different from what we have considered so far in that individuals might condition their actions on it. More specifically, organizations might reward good performance with biases that depend on whether success was achieved with or without an exogenous advantage, and agents’ incentives to exert efforts might vary with their identity. We examine the long-run equilibrium consequences of identity-based advantages by focusing on the *persistence of societal luck*, defined as the probability that the initially advantaged agent is ultimately selected.

Our analysis in this section focuses on the case  $q = \frac{1}{2}$ . This allows us to separate the effect of the commonly observed societal advantage  $\alpha$  from that of agents’ information (as captured by  $q > \frac{1}{2}$ ).<sup>14</sup> Hence there is neither a favorite nor an underdog (in the sense of Section 3), and we assume, without loss of generality, that agent  $A$  receives the advantage  $\alpha$ . We distinguish two scenarios. Under *identity-dependent* (ID) biases, the principal can condition her second-stage bias  $\beta_i$  on the first-stage winner’s identity  $i \in \{A, B\}$ . In contrast, under *identity-independent* (II) bias, the principal is required to set  $\beta_A = \beta_B$ . II bias might be a consequence of legislation aimed at preventing discriminatory practices.<sup>15</sup>

---

<sup>14</sup>Section 5.1 analyzes the impact of  $q$  on the persistence of societal luck. Corollary 3 generalizes to arbitrary  $q$ , for small  $\alpha$ .

<sup>15</sup>Title VII of the 1964 Civil Rights Act declares as “an unlawful employment practice [...] to discriminate against any individual because of his race, color, religion, sex, or national origin in admission to, or employment in, any program established to provide apprenticeship or other training.”

Alternatively, there may be behavioral reasons why advantages are not accounted for, even when they are known to exist.<sup>16</sup> Our objective here is to determine how the persistence of societal luck differs between the equilibria in the regimes of ID and II biases.

In each scenario (ID or II), agents choose efforts optimally in response to the bias(es) they anticipate. For the by-now familiar reasons, second-stage efforts are identical across agents, so we can focus on the agents' first-stage effort differential  $\Delta e_1 = e_{A,1} - e_{B,1}$ . Suppose that under ID biases, agents' optimization results in  $\Delta e_1 = \Delta e_1^*(\beta_A, \beta_B)$ , and let  $\Delta e_1 = \Delta e_1^*(\beta)$  be the analogous notation under II bias. The principal's optimal choice of bias, in each regime, depends on the conjectured *net advantage*,  $\tilde{\alpha} = \alpha + \Delta e_1$ , of the advantaged agent. Denote the principal's optimal biases in the ID regime by  $\beta_A^*(\tilde{\alpha})$  and  $\beta_B^*(\tilde{\alpha})$ , and let  $\beta^*(\tilde{\alpha})$  be her optimal II bias. In the ID regime, an equilibrium is a combination of biases and net advantage  $(\beta_A^{ID}, \beta_B^{ID}, \tilde{\alpha}^{ID})$  that are mutual best responses, that is,  $\beta_A^{ID} = \beta_A^*(\tilde{\alpha}^{ID})$ ,  $\beta_B^{ID} = \beta_B^*(\tilde{\alpha}^{ID})$ , and  $\tilde{\alpha}^{ID} = \alpha + \Delta e_1^*(\beta_A^{ID}, \beta_B^{ID})$ . Similarly, in the II regime, an equilibrium  $(\beta^{II}, \tilde{\alpha}^{II})$  satisfies  $\beta^{II} = \beta^*(\tilde{\alpha}^{II})$  and  $\tilde{\alpha}^{II} = \alpha + \Delta e_1^*(\beta^{II})$ . We use  $\Delta e_1^{ID}$  and  $\Delta e_1^{II}$  to denote  $\Delta e_1^*(\beta_A^{ID}, \beta_B^{ID})$  and  $\Delta e_1^*(\beta^{II})$ , respectively.

For arbitrary values of biases  $\beta_A$  and  $\beta_B$ , and net advantage  $\tilde{\alpha}$ , selective efficiency is:

$$\begin{aligned} S(\beta_A, \beta_B, \tilde{\alpha}) &= \frac{1}{2}[G(h + \tilde{\alpha})G(h + \beta_A) + G(-h - \tilde{\alpha})G(h - \beta_B)] \\ &+ \frac{1}{2}[G(h - \tilde{\alpha})G(h + \beta_B) + G(-h + \tilde{\alpha})G(h - \beta_A)]. \end{aligned} \quad (14)$$

The terms in the first (respectively, second) square brackets are the probability that the better agent is selected, conditional on being advantaged (respectively, disadvantaged). The principal's optimal ID biases  $\beta_A^*$  and  $\beta_B^*$  solve the FOCs

$$\frac{G(h + \tilde{\alpha})}{G(-h + \tilde{\alpha})} = \frac{g(h - \beta_A^*)}{g(h + \beta_A^*)} \quad \text{and} \quad \frac{G(h - \tilde{\alpha})}{G(-h - \tilde{\alpha})} = \frac{g(h - \beta_B^*)}{g(h + \beta_B^*)}. \quad (15)$$

The principal's optimal II bias  $\beta^*$  solves  $\frac{\partial S}{\partial \beta} = 0$  under the constraint that  $\beta_A = \beta_B = \beta$ :

$$\frac{G(h + \tilde{\alpha}) + G(h - \tilde{\alpha})}{G(-h + \tilde{\alpha}) + G(-h - \tilde{\alpha})} = \frac{g(h - \beta^*)}{g(h + \beta^*)}. \quad (16)$$

Equations (15) and (16) are the analogues of (5). They show that in the II regime, the principal is restricted to set bias to match the “average” informativeness of a first-stage win, whereas ID biases allow the principal to adapt to whether such a win was achieved

---

<sup>16</sup>Exley and Nielsen (2024) document that evaluators correctly expect women to be less confident than men in the assessment of their own abilities but fail themselves to account for this gender gap in their evaluations. We can show that, for small  $\alpha$ , optimal II bias becomes insensitive to  $\alpha$ , so that our analysis of the II-case approximates the case where an advantage exists but is neglected by the principal.

with a net advantage  $\tilde{\alpha}$  or *against* it. The log-concavity of  $g$  implies that for all  $\tilde{\alpha} > 0$ ,

$$\beta_B^*(\tilde{\alpha}) > \beta^*(\tilde{\alpha}) > \beta_A^*(\tilde{\alpha}) > 0. \quad (17)$$

Intuitively, a first-stage win despite being disadvantaged is a stronger positive signal about the winner's ability than a first-stage win with a net advantage in one's favor. While comparing biases for *given* net advantage  $\tilde{\alpha}$  is straightforward, a full comparison of the equilibria in the two scenarios requires characterizing the equilibrium effort differentials:

**Proposition 2 (Equilibrium with societal luck)** *Let  $q = \frac{1}{2}$  and suppose agent  $A$ 's identity augments his first-stage performance by  $\alpha > 0$ .*

- (i) *Whether or not bias can condition on agents' identities, in equilibrium agent  $A$  exerts a lower first-stage effort than agent  $B$  but maintains a strict net advantage:*

$$-\alpha < \Delta e_1^{II} < 0 \quad \text{and} \quad -\alpha < \Delta e_1^{ID} < 0.$$

- (ii) *If the agents' ability difference  $h$  is sufficiently small, then in equilibrium*

$$\tilde{\alpha}^{ID} = \alpha + \Delta e_1^{ID} < \alpha + \Delta e_1^{II} = \tilde{\alpha}^{II},$$

*that is, making bias identity-dependent reduces agent  $A$ 's equilibrium net advantage.*

Agent  $A$ 's societal advantage  $\alpha$  makes the competition in the first stage asymmetric. Similarly to Section 3 (where for  $q > \frac{1}{2}$ , the first stage was asymmetric due to agents' information about relative abilities), the agents' first-stage effort differential arises exclusively from the impact on second-stage effort costs of which agent wins the first stage. Here, because a net advantage  $\alpha + \Delta e_1 > 0$  makes  $A$  more likely to win the first stage, both the agents and the principal are less confident in the first-stage winner's ability when it is  $A$  compared to when it is  $B$ . Under II bias, the agents will therefore expect the biased second-stage competition to be more balanced, and consequently to induce greater efforts, after a first-stage win by  $A$  than after a first-stage win by  $B$ . This difference in second-stage effort costs gives the advantaged agent,  $A$ , a weaker incentive than his rival to exert first-stage effort, resulting in  $\Delta e_1 < 0$ . Identity-dependent biases augment this "future effort-cost effect", reducing the induced  $\Delta e_1$  further below zero, because the principal will optimally choose  $\beta_A < \beta_B$  for any anticipated  $\alpha + \Delta e_1 > 0$ . As long as the ability difference  $h$  is not too large, a reduction in  $\beta_A$  makes the second-stage competition even more balanced following a win by  $A$ , and an increase in  $\beta_B$  makes the second stage even less balanced following a win by  $B$ .

We now compare the equilibria under II and ID biases with respect to both selective efficiency and the persistence of societal luck. Note first that, because the principal cannot commit to the level of bias, it is unclear a priori whether selective efficiency will be higher in the equilibrium under ID biases, even though she is less constrained here than under II bias. Yet it follows from Proposition 2(ii) that selective efficiency is always higher under ID biases than under II bias. This is because, by the envelope theorem, maximized selective efficiency under II bias is decreasing in the net advantage, and because under ID biases the net advantage is reduced. Thus, agents' effort responses *augment* the direct benefits of ID biases for selective efficiency.

Our main interest is the comparison across regimes of the persistence of societal luck, given by the probability  $P_\alpha$  of ultimately selecting the initially advantaged agent:

$$P_\alpha(\beta_A, \beta_B, \tilde{\alpha}) = \frac{1}{2} \sum_{\Delta a \in \{-h, h\}} [G(\Delta a + \tilde{\alpha})G(\Delta a + \beta_A) + G(-\Delta a - \tilde{\alpha})G(\Delta a - \beta_B)]. \quad (18)$$

Under II bias, societal luck is *always* made persistent in equilibrium: it always holds that  $P_\alpha(\beta^{II}, \beta^{II}, \alpha + \Delta e_1^{II}) > \frac{1}{2}$ . To see this, write persistence for  $\beta_A = \beta_B = \beta$  as

$$P_\alpha(\beta, \beta, \tilde{\alpha}) = \frac{1}{2} \{1 + [G(h + \tilde{\alpha}) - G(h - \tilde{\alpha})][G(h + \beta) - G(h - \beta)]\}, \quad (19)$$

and note that in equilibrium, both the net advantage,  $\tilde{\alpha}^{II} = \alpha + \Delta e_1^{II}$ , and the principal's choice of bias,  $\beta^{II}$ , are strictly positive, as shown by Proposition 2 and (17). Intuitively, the advantaged agent is selected with higher probability than his rival, because he is more likely to win the first stage (despite his lower effort), and with II bias, the second stage is biased by the *same* amount, no matter the identity of the first-stage winner.

In contrast, allowing for ID biases may completely eliminate the persistence of societal luck. For example, when the difference in agents' noise terms has a logistic distribution, then  $P_\alpha(\beta_A^{ID}, \beta_B^{ID}, \alpha + \Delta e_1^{ID}) = \frac{1}{2}$  for *all* values of exogenous advantage  $\alpha > 0$ .<sup>17</sup>

The following corollary provides a general comparison of the persistence of societal luck across the equilibria in the II and ID regimes. It also compares the expected utility difference  $\Delta U$  between the advantaged and the disadvantaged agent in these equilibria:

$$\Delta U(\beta_A, \beta_B, \alpha + \Delta e_1) \equiv [2P_\alpha(\beta_A, \beta_B, \alpha + \Delta e_1) - 1] - [C_1(e_{A,1}) - C_1(e_{B,1})]. \quad (20)$$

**Corollary 3 (Persistence of societal luck)** *Under the assumptions of Proposition 2 (ii) and for exogenous advantage  $\alpha > 0$  sufficiently small, identity-dependent biases*

---

<sup>17</sup>The logistic distribution does not satisfy part (iv) of Assumption 1, yet all our results remain valid.



(i) *reduce the persistence of societal luck in equilibrium:*

$$P_{\alpha}^{ID} \equiv P_{\alpha}(\beta_A^{ID}, \beta_B^{ID}, \alpha + \Delta e_1^{ID}) < P_{\alpha}(\beta_A^{II}, \beta_B^{II}, \alpha + \Delta e_1^{II}) \equiv P_{\alpha}^{II}$$

(ii) *reduce the expected utility difference between the advantaged and the disadvantaged agent in equilibrium:*

$$\Delta U(\beta_A^{ID}, \beta_B^{ID}, \alpha + \Delta e_1^{ID}) < \Delta U(\beta_A^{II}, \beta_B^{II}, \alpha + \Delta e_1^{II}).$$

We stress that ID biases reduce the persistence of societal luck via two distinct channels. For any *given* net advantage  $\tilde{\alpha}$ , persistence is reduced by ID biases because  $\beta_A^*(\tilde{\alpha}) < \beta^*(\tilde{\alpha}) < \beta_B^*(\tilde{\alpha})$ , and  $P_{\alpha}$  in (18) is increasing in  $\beta_A$  and decreasing in  $\beta_B$ . This channel reflects the fact that, whatever the first-stage outcome, ID biases effectively penalize, in the second stage, the agent who benefited from the first-stage advantage.

The second channel through which ID biases reduce persistence is via their effect on agents' effort incentives. As shown by Proposition 2(ii), ID biases induce the disadvantaged agent to compensate *even more* for his disadvantage through higher (relative) first-stage effort than under II bias. This effort response generates a further reduction in the persistence of societal luck, as long as  $\alpha$  is sufficiently small, because for small  $\alpha$ , a reduction in  $\tilde{\alpha}$  is guaranteed to lower  $P_{\alpha}(\beta^*(\tilde{\alpha}), \beta^*(\tilde{\alpha}), \tilde{\alpha})$ .<sup>18</sup> In fact, part (ii) of Corollary 3 shows that this effort response reduces persistence by so much that it even reduces the utility difference between the advantaged and the disadvantaged agent.

The results in this section can be interpreted as offering support for *affirmative action*, in the form of selection processes that allow organizations to condition biases on agents' identities—e.g., through gender-specific mentoring or grants accounting for socioeconomic backgrounds. Roemer (2000) argues that an *equal opportunity principle* should be applied at the entry level of careers, e.g. for admissions to medical school, while a *non-discrimination principle* should govern selection into final positions, such as the licensing of surgeons. Corollary 3i) shows that applying a non-discrimination principle (requiring bias to be identity-independent) in the selection for senior positions can backfire, by propagating disadvantages stemming from a failure to establish equal opportunity upon entry. Moreover, Corollary 3ii) shows that in our setting, affirmative action is in fact doubly beneficial, because identity-dependent biases not only improve selective efficiency, but also decrease inequality. Importantly, our analysis highlights that the incentive effects

---

<sup>18</sup> For sufficiently large  $\alpha$ , the persistence of societal luck under II bias is decreasing in net advantage because as  $\alpha \rightarrow \infty$ , the principal will ignore the first-stage outcome and select the winner of an unbiased second stage, in which  $\alpha$  has no effect on performance; persistence thus eventually falls towards  $\frac{1}{2}$ .

of such policies do not hinder their effectiveness but rather amplify their benefits.

## 5. DRIVERS OF THE PERSISTENCE OF LUCK

Our analysis in Sections 3 and 4 has shown that organizational selection induces early-career luck and societal luck to have a persistent effect on individual success. But in which careers is luck made most persistent, or in other words, what job characteristics increase the impact of initial luck on final success? In this section, we discuss several comparative statics and extensions of our model to shed light on this question.

Testing our theory would require observing individual workers' performance at various stages of their careers and determining how final success correlates with either initial performance or initial advantages derived from workers' identities. In this section, we examine the predicted impact on this correlation of three features of careers: (1) Workers' information about their relative abilities; (2) the coarseness of performance measurement; and (3) the intensity of competition for selection.

In Sections 5.1 and 5.2 we show that the persistence of early-career luck and the persistence of societal luck are both increasing with workers' information about relative abilities (Proposition 3) and with the coarseness of performance measurement (Proposition 4). In Section 5.3 we show that the effect of stronger competition for selection depends on the kind of luck under consideration (Proposition 5): Although an increase in the intensity of competition (e.g. due to lower costs of effort or larger benefits of being selected) makes early-career luck more persistent, it reduces the persistence of societal luck. Besides offering testable predictions, these results relate our theory to the continuing debate about the relative importance of luck versus merit as determinants of economic outcomes, by linking the issue with observable features of individual careers.

### 5.1. Information about relative ability

How does agents' information about their relative abilities affect the persistence of luck? To answer this, we perform comparative statics with respect to the parameter  $q$ : A rise in  $q$  represents an improvement in agents' prior information about their relative abilities.

Throughout Section 5, for simplicity and to facilitate comparisons between the persistence of early-career luck and that of societal luck, we focus on the setting where  $h$  is small when analyzing both forms of persistence.<sup>19</sup>

---

<sup>19</sup>In equilibrium, both forms of persistence are in fact locally insensitive to  $h$  as  $h$  rises from 0.

**Proposition 3 (Agents’ information)** *Both early-career luck and societal luck are more persistent in careers where agents are better informed about their relative abilities. Formally, for  $q > \frac{1}{2}$ ,  $P_0^*$  is strictly increasing in  $q$ ,  $P_\alpha^{II}$  is strictly increasing in  $q$  for small  $h$ , and  $P_\alpha^{ID}$  is strictly increasing in  $q$  for small  $h$  and  $2q - 1$  if  $L(y)$  is convex for  $y \geq 0$ .*

Proposition 3 highlights that the persistence of luck results from an equilibrium interaction: Luck is made more persistent in careers where organizations favor early winners more strongly to capitalize on agents’ strategic responses to their information about relative abilities. Empirical studies suggest that agents’ information improves when workers collaborate in teams, where interpersonal observation is key for mutual learning (Hamilton et al., 2003) and peer effects (Mas and Moretti, 2009). Proposition 3 thus suggests that initial luck has a greater impact on final success in careers with more teamwork, which is in line with the observation that in highly collaborative industries (e.g. science, management consulting), biases in the form of grants and fast tracks are abundant.<sup>20</sup>

To understand Proposition 3, note that the persistence  $P_0^* = G(\beta_0^*)$  of early-career luck strictly increases in  $\beta_0^*$ . Equation (7), determining equilibrium bias in the limit, shows  $\beta_0^*$  increases in  $q$ , for two reasons. When agents are better informed, the one exerting higher stage-one effort is more likely to be the more able one, and, from (6), the effort differential itself rises faster with  $h$  for  $h$  small. Both effects make the informativeness of the first-stage outcome rise more rapidly with  $h$  and hence increase the limiting equilibrium bias.

For societal luck, Proposition 3 generalizes the analysis in Section 4 to arbitrary  $q \geq \frac{1}{2}$ . The stage-one effort differential now depends on whether the favorite or the underdog received the advantage  $\alpha$ . Yet under II bias, as  $h$  gets small, both these differentials converge to zero. Hence, from (18), the limiting persistence of societal luck under II bias is  $\lim_{h \rightarrow 0} P_\alpha^{II} = G(\alpha)G(\beta_0^{II}) + G(-\alpha)G(-\beta_0^{II})$  which, for any  $\alpha > 0$ , is increasing in the limiting equilibrium bias. This bias,  $\beta_0^{II}$ , solves an equation ((A46) in the Appendix) analogous to (7), which shows that when agents are better informed about their relative abilities,  $\beta_0^{II}$  increases, because of the same two forces that cause  $\beta_0^*$  to increase in  $q$ .

Finally, for the case of ID bias, the limiting persistence of societal luck is  $\lim_{h \rightarrow 0} P_\alpha^{ID} = G(\tilde{\alpha})G(\beta_{A0}^{ID}) + G(-\tilde{\alpha})G(-\beta_{B0}^{ID})$ . For  $q$  sufficiently close to  $\frac{1}{2}$  (i.e.  $2q - 1$  small), both  $\beta_{A0}^{ID}$  and  $\beta_{B0}^{ID}$  are increasing in  $q$ , paralleling the result for II bias. However, the fact that  $\beta_{B0}^{ID} > \beta_{A0}^{ID}$  reduces the rate at which persistence changes with  $q$ . Nevertheless, persistence under ID

---

<sup>20</sup>Cole et al. (1981) documents the relevance of luck for scientific careers using a field experiment on NSF grant proposals and concludes that “the fate of a particular grant application is roughly half determined [...] by apparently random elements which might be characterized as the luck of the reviewer draw.” That grants make luck persistent is suggested by Zhang et al.’s (2022) finding that the superior publication records of scientists at elite universities can be largely explained by accumulated resource advantages rather than inherent differences in talent.

biases rises with  $q$  if the noise distribution is normal or has thinner tails ( $L$  convex), because this condition limits the rate at which  $\beta_{B0}^{ID}$  rises with  $q$  relative to  $\beta_{A0}^{ID}$ .

## 5.2. Coarseness of performance measurement

How does the principal's ability to measure agents' performance affect the persistence of luck? We extend our model to allow performance measurement to be less coarse than ordinal. For simplicity, we assume  $q = \frac{1}{2}$  and focus on the regime of ID biases. For any strictly increasing series of cutoffs  $d_n > 0$ , let  $I_N = \{[0, d_1), [d_1, d_2), \dots, [d_{N-1}, d_N), [d_N, \infty)\}$  define an information partition. The principal observes not only the sign of  $\Delta x_1$  but also which interval  $[d_{n-1}, d_n)$  contains the winning margin  $|\Delta x_1|$ , and can condition the bias on this interval. We call  $I_{N-1}$  a *coarsening* of  $I_N$  when  $I_{N-1}$  arises from  $I_N$  by merging two adjacent intervals. As a shorthand, let  $I_{ord}$  and  $I_{card}$  denote the extreme cases of purely ordinal and fully cardinal performance measurement.

When agents have the same identity ( $\alpha = 0$ ), the principal, if she observes  $|\Delta x_1| \in [d_{n-1}, d_n)$ , chooses  $\beta^n$  to maximize selective efficiency, where the terms in  $\beta^n$  are

$$S^n = [G(h - d_{n-1}) - G(h - d_n)]G(h + \beta^n) + [G(h + d_n) - G(h + d_{n-1})]G(h - \beta^n). \quad (21)$$

For  $h \rightarrow 0$ , the limiting equilibrium biases  $\beta_0^1, \dots, \beta_0^{N+1}$  are uniquely determined by

$$\frac{g(d_{n-1}) - g(d_n)}{G(d_n) - G(d_{n-1})} = L(\beta_0^n) \quad \text{for } n = 1, \dots, N+1. \quad (22)$$

Equations (22) generalize (7) to less coarse information, for  $q = \frac{1}{2}$ . By Assumption 1, each of the limiting biases is strictly positive, and since as  $h \rightarrow 0$ ,  $|\Delta x_1|$  falls in  $[d_{n-1}, d_n)$  with probability  $2[G(d_n) - G(d_{n-1})]$ , the *expected* persistence of early-career luck is

$$P_0^*(I_N) = \mathbb{E}[G(\beta_0^n)] = \sum_{n=1}^{N+1} 2[G(d_n) - G(d_{n-1})]G(\beta_0^n). \quad (23)$$

When agents differ in their identities ( $\alpha > 0$ ), we let  $P_\alpha^{ID}(I_N)$  denote the expected persistence of societal luck when the principal can condition bias on her information about  $|\Delta x_1|$  and on the first-stage winner's identity.

**Proposition 4 (Performance measurement)** *Suppose that  $q = \frac{1}{2}$  and  $L(y)$  is convex for  $y \geq 0$ . Both early-career luck and societal luck are more persistent in careers where measurement of relative performance is coarser. Formally, given any  $I_N$ ,  $P_0^*(I_{N-1}) > P_0^*(I_N)$  for any coarsening  $I_{N-1}$  of  $I_N$ , and for small  $h$ ,  $P_\alpha^{ID}(I_{ord}) > P_\alpha^{ID}(I_{card})$ .*

Proposition 4 identifies the coarseness of performance measurement as a driver of persistence: Any coarsening makes early-career luck more persistent, and societal advantages are more persistent when performance measurement is ordinal rather than cardinal. Lazear (2000) documents that for managers, piece rates are employed ten times less frequently than for operatives, and attributes this difference to the absence of a cardinal measure of managerial performance. Our theory thus predicts initial luck to be particularly relevant for long-run success in careers with a high degree of management duties. Moreover, since an increase in automation and use of AI shifts task and job composition towards more managerial and non-routine activities (Deming, 2021; Acemoglu et al., 2022), our results also suggest these trends should increase the persistence of initial luck in labor markets.

Proposition 4 resonates well with the frequent use of fast tracks in both private companies and public agencies. For instance, Unilever’s Future Leaders Programme or Singapore’s Public Service Leadership Programme boost the management careers of a selected group of college graduates through designated job assignments, mentoring, and leadership workshops.<sup>21</sup> Competition for these programs occurs through structured competency interviews, resulting in ordinal rankings of candidates. Our theory rationalizes why, in these settings, being lucky or socially advantaged during college can be expected to have a particularly strong effect on the likelihood of becoming a top manager.

To understand Proposition 4, observe first that for any information partition  $I_N$ , it follows from (22) that the *expected value* of  $L(\beta_0)$  is a constant,  $2g(0)$ :

$$\mathbb{E}[L(\beta_0)] = \sum_{n=1}^{N+1} 2[G(d_n) - G(d_{n-1})]L(\beta_0^n) = \sum_{n=1}^{N+1} 2[g(d_{n-1}) - g(d_n)] = 2g(0). \quad (24)$$

Thus, the (degenerate) distribution of  $L(\beta_0)$  under ordinal information is a *mean-preserving contraction* of the corresponding distribution of  $L(\beta_0)$  under any less coarse information structure.<sup>22</sup> Similar logic shows that *any* coarsening leads to a mean-preserving contraction of the distribution of  $L(\beta_0)$ . Now since  $G(\beta) = G(L^{-1}(L(\beta)))$ , expected persistence in (23) can be expressed as the expectation of  $G(L^{-1}(\cdot))$ , evaluated with respect to the distribution of  $L(\beta_0)$ . If  $L$  is convex,  $G(L^{-1}(\cdot))$  is concave, so any coarsening of the principal’s information increases the expected persistence of early-career luck.<sup>23</sup>

<sup>21</sup>For a description of these programs see <https://graduateships.com/companies/unilever/> and <https://www.psd.gov.sg/leadership/public-service-leadership-careers>.

<sup>22</sup>Equation (24) generalizes (13) in Section 3.3. For  $q = \frac{1}{2}$ , the last inequality in (13) becomes an equality and hence,  $L(\beta_0^*)$  has the same mean as  $L(|\Delta\epsilon_1|)$ , where  $|\Delta\epsilon_1|$  is the bias under cardinal information, the least coarse of all information structures.

<sup>23</sup>A similar argument shows that, in the presence of societal advantages ( $\alpha > 0$ ), the use of identity-dependent biases reduces the persistence of early-career luck compared to when bias must be identity-independent. The general lesson is that, for  $L$  convex, the more constrained the principal is in adjusting

Concerning societal luck, when the principal has full cardinal knowledge of  $|\Delta x_1|$  and can condition her bias on whether the advantaged or the disadvantaged agent won the first stage, then she will optimally choose the biases  $\beta_A^* = |\Delta x_1| - \tilde{\alpha}$  and  $\beta_B^* = |\Delta x_1| + \tilde{\alpha}$ . In this way, she perfectly corrects for the net advantage or disadvantage arising from societal luck and the induced effort differential. Hence, when performance measurement is cardinal, societal advantages cease to be persistent, and  $P_\alpha^{ID}(I_{card}) = \frac{1}{2}$ . In contrast,  $\lim_{h \rightarrow 0} P_\alpha^{ID}(I_{ord}) = G(\tilde{\alpha})G(\beta_{A0}^{ID}) + G(-\tilde{\alpha})G(-\beta_{B0}^{ID})$ , where  $\beta_{B0}^{ID} > \beta_{A0}^{ID}$ . Convexity of  $L$ , as noted in Section 5.1, limits the size of  $\beta_{B0}^{ID}$  relative to  $\beta_{A0}^{ID}$  and is here sufficient to ensure that  $\lim_{h \rightarrow 0} P_\alpha^{ID}(I_{ord}) > \frac{1}{2}$ . Hence, organizational selection makes societal advantages more persistent when performance measurement is ordinal rather than cardinal.<sup>24</sup>

### 5.3. Intensity of competition for selection

Does initial luck have a stronger or weaker effect on final success in more competitive careers? We show that, in contrast to our findings in Sections 5.1 and 5.2, how the persistence of luck changes with more intense competition for selection depends on the kind of luck under consideration. We interpret an increase in competition for selection as an increase in agents' incentives to exert effort, which can arise from either a reduction in the marginal costs of effort or an increase in the value of being selected. Because the value of being selected is normalized to 1, we identify an increase in the intensity of competition with a reduction in the parameters  $c_t$  in the quadratic cost function  $C(e_t) = \frac{c_t}{2}e_t^2$ .

**Proposition 5 (Competition)** *The effect of more intense competition for selection on the persistence of luck depends on whether luck is associated with an agent's identity:*

- (i) *Competition increases the persistence of early-career luck. Formally,  $P_0^*$  is decreasing in  $c_t$  and strictly so for  $q > \frac{1}{2}$ .*
- (ii) *Competition can decrease the persistence of societal luck. Formally, for  $q = \frac{1}{2}$  and for small  $h$  and  $\alpha$ ,  $P_\alpha^{ID}$  is strictly increasing in  $c_t$  if  $L(y)$  is convex for  $y \geq 0$ .*

To understand Proposition 5i), recall from Section 3 that, when  $q > \frac{1}{2}$ , agents' efforts make their stage-one performance ranking more aligned with their ability ranking. Specifically, the proportionality of (6) with respect to  $\frac{1}{c_1 c_2}$  shows that intensifying competition makes the *positive* effort differential between the favorite and the underdog rise faster with heterogeneity  $h$  as  $h$  rises from 0. Hence, from (7), organizations will choose larger biases when noise swamps ability differences, making early-career luck more persistent.

---

her bias to the information contained in the first-stage outcome, the more persistent is early-career luck.

<sup>24</sup>For "intermediate" information structures  $I_1(d_1)$ , we have used our family of exponential power distributions to show numerically that  $P_\alpha^{ID}(I_{ord}) > P_\alpha^{ID}(I_1) > P_\alpha^{ID}(I_{card})$  for  $\alpha < \frac{d_1}{2}$  and  $L$  convex.

For societal luck, greater competition between uninformed agents ( $q = \frac{1}{2}$ ) has the opposite effect on persistence, because as shown in Section 4, the disadvantaged agent will partially compensate for his disadvantage by working harder than his rival, i.e. the effort differentials  $\Delta e_1^{ID}$  and  $\Delta e_1^{II}$  are *negative*. Intensifying competition will amplify this effect, further reducing the net advantage of the advantaged agent. For ID biases,  $P_\alpha^{ID}$  is increasing in net advantage for small  $\alpha$ , as long as  $h$  is small and  $L$  is convex, so under these conditions agents' effort response to greater competition reduces  $P_\alpha^{ID}$ .<sup>25</sup>

There is extensive evidence of ever-toughening competition for jobs in internal and external labor markets, especially when strong sensitivity to quality differences generates “superstar” effects (Rosen, 1981). Inside firms, competition has been intensified by increasing wage inequality (see, e.g., Acemoglu and Autor, 2011, for a review)—especially across the top tier of the hierarchy—and by the rising homogeneity of workers resulting from a more assortative matching between high-skilled workers and high-wage firms (Card et al., 2013). Across firms, small differences in managerial talent are found to translate into large pay differences for CEOs (Gabaix and Landier, 2008), and growing firm heterogeneity has intensified competition between managers vying for the top employers (Autor et al., 2020; Bao et al., 2026). Although estimates of the resources workers spend competing are difficult to obtain, higher rates of turnover and incidence of burnouts are indications of larger efforts being exerted.

Proposition 5 suggests that the growing intensity of competition for jobs has made early-career luck increasingly persistent. A silver lining is that an increase in competition can *reduce* the persistence of advantages individuals derive from their identity. This latter finding relates our theory to the literature on relative intergenerational mobility, which provides an empirical proxy for the persistence of societal luck. The literature has identified a negative correlation between relative mobility and income inequality (“The Great Gatsby Curve”, Corak, 2013), which may arise from the well-established fact that inequality of income translates into inequality of opportunity (e.g. education), particularly at early ages (Currie and Almond, 2011). Proposition 5 shows that, on the contrary, large differentials in potential earnings may create incentives that improve mobility by raising the efforts of the disadvantaged relative to the advantaged. This might help understand why mobility fails to correlate with the earnings share of the Top 1% (Chetty et al., 2014). An increase in the earnings share of top earners, while changing opportunities for a few, might boost the incentives of many, so that the incentive effects identified by our theory might offset the opportunity effects highlighted by the literature.

---

<sup>25</sup>For II bias, it can be shown that agents' effort response to increased competition disappears as  $h$  gets small, but for arbitrary  $h$ ,  $P_\alpha^{II}$  will fall as competition intensifies, as long as  $\alpha$  is small enough to guarantee that  $P_\alpha^{II}$  is increasing in net advantage.



## 6. CONCLUSION

When the careers of National Hockey League players (Deaner et al., 2013) or S&P 500 CEOs (Du et al., 2012) are kick-started by the proximity of their birthdays to a cut-off, luck seems to play an unjustified role in talent selection. Such findings might seem to support recent critiques of a meritocratic worldview (e.g. Piketty, 2014; Sandel, 2020) which argue that meritocracy is a myth, used to justify exorbitant degrees of economic and social inequality.

The main contribution of this paper is to show that making early-career luck have a persistent effect on individual careers is consistent with—if not a necessary feature of—a society aiming to allocate resources and decision-making power to the most able individuals. Our theory illuminates a basic mechanism behind economic inequality by rationalizing the persistence of luck as an equilibrium outcome of the strategic interaction between an organization aiming to maximize selective efficiency and heterogeneous agents choosing costly efforts to influence their chances of being selected.

We have also analyzed how organizational learning contributes to the persistence of advantages that certain individuals initially derive from their identities (e.g. gender, race, socio-economic background). Our results suggest that non-discrimination policies that constrain an organization’s selection process may backfire, by propagating disadvantages stemming from unequal initial opportunities.

Finally, we have characterized the settings where our theory predicts the impact of initial luck on final success to be largest. This happens when agents are well informed about their relative abilities and the organization is restricted to using ordinal rather than cardinal performance information. To the extent that the rise of teamwork and the proliferation of management duties have improved workers’ information about co-workers’ abilities while hampering the measurement of individual contributions and relative performance, our theory suggests an increasing persistence of luck in careers. Although more intense competition for top jobs and for jobs at superstar employers reinforces this trend by making early-career luck more persistent, the incentive effects of greater competition can be helpful in mitigating the persistence of advantages associated with agents’ identities.

## APPENDIX - PROOFS OF MAIN RESULTS

### Proof of Lemma 1

Use superscripts  $w$  and  $l$ , respectively, to distinguish the cases where agent  $A$  won and lost 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  $\Delta e_2^l = e_{A,2}^l - e_{B,2}^l$ . Let  $q^w(\Delta e_1, q)$  and  $q^l(\Delta e_1, q)$  denote the agents’ posterior probabilities that the *winner* of

the first stage is the more able agent, given  $q$  and  $\Delta e_1$ .

We first show that agents exert identical effort in the second stage and that this holds *independently* of  $q$  and  $\Delta e_1$ . In case  $w$ ,  $A$ 's and  $B$ 's second-stage efforts solve:

$$\begin{aligned} C_2'(e_{A,2}^w) &= q^w(\Delta e_1, q)g(h + \beta + \Delta e_2^w) + (1 - q^w(\Delta e_1, q))g(-h + \beta + \Delta e_2^w) \\ C_2'(e_{B,2}^w) &= q^w(\Delta e_1, q)g(-h - \beta - \Delta e_2^w) + (1 - q^w(\Delta e_1, q))g(h - \beta - \Delta e_2^w). \end{aligned}$$

Given the symmetry of  $g$  about 0, the marginal returns to effort are identical for  $A$  and  $B$ , so  $e_{A,2}^w = e_{B,2}^w$ . An analogous argument for case  $l$  shows that  $e_{A,2}^l = e_{B,2}^l$ .

Now consider the first stage. For  $\beta \leq 0$ , the agents have no incentives to win the first stage and hence  $e_{A,1} = e_{B,1} = 0$ . Consider  $\beta > 0$ . The overall utility of agent  $A$  is:

$$\begin{aligned} -C_1(e_{A,1}) &+ q\{G(h + \Delta e_1)[G(h + \beta + \Delta e_2^w) - C_2(e_{A,2}^w)] \\ &+ [1 - G(h + \Delta e_1)][G(h - \beta + \Delta e_2^l) - C_2(e_{A,2}^l)]\} \\ &+ (1 - q)\{G(-h + \Delta e_1)[G(-h + \beta + \Delta e_2^w) - C_2(e_{A,2}^w)] \\ &+ [1 - G(-h + \Delta e_1)][G(-h - \beta + \Delta e_2^l) - C_2(e_{A,2}^l)]\}. \end{aligned}$$

A change in  $e_{A,1}$  does not affect  $e_{B,2}^w$ ,  $e_{B,2}^l$ , or  $\beta$ , 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  $\Delta e_2^w = \Delta e_2^l = 0$  and the symmetry of  $g$ , the FOC for  $e_{A,1}$  is

$$\begin{aligned} C_1'(e_{A,1}) &= [qg(h + \Delta e_1) + (1 - q)g(-h + \Delta e_1)] \\ &\cdot \{[G(h + \beta) - G(h - \beta)] - [C_2(e_2^w) - C_2(e_2^l)]\}. \end{aligned} \tag{A1}$$

Analogously, for agent  $B$ , the FOC for  $e_{B,1}$  is

$$\begin{aligned} C_1'(e_{B,1}) &= [(1 - q)g(h - \Delta e_1) + qg(-h - \Delta e_1)] \\ &\cdot \{[G(h + \beta) - G(h - \beta)] + [C_2(e_2^w) - C_2(e_2^l)]\}. \end{aligned} \tag{A2}$$

Subtracting (A2) from (A1), and using the symmetry of  $g$ , gives

$$\frac{C_1'(e_{A,1}) - C_1'(e_{B,1})}{C_2(e_2^l) - C_2(e_2^w)} = 2[qg(h + \Delta e_1) + (1 - q)g(-h + \Delta e_1)]. \tag{A3}$$

Since the effort cost functions are strictly increasing and strictly convex, we conclude that in equilibrium,  $\Delta e_1 = e_{A,1} - e_{B,1}$  and  $e_2^l - e_2^w$  must have the same sign.

To determine the sign of  $e_2^l - e_2^w$ , compare agent  $A$ 's FOCs for second-stage effort,

after a first-stage win by  $A$  vs. after a first-stage loss by  $A$ , respectively:

$$C'_2(e_2^w) = q^w(\Delta e_1, q)g(h + \beta) + (1 - q^w(\Delta e_1, q))g(-h + \beta), \quad (\text{A4})$$

$$C'_2(e_2^l) = q^l(\Delta e_1, q)g(-h - \beta) + (1 - q^l(\Delta e_1, q))g(h - \beta). \quad (\text{A5})$$

Subtracting the second FOC from the first, and using the symmetry of  $g$ , gives

$$C'_2(e_2^w) - C'_2(e_2^l) = [q^w(\Delta e_1, q) - q^l(\Delta e_1, q)][g(h + \beta) - g(-h + \beta)]. \quad (\text{A6})$$

The strict log-concavity and symmetry of  $g$  imply that for any  $\beta > 0$ ,  $g(h + \beta) - g(-h + \beta) < 0$ , while for  $\beta = 0$ ,  $g(h + \beta) - g(-h + \beta) = 0$ . Hence, since costs are strictly convex,

$$e_2^l - e_2^w \geq 0 \iff q^w(\Delta e_1, q) - q^l(\Delta e_1, q) \geq 0. \quad (\text{A7})$$

The agents' posterior beliefs  $q^w(\Delta e_1, q)$  and  $q^l(\Delta e_1, q)$  are given by

$$q^w(\Delta e_1, q) = \frac{qG(h + \Delta e_1)}{qG(h + \Delta e_1) + (1 - q)G(-h + \Delta e_1)}, \quad (\text{A8})$$

$$q^l(\Delta e_1, q) = \frac{(1 - q)G(h - \Delta e_1)}{(1 - q)G(h - \Delta e_1) + qG(-h - \Delta e_1)}. \quad (\text{A9})$$

Observe that  $q^w$  and  $q^l$  are, respectively, strictly decreasing and strictly increasing in  $\Delta e_1$ . For  $q = \frac{1}{2}$ , they are equal at  $\Delta e_1 = 0$ , while for  $q > \frac{1}{2}$ , they are equal at some  $\Delta e_1 > 0$ .

Consider  $q = \frac{1}{2}$ . Suppose, for contradiction, that  $\Delta e_1 < 0$ . Then  $q^w(\Delta e_1, q) - q^l(\Delta e_1, q) > 0$  since winning despite an effort disadvantage is a stronger signal than winning with an effort advantage. Then, by (A7),  $e_2^l > e_2^w$ . In turn, this implies, using (A3), that  $\Delta e_1 > 0$ , which is a contradiction. Analogously, assuming that  $\Delta e_1 > 0$  also leads to a contradiction. Hence, for  $q = \frac{1}{2}$ , agents choose  $e_{A,1} = e_{B,1}$ .

Consider now  $q > \frac{1}{2}$ . Suppose, for contradiction, that  $\Delta e_1 \leq 0$ . Then from (A8) and (A9),  $q^w(\Delta e_1, q) - q^l(\Delta e_1, q) > 0$ : the win by agent  $A$  despite his effort disadvantage and given the prior  $q > \frac{1}{2}$  in his favor is a stronger signal than the win by agent  $B$  with his effort advantage and the prior against him. By (A7), it follows that  $e_2^l > e_2^w$ . In turn, this implies, using (A3), that  $\Delta e_1 > 0$ , which is a contradiction. ■

## Proof of Lemma 2

Equation (5) is the FOC for maximizing (4) with respect to  $\beta$ . To check the SOC, consider

$$\frac{\partial^2 S(\beta, \Delta e_1; h)}{\partial \beta^2} = \kappa(\Delta e_1; h)g'(h + \beta) + [1 - \kappa(\Delta e_1; h)]g'(h - \beta).$$

Using (5) and the symmetry of  $g$ , we get

$$\frac{\partial^2 S(\beta, \Delta e_1; h)}{\partial \beta^2} \propto \frac{g'(h + \beta)}{g(h + \beta)} + \frac{g'(h - \beta)}{g(h - \beta)} = \frac{g'(\beta + h)}{g(\beta + h)} - \frac{g'(\beta - h)}{g(\beta - h)} < 0,$$

where the inequality is equivalent to  $g$ 's strict log-concavity. To see that (5) has a unique, strictly positive solution, note from (3) that for  $\Delta e_1 \geq 0$ ,  $G$  increasing and  $q \geq \frac{1}{2}$  imply

$$\kappa(\Delta e_1; h) > qG(\Delta e_1) + (1 - q)G(-\Delta e_1) \geq \frac{1}{2} [G(\Delta e_1) + G(-\Delta e_1)] = \frac{1}{2}.$$

Hence, the LHS of (5),  $\frac{\kappa(\Delta e_1; h)}{1 - \kappa(\Delta e_1; h)} > 1$ . The RHS,  $\frac{g(h - \beta)}{g(h + \beta)}$ , equals 1 at  $\beta = 0$ , is strictly increasing in  $\beta$  by strict log-concavity of  $g$ , and  $\rightarrow \infty$  as  $\beta \rightarrow z - h$ . To see this for  $z = +\infty$ , recall Assumption 1(iv) and note that  $L(y) = -(\ln g(y))'$  implies

$$\ln \frac{g(h - \beta)}{g(h + \beta)} = \ln g(\beta - h) - \ln g(\beta + h) = \int_{\beta - h}^{\beta + h} L(y) dy > 2hL(\beta - h),$$

where the inequality holds since  $L$  is strictly increasing. ■

### Proof of Proposition 1

To characterize the principal's equilibrium choice of bias in the limit  $h \rightarrow 0$ , denoted as  $\beta_0^* = \lim_{h \rightarrow 0} \beta^*(h)$ , we first analyze agents' efforts. For  $h \rightarrow 0$ , (A6) implies  $(e_2^w - e_2^l) \rightarrow 0$ , so the RHSs of (A1) and (A2) become equal. Hence,  $\lim_{h \rightarrow 0} \Delta e_1^*(h) = 0$ .

Given the principal's equilibrium conjecture  $\Delta e_1^*(h)$ , her optimal bias  $\beta^*(\Delta e_1^*, h)$  solves the FOC corresponding to (4):

$$\begin{aligned} S_\beta(\beta, \Delta e_1^*; h) &= [qG(h + \Delta e_1^*) + (1 - q)G(h - \Delta e_1^*)]g(h + \beta) \\ &\quad - \{q[1 - G(h + \Delta e_1^*)] + (1 - q)[1 - G(h - \Delta e_1^*)]\}g(h - \beta) = 0. \end{aligned} \quad (\text{A10})$$

Since  $\lim_{h \rightarrow 0} \Delta e_1^* = 0$ ,  $\lim_{h \rightarrow 0} S_\beta(\beta, \Delta e_1^*; h) = 0$  for all  $\beta$ . Hence, to characterize  $\beta_0^*$  we need to differentiate  $S_\beta(\beta^*(h), \Delta e_1^*(h); h) = 0$  totally with respect to  $h$  and let  $h \rightarrow 0$ . Given  $\lim_{h \rightarrow 0} S_\beta(\beta, \Delta e_1^*; h) = 0$  for all  $\beta$ ,  $\lim_{h \rightarrow 0} S_{\beta\beta}(\beta^*, \Delta e_1^*; h) = 0$ , and we get

$$0 = 2g(0)g(\beta_0^*) \left[ 1 + (2q - 1) \lim_{h \rightarrow 0} \frac{\partial \Delta e_1^*(\beta_0^*; h)}{\partial h} \right] + g'(\beta_0^*). \quad (\text{A11})$$

(A11) is equivalent to (7), so it remains to show that (A11) has a unique solution, which is strictly positive. For this purpose, we compute  $\lim_{h \rightarrow 0} \frac{\partial \Delta e_1^*(\beta^*(h); h)}{\partial h}$ , making use of the fact that  $\lim_{h \rightarrow 0} \frac{\partial \beta^*(h)}{\partial h} = 0$ , which we prove below. First, multiplying (A3) by  $C_2(e_2^l) - C_2(e_2^w)$ ,

then differentiating both sides with respect to  $h$  and taking  $h \rightarrow 0$  gives

$$C_1'''(e_{A,1}) \lim_{h \rightarrow 0} \frac{\partial \Delta e_1^*(\beta; h)}{\partial h} = 2g(0) \lim_{h \rightarrow 0} C_2'(e_2^w) \frac{\partial(e_2^l - e_2^w)}{\partial h}, \quad (\text{A12})$$

where all efforts denote their optimal values as  $h \rightarrow 0$  and we used that as  $h \rightarrow 0$ ,  $\Delta e_1^* = 0$  and  $e_2^w = e_2^l$ . Next, differentiating (A6) with respect to  $h$  and letting  $h \rightarrow 0$  gives

$$\lim_{h \rightarrow 0} \frac{\partial(e_2^l - e_2^w)}{\partial h} = -2(2q - 1)g'(\beta)/C_2''(e_2^w) > 0. \quad (\text{A13})$$

Using  $C_t''(e) = c_t$  and  $\lim_{h \rightarrow 0} C_2'(e_2^w) = g(\beta_0^*)$  from (A4), combining (A12) and (A13) gives

$$\lim_{h \rightarrow 0} \frac{\partial \Delta e_1^*(\beta; h)}{\partial h} = -4 \frac{(2q - 1)}{c_1 c_2} g(0) g(\beta) g'(\beta). \quad (\text{A14})$$

Substituting into (A11) leads to

$$2g(0) = L(\beta_0^*) \left[ 1 - \frac{8(2q - 1)^2}{c_1 c_2} g(0)^2 g(\beta_0^*)^2 \right]. \quad (\text{A15})$$

Because  $L$  and the term in square brackets are strictly increasing in  $\beta$ , and because  $L(0) = 0$  and  $\lim_{y \rightarrow z} L(y) = \infty$  by Assumption 1, there is a unique  $\beta_0^* > 0$  solving (A15).

Finally, to show that  $\lim_{h \rightarrow 0} \frac{\partial \beta^*(h)}{\partial h} = 0$ , suppose agent  $A$  is likely to be “better” than agent  $B$  by  $-h$ , which effectively switches agents’ “names”. Irrelevance of naming implies  $\Delta e_1^*(-h) = -\Delta e_1^*(h)$ , and it follows that

$$\lim_{h \rightarrow 0} \frac{\partial^2 \Delta e_1^*(h)}{\partial h^2} = 0. \quad (\text{A16})$$

Now  $S(\beta, \Delta e_1^*(-h); -h)$  is the probability of selecting the worse agent, and since one of the two agents is always selected,

$$S(\beta, \Delta e_1^*(-h); -h) + S(\beta, \Delta e_1^*(h); h) = 1. \quad (\text{A17})$$

Indeed, from the symmetry of  $g$  we have  $G(-x) = 1 - G(x)$  and hence,  $\kappa(\Delta e_1^*(-h); -h) + \kappa(\Delta e_1^*(h); h) = 1$ , so that (4) leads to (A17). Differentiating (A17) twice with respect to  $h$ , taking the limit  $h \rightarrow 0$ , and using (A16) yields  $\lim_{h \rightarrow 0} S_{hh}(\beta, \Delta e_1^*(h); h) = 0$ . Hence,  $\lim_{h \rightarrow 0} S_{\beta hh}(\beta, \Delta e_1^*(h); h) = 0$  for any  $\beta$ , and it follows that  $\lim_{h \rightarrow 0} \frac{\partial \beta^*(h)}{\partial h} = 0$ . ■

## Proof of Corollary 2

Given equation (12), we must show that convexity of  $L(y)$  for  $y \geq 0$  implies  $P_0^* = G(\beta_0^*) > \frac{3}{4}$ . Write  $k = |\Delta e_1|$ , and note that  $k$  is distributed with density  $2g(k)$  on  $[0, z]$ . We first

prove that  $\beta_0^* \geq \mathbb{E}[k]$ . To show this, note first that (6) and (7) imply  $L(\beta_0^*) \geq 2g(0)$ . Also,

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

using  $g(z) = 0$  by Assumption 1(iv). Convexity of  $L(y)$  for  $y \geq 0$  implies  $L(\beta_0^*) \geq \mathbb{E}[L(k)] \geq L(\mathbb{E}[k])$ , and since  $L$  is increasing by Assumption 1(ii), we have  $\beta_0^* \geq \mathbb{E}[k]$ .

Now, since  $G(y)$  is increasing and strictly concave for  $y > 0$  by Assumption 1,  $G(\beta_0^*) \geq G(\mathbb{E}[k]) > \mathbb{E}[G(k)]$ . Evaluating

$$\mathbb{E}[G(k)] = 2 \int_0^z G(k)g(k)dk = 2 \left( G(k)^2 \Big|_0^z - \int_0^z G(k)g(k)dk \right),$$

and using  $G(k)^2 \Big|_0^z = \frac{3}{4}$ , yields  $\mathbb{E}[G(k)] = \frac{3}{4}$ . Hence  $P_0^* = G(\beta_0^*) > \frac{3}{4}$ . ■

### Proof of Proposition 2

Before proving claims (i) and (ii), we note that  $\beta_A^*(\tilde{\alpha})$ ,  $\beta_B^*(\tilde{\alpha})$ , and  $\beta^*(\tilde{\alpha})$  are all strictly positive, because the LHSs of the FOCs (15) and (16) are strictly larger than one, while the RHSs equal one when bias is zero and strictly increase in bias by the log-concavity of  $g$ . Moreover, (17) holds because for all  $\tilde{\alpha} > 0$ ,

$$\frac{G(h + \tilde{\alpha})}{G(-h + \tilde{\alpha})} < \frac{G(h + \tilde{\alpha}) + G(h - \tilde{\alpha})}{G(-h + \tilde{\alpha}) + G(-h - \tilde{\alpha})} < \frac{G(h - \tilde{\alpha})}{G(-h - \tilde{\alpha})}. \quad (\text{A18})$$

Since the first two ratios in (A18) are decreasing in  $\tilde{\alpha}$  whereas the third is increasing,  $\beta_A^*(\tilde{\alpha})$  and  $\beta^*(\tilde{\alpha})$  are strictly decreasing whereas  $\beta_B^*(\tilde{\alpha})$  is strictly increasing. Since all three ratios in (A18) become equal as  $\tilde{\alpha} \rightarrow 0$ , it follows that  $\lim_{\alpha \rightarrow 0} \beta_A^*(\tilde{\alpha}) = \lim_{\alpha \rightarrow 0} \beta_B^*(\tilde{\alpha}) = \lim_{\alpha \rightarrow 0} \beta^*(\tilde{\alpha})$ . Differentiating the LHS of (16) with respect to  $\tilde{\alpha}$  gives

$$\frac{2[g(h + \tilde{\alpha}) - g(h - \tilde{\alpha})]}{[G(-h + \tilde{\alpha}) + G(-h - \tilde{\alpha})]^2},$$

which converges to zero as  $\tilde{\alpha} \rightarrow 0$ , thus implying that  $\lim_{\tilde{\alpha} \rightarrow 0} \frac{\partial \beta^*}{\partial \tilde{\alpha}} = 0$ . Since the FOCs (15) are identical except for the sign of  $\tilde{\alpha}$ ,  $\lim_{\tilde{\alpha} \rightarrow 0} \frac{\partial \beta_A^*}{\partial \tilde{\alpha}} = -\lim_{\tilde{\alpha} \rightarrow 0} \frac{\partial \beta_B^*}{\partial \tilde{\alpha}}$ . Finally, since the first ratio in (A18) is strictly decreasing in  $\tilde{\alpha}$  even as  $\tilde{\alpha} \rightarrow 0$ , we have  $\lim_{\tilde{\alpha} \rightarrow 0} \frac{\partial \beta_A^*}{\partial \tilde{\alpha}} < 0$ .

**Part (i)** This proof treats jointly the cases of ID and II biases—for II bias, simply impose  $\beta_A = \beta_B = \beta$ . In stage two, agents exert identical efforts, as follows from arguments analogous to those for Lemma 1. Let  $e_2^w$  and  $e_2^l$  denote agents' stage-two efforts after the

advantaged agent  $A$  won or lost stage one, respectively. The FOCs for  $e_{A,1}$  and  $e_{B,1}$  are

$$2C'_1(e_{A,1}) = \sum_{\Delta a} g(\Delta a + \tilde{\alpha})[G(\Delta a + \beta_A) - G(\Delta a - \beta_B) + C_2(e_2^l) - C_2(e_2^w)], \quad (\text{A19})$$

$$2C'_1(e_{B,1}) = \sum_{\Delta a} g(\Delta a + \tilde{\alpha})[G(-\Delta a + \beta_B) - G(-\Delta a - \beta_A) + C_2(e_2^w) - C_2(e_2^l)], \quad (\text{A20})$$

where the summations are over  $\Delta a \in \{-h, h\}$ . Since  $G(-x) = 1 - G(x)$  by symmetry of  $g$ , the marginal effect of effort on the probability of selection is the same for  $A$  and  $B$ . Subtracting  $B$ 's FOC from  $A$ 's yields:

$$\frac{C'_1(e_{A,1}) - C'_1(e_{B,1})}{C_2(e_2^l) - C_2(e_2^w)} = \sum_{\Delta a \in \{-h, h\}} g(\Delta a + \alpha + \Delta e_1). \quad (\text{A21})$$

Given  $C_1$  and  $C_2$  are increasing and convex, in equilibrium  $\Delta e_1 = e_{A,1} - e_{B,1}$  and  $e_2^l - e_2^w$  must have the same sign. To determine this sign, consider  $A$ 's expected utility in stage two, separately for the two cases where  $A$  won ( $w$ ) or lost ( $l$ ) stage one, respectively:

$$q^w G(h + \beta_A) + (1 - q^w)G(-h + \beta_A) - C_2(e_2^w) \text{ and } q^l G(-h - \beta_B) + (1 - q^l)G(h - \beta_B) - C_2(e_2^l).$$

Here we have introduced

$$q^w = \frac{G(h + \tilde{\alpha})}{G(h + \tilde{\alpha}) + G(-h + \tilde{\alpha})} \quad \text{and} \quad q^l = \frac{G(h - \tilde{\alpha})}{G(h - \tilde{\alpha}) + G(-h - \tilde{\alpha})}$$

to denote all players' posterior probabilities that the winner of the first stage is the more able agent. The corresponding FOCs determining  $e_2^w$  and  $e_2^l$  are

$$C'_2(e_2^w) = q^w g(h + \beta_A) + (1 - q^w)g(-h + \beta_A), \quad (\text{A22})$$

$$C'_2(e_2^l) = q^l g(-h - \beta_B) + (1 - q^l)g(h - \beta_B). \quad (\text{A23})$$

Note that  $q^w$  (resp.  $q^l$ ) decreases (resp. increases) in  $A$ 's net advantage  $\tilde{\alpha} = \alpha + \Delta e_1$ , and  $q^l > q^w \Leftrightarrow \tilde{\alpha} > 0$ . With ID biases,  $\beta_A - \beta_B$  and  $q^w - q^l$  have the same sign, since the principal sets a larger bias the more confident she is that the stage-one winner is more able.

We now argue, by contradiction, that  $-\alpha < \Delta e_1 < 0$ . Suppose first that, instead,  $\Delta e_1 \leq -\alpha$ . Then  $\alpha + \Delta e_1 \leq 0$  implies that  $q^l \leq q^w$  and thus  $\beta_A \geq \beta_B$ . (For identity-



independent bias, this condition holds trivially.) We have, for all  $\beta \in (0, \beta_A]$ ,

$$\frac{q^w}{1 - q^w} = \frac{g(h - \beta_A)}{g(h + \beta_A)} \geq \frac{g(h - \beta)}{g(h + \beta)} > \frac{g'(h - \beta)}{g'(h + \beta)} = -\frac{g'(-h + \beta)}{g'(h + \beta)},$$

where the first equality is the principal's first-order condition for  $\beta_A$ , the two inequalities follow from  $\beta \in (0, \beta_A]$  and the strict log-concavity of  $g$ , and the second equality holds because  $g$  is symmetric. Hence, for  $\beta \in (0, \beta_A]$ ,  $q^w g'(h + \beta) + (1 - q^w)g'(-h + \beta) < 0$ , so

$$q^w g(h + \beta_A) + (1 - q^w)g(-h + \beta_A) \leq q^w g(h + \beta_B) + (1 - q^w)g(-h + \beta_B), \quad (\text{A24})$$

with strict inequality if  $\beta_B < \beta_A$ . Since  $\beta_B > 0$  and, under the hypothesis,  $q^l \leq q^w$ , the RHS of (A24) is less than or equal to  $q^l g(h + \beta_B) + (1 - q^l)g(-h + \beta_B)$ . Hence (A22) and (A23) imply  $C'_2(e_2^w) \leq C'_2(e_2^l)$ , and convexity of  $C_2$  implies  $e_2^w \leq e_2^l$ . Since  $\Delta e_1$  must have the same sign as  $e_2^l - e_2^w \geq 0$ , we have a contradiction to our assumption that  $\Delta e_1 \leq -\alpha < 0$ .

Similarly, if  $\Delta e_1 \geq 0$ , then it follows from  $\alpha + \Delta e_1 > 0$  that  $q^l > q^w$ , so  $\beta_B > \beta_A$ . Now we have, for all  $\beta \in (0, \beta_B)$ ,

$$\frac{q^l}{1 - q^l} = \frac{g(h - \beta_B)}{g(h + \beta_B)} > \frac{g(h - \beta)}{g(h + \beta)} > \frac{g'(h - \beta)}{g'(h + \beta)} = -\frac{g'(-h + \beta)}{g'(h + \beta)},$$

and thus  $q^l g'(h + \beta) + (1 - q^l)g'(-h + \beta) < 0$ . Hence, since  $\beta_B > \beta_A > 0$ ,

$$q^l g(h + \beta_B) + (1 - q^l)g(-h + \beta_B) < q^l g(h + \beta_A) + (1 - q^l)g(-h + \beta_A),$$

and the RHS is strictly smaller than  $q^w g(h + \beta_A) + (1 - q^w)g(-h + \beta_A)$  because  $q^l > q^w$ . It follows from (A22) and (A23) that  $C'_2(e_2^w) > C'_2(e_2^l)$  and thus  $e_2^w > e_2^l$ . Since in equilibrium,  $\Delta e_1$  must have the same sign as  $e_2^l - e_2^w < 0$ , we obtain a contradiction.

**Part (ii)** Let  $(\beta^{II}, \Delta e_1^{II})$  and  $(\beta_A^{ID}, \beta_B^{ID}, \Delta e_1^{ID})$  denote the unique equilibria with identity-independent and identity-dependent biases, respectively.<sup>26</sup> Assume that  $h$  is sufficiently small such that  $-2h + \beta_A^{ID} \geq 0$ . An interval of such  $h$  values exists because, by an argument analogous to the proof of Proposition 1,  $\lim_{h \rightarrow 0} \beta_A^{ID} > 0$ . We now show that

$$\Delta e_1^{ID} < \Delta e_1^{II}. \quad (\text{A25})$$

For contradiction, assume  $\Delta e_1^{ID} \geq \Delta e_1^{II}$ . Starting from  $(\beta_A^{ID}, \beta_B^{ID}, \Delta e_1^{ID})$ , suppose the

---

<sup>26</sup>Recall that we assumed, in footnote 9, that the slopes of the marginal cost of effort curves in each period are large enough to ensure the existence and uniqueness of equilibria.

principal's optimal bias given  $\Delta e_1^{ID}$  must now be identity-independent, with  $\beta_A = \beta_B = \hat{\beta} \equiv \beta^*(\alpha + \Delta e_1^{ID})$ . Consider the agents' optimal effort response  $\Delta e_1^*(\hat{\beta}; \hat{\beta})$  to  $\hat{\beta}$ . As costs are quadratic, it follows from (A21) that  $\Delta e_1^*(\hat{\beta}; \hat{\beta})$  satisfies the implicit equation

$$c_1 \Delta e_1 - [C_2(e_2^l) - C_2(e_2^w)] \sum_{\Delta a \in \{-h, h\}} g(\Delta a + \alpha + \Delta e_1) = 0, \quad (\text{A26})$$

where, using (A22) and (A23),

$$\begin{aligned} C_2(e_2^l) - C_2(e_2^w) = & \frac{1}{c_2} [q^l g(-h - \beta_B) + (1 - q^l) g(h - \beta_B)]^2 \\ & - \frac{1}{c_2} [q^w g(h + \beta_A) + (1 - q^w) g(-h + \beta_A)]^2, \end{aligned}$$

with  $\beta_A = \beta_B = \hat{\beta}$ . Because  $\beta_A^{ID} < \hat{\beta} < \beta_B^{ID}$ , as shown at the start of the proof of this proposition, the move from  $\beta_A = \beta_A^{ID}$  and  $\beta_B = \beta_B^{ID}$  to  $\beta_A = \beta_B = \hat{\beta}$  decreases  $g(2h + \beta_A)$  and increases  $g(-2h - \beta_B)$  and, given  $-2h + \beta_A^{ID} \geq 0$  (which implies  $h - \beta_B^{ID} < 0$ ), it also decreases  $g(-2h + \beta_A)$  and increases  $g(2h - \beta_B)$ . The move from  $(\beta_A^{ID}, \beta_B^{ID})$  to  $\hat{\beta}$  thus reduces the LHS of (A26) for any fixed  $\Delta e_1$  by increasing  $C_2(e_2^l) - C_2(e_2^w)$ , which is negative, as shown in the proof of claim (i). Given that the LHS of (A26) is negative for  $\Delta e_1 = -\alpha$  and positive for  $\Delta e_1 = 0$  and that equilibrium is unique, the move from  $(\beta_A^{ID}, \beta_B^{ID})$  to  $\hat{\beta}$  thus leads to an increase in  $\Delta e_1$ , i.e. we have shown that  $\Delta e_1^*(\hat{\beta}, \hat{\beta}) > \Delta e_1^*(\beta_A^{ID}, \beta_B^{ID}) = \Delta e_1^{ID}$ .

To see that this leads to a contradiction, define the function  $\gamma(\beta) = (\beta^*)^{-1}(\beta) - \alpha$ . That is,  $\gamma(\beta)$  gives the effort differential  $\Delta e_1$  that makes  $\beta$  the principal's optimal  $\Pi$  bias. Given uniqueness of the equilibrium  $(\beta^{II}, \Delta e_1^{II})$ , the curves  $\gamma(\beta)$  and  $\Delta e_1^*(\beta, \beta)$  intersect exactly once. And because  $\Delta e_1^*(\beta, \beta)$  goes to zero for  $\beta \rightarrow 0$  and for  $\beta \rightarrow \infty$ , and  $\gamma(\beta)$  is strictly decreasing,  $\Delta e_1^*(\beta, \beta)$  must cross  $\gamma(\beta)$  from below. In particular, for any  $\beta < \beta^{II}$  it must hold that  $\gamma(\beta) > \Delta e_1^*(\beta, \beta)$ . Now  $\hat{\beta} = \beta^*(\alpha + \Delta e_1^{ID}) \leq \beta^*(\alpha + \Delta e_1^{II}) = \beta^{II}$ , because  $\beta^*$  is decreasing and we have assumed that  $\Delta e_1^{ID} \geq \Delta e_1^{II}$ . Hence  $\gamma(\hat{\beta}) \geq \Delta e_1^*(\hat{\beta}, \hat{\beta})$ . Because by definition  $\hat{\beta} = \beta^*(\alpha + \Delta e_1^{ID})$  and  $\gamma(\hat{\beta}) = \Delta e_1^{ID}$ , this inequality is equivalent to  $\Delta e_1^{ID} \geq \Delta e_1^*(\hat{\beta}, \hat{\beta})$ , thus contradicting our earlier result. ■

### Proof of Corollary 3

This proof assumes that  $\alpha$  is sufficiently small such that  $P_\alpha(\beta^*(\tilde{\alpha}), \beta^*(\tilde{\alpha}), \tilde{\alpha})$  is increasing in  $\tilde{\alpha}$  for all  $\tilde{\alpha} < \alpha + \Delta e_1^{II}$ . An interval of such  $\alpha$  values exists because i) (19) shows that  $P_\alpha(\beta, \beta, \tilde{\alpha})$  is increasing in  $\tilde{\alpha}$ , ii)  $\lim_{\alpha \rightarrow 0} \Delta e_1^{II} = 0$ , so  $\lim_{\alpha \rightarrow 0} \tilde{\alpha} = 0$ , and iii) as shown in the proof of Proposition 2,  $\lim_{\tilde{\alpha} \rightarrow 0} \frac{\partial \beta^*}{\partial \tilde{\alpha}} = 0$ .

**Part (i)** It follows from (17) that  $\beta_A^{ID} = \beta_A^*(\alpha + \Delta e_1^{ID}) < \beta^*(\alpha + \Delta e_1^{ID}) < \beta_B^*(\alpha +$

$\Delta e_1^{ID} = \beta_B^{ID}$ . Therefore, since  $P_\alpha$  is increasing in  $\beta_A$  but decreasing in  $\beta_B$ , it holds that  $P_\alpha(\beta_A^{ID}, \beta_B^{ID}, \alpha + \Delta e_1^{ID}) < P_\alpha(\beta^*(\alpha + \Delta e_1^{ID}), \beta^*(\alpha + \Delta e_1^{ID}), \alpha + \Delta e_1^{ID})$ . Furthermore (A25), along with the assumption that  $P_\alpha(\beta^*(\tilde{\alpha}), \beta^*(\tilde{\alpha}), \tilde{\alpha})$  is increasing in  $\tilde{\alpha}$  for all  $\tilde{\alpha} < \alpha + \Delta e_1^{II}$ , implies that  $P_\alpha(\beta^*(\alpha + \Delta e_1^{ID}), \beta^*(\alpha + \Delta e_1^{ID}), \alpha + \Delta e_1^{ID}) < P_\alpha(\beta^*(\alpha + \Delta e_1^{II}), \beta^*(\alpha + \Delta e_1^{II}), \alpha + \Delta e_1^{II}) = P_\alpha(\beta^{II}, \beta^{II}, \alpha + \Delta e_1^{II})$ . Hence,  $P_\alpha(\beta_A^{ID}, \beta_B^{ID}, \alpha + \Delta e_1^{ID}) < P_\alpha(\beta^{II}, \beta^{II}, \alpha + \Delta e_1^{II})$ .

**Part (ii)** Since when  $\alpha = 0$ , the expected utility difference (20) is equal to 0 both for ID and II biases, we will prove the claim for small  $\alpha$  by showing that

$$\lim_{\alpha \rightarrow 0} \frac{d(\Delta U^{ID} - \Delta U^{II})}{d\alpha} < 0. \quad (\text{A27})$$

For each bias regime (evaluating at equilibrium values of the biases and  $\Delta e_1$ ), we have

$$\lim_{\alpha \rightarrow 0} \frac{d}{d\alpha} \Delta U(\beta_A, \beta_B, \alpha + \Delta e_1) = 2 \frac{dP_\alpha}{d\alpha} \Big|_{\alpha=0} - c_1 e_1^* \frac{\partial \Delta e_1}{\partial \alpha} \Big|_{\alpha=0}. \quad (\text{A28})$$

Here we used that costs are quadratic and  $\lim_{\alpha \rightarrow 0} \Delta e_1 = 0$ . The common limiting value of effort,  $e_1^*$ , is the same in both bias regimes and solves (A19) and (A20) for  $\alpha \rightarrow 0$ :

$$c_1 e_1^* = g(h)[G(h + \beta^*(0)) - G(h - \beta^*(0))], \quad (\text{A29})$$

where  $\beta^*(0)$  denotes the equilibrium bias, as  $\alpha \rightarrow 0$ , in both bias regimes.

The proof of Proposition 2 showed that  $\lim_{\alpha \rightarrow 0} \frac{\partial \beta_A^*}{\partial \alpha} = -\lim_{\alpha \rightarrow 0} \frac{\partial \beta_B^*}{\partial \alpha} < 0$  and that  $\lim_{\alpha \rightarrow 0} \frac{\partial \beta^*}{\partial \alpha} = 0$ . Using these results, as well as (18), (19), and (A29), to evaluate (A28) for each bias regime, we can express the left-hand side of (A27), after simplifying, as

$$\begin{aligned} \lim_{\alpha \rightarrow 0} \frac{d(\Delta U^{ID} - \Delta U^{II})}{d\alpha} &= g(h)[G(h + \beta^*(0)) - G(h - \beta^*(0))] \frac{d(\Delta e_1^{ID} - \Delta e_1^{II})}{d\alpha} \Big|_{\alpha=0} \\ &\quad + 2[G(h)g(h + \beta^*(0)) + G(-h)g(-h + \beta^*(0))] \frac{d\beta_A^*}{d\alpha} \Big|_{\alpha=0}. \end{aligned}$$

This whole expression is strictly negative, because Proposition 2 implies that  $\Delta e_1^{ID} - \Delta e_1^{II}$ , which is 0 at  $\alpha = 0$ , is non-increasing in  $\alpha$  for small  $\alpha$ , and because  $\frac{d\beta_A^*}{d\alpha} \Big|_{\alpha=0} < 0$ . ■

## REFERENCES

- ACEMOGLU, D. AND D. AUTOR (2011): “Skills, Tasks and Technologies: Implications for Employment and Earnings.” in *Handbook of Labor Economics*, ed. by D. Card and O. Ashenfelter, Elsevier, vol. 4, 1043–1171.
- ACEMOGLU, D., D. AUTOR, J. HAZELL, AND P. RESTREPO (2022): “Artificial Intelligence and Jobs: Evidence from Online Vacancies,” *J. Labor Econ.*, 40, S293–S340.

- ALESINA, A. AND G. M. ANGELETOS (2005): “Fairness and redistribution,” *Am. Econ. Rev.*, 95, 960–980.
- ALESINA, A., S. STANTCHEVA, AND E. TESO (2018): “Intergenerational Mobility and Preferences for Redistribution,” *Am. Econ. Rev.*, 108, 521–554.
- ALMÅS, I., A. W. CAPPELEN, AND B. TUNGODDEN (2020): “Cutthroat capitalism versus cuddly socialism: Are americans more meritocratic and efficiency-seeking than Scandinavians?” *J. Polit. Econ.*, 128, 1753–1788.
- ARIGA, K., Y. OHKUSA, AND G. BRUNELLO (1999): “Fast track: Is it in the genes? The promotion policy of a large Japanese firm,” *Journal of Economic Behavior and Organization*, 38, 385–402.
- ASCH, S. E. (1946): “Forming impressions of personality,” *Journal of Abnormal and Social Psychology*, 41, 258–290.
- AUTOR, D., D. DORN, L. F. KATZ, C. PATTERSON, AND J. VAN REENEN (2020): “The Fall of the Labor Share and the Rise of Superstar Firms,” *The Q. J. Econ.*, 135, 645–709.
- BAGUES, M. F. AND B. ESTEVE-VOLART (2010): “Can Gender Parity Break the Glass Ceiling? Evidence from a Repeated Randomized Experiment,” *The Rev. Econ. Stud.*, 77, 1301–1328.
- BAKER, G., M. GIBBS, AND B. HOLMSTRÖM (1994): “The Internal Economics of the Firm: Evidence from Personnel Data,” *Q. J. Econ.*, 109, 881–919.
- BAO, R., J. DE LOECKER, AND J. EECKHOUT (2026): “Manager Pay Inequality and Market Power,” *Rev. Econ. Stud.*, forthcoming.
- BARDHI, A., Y. GUO, AND B. STRULOVICI (2025): “Early-Career Discrimination: Spiraling or Self-Correcting?” [https://www.arjadabardhi.com/s/ReStud\\_ECD\\_draft45.pdf](https://www.arjadabardhi.com/s/ReStud_ECD_draft45.pdf).
- BEDARD, K. AND E. DHUEY (2006): “The Persistence of Early Childhood Maturity: International Evidence of Long-Run Age Effects,” *The Q. J. Econ.*, 121, 1437–1472.
- BELZIL, C. AND M. BOGNANNO (2008): “Promotions, demotions, halo effects, and the earnings dynamics of American executives,” *J. Labor Econ.*, 26, 287–310.
- BENABOU, R. AND J. TIROLE (2006): “Belief in a Just World and Redistributive Politics,” *Q. J. Econ.*, 121, 699–746.
- BENSON, A., D. LI, AND K. SHUE (2026): ““Potential” and the Gender Promotion Gap,” *Am. Econ. Rev.*, 116, 375–417.
- BLANK, R. M. (2005): “Tracing the Economic Impact of Cumulative Discrimination,” *Am. Econ. Rev.*, 95, 99–103.
- BOHREN, J. A., P. HULL, AND A. IMAS (2025): “Systemic Discrimination: Theory and Measurement,” *Q. J. Econ.*, 140, 1743–1799.
- BOHREN, J. A., A. IMAS, AND M. ROSENBERG (2019): “The dynamics of discrimination: Theory and evidence,” *Am. Econ. Rev.*, 109, 3395–3436.
- CAPPELEN, A. W., J. KONOW, E. Ø. SØRENSEN, AND B. TUNGODDEN (2013): “Just luck: An experimental study of risk-taking and fairness,” *Am. Econ. Rev.*, 103, 1398–1413.
- CARD, D., J. HEINING, AND P. KLINE (2013): “Workplace Heterogeneity and the Rise of West German Wage Inequality,” *Q. J. Econ.*, 128, 967–1015.

- CATTAN, S., K. G. SALVANES, AND E. TOMINEY (2025): “First-Generation Elite: The Role of School Social Networks,” *Am. Econ. Rev.*, 115, 4369–4403.
- CHETTY, R., N. HENDREN, P. KLINE, AND E. SAEZ (2014): “Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States,” *Q. J. Econ.*, 129, 1553–1623.
- CIOCCA-ELLER, C. (2023): “What Makes a Quality College? Reexamining the Equalizing Potential of Higher Education in the United States,” *American Journal of Sociology*, 129, 637–714.
- COLE, S., J. R. COLE, AND G. A. SIMON (1981): “Chance and Consensus in Peer Review,” *Science*, 214, 881–886.
- CONG, L. W. AND Y. XIAO (2022): “Persistent Blessings of Luck: Theory and an Application to Venture Capital,” *Review of Financial Studies*, 35, 1183–1221.
- CORAK, M. (2013): “Income Inequality, Equality of Opportunity, and Intergenerational Mobility,” *J. Econ. Perspect.*, 27, 79–102.
- CURRIE, J. AND D. ALMOND (2011): “Human capital development before age five,” in *Handbook of Labor Economics*, ed. by D. Card and O. Ashenfelter, Elsevier, vol. 4, Part B, ch. 15, 1315–1486.
- DEANER, R. O., A. LOWEN, AND S. COBLEY (2013): “Born at the Wrong Time: Selection Bias in the NHL Draft,” *PLoS ONE*, 8.
- DEMING, D. J. (2021): “The Growing Importance of Decision-Making on the Job,” *NBER Working Paper 28733*.
- DIPRETE, T. A. AND G. M. EIRICH (2006): “Cumulative advantage as a mechanism for inequality: A review of theoretical and empirical developments,” *Annual Review of Sociology*, 32, 271–297.
- DU, Q., H. GAO, AND M. D. LEVI (2012): “The relative-age effect and career success: Evidence from corporate CEOs,” *Economics Letters*, 117, 660–662.
- DURANDARD, T. (2023): “Dynamic delegation in promotion contests,” <https://arxiv.org/abs/2308.05668>.
- EXLEY, C. AND K. NIELSEN (2024): “The Gender Gap in Confidence: Expected But Not Accounted For,” *Am. Econ. Rev.*, 114, 851–885.
- FABBRI, A. AND M. PELLIZZARI (2025): “First-Generation Graduates in OECD Countries,” *CEPR Discussion Paper 20796*.
- FARBER, H. AND R. GIBBONS (1996): “Learning and Wage Dynamics,” *Q. J. Econ.*, 111, 1007–1047.
- FONG, C. (2001): “Social preferences, self-interest, and the demand for redistribution,” *J. Public Econ.*, 82, 225–246.
- FRANK, R. H. (2016): *Success and Luck: Good Fortune and the Myth of Meritocracy*, Princeton University Press.
- FRYER, R. G. (2007): “Belief flipping in a dynamic model of statistical discrimination,” *J. Public Econ.*, 91, 1151–1166.
- GABAIX, X. AND A. LANDIER (2008): “Why has CEO Pay Increased So Much?” *Q. J. Econ.*, 123, 49–100.
- GAMORAN, A. AND R. MARE (1989): “Secondary school tracking and educational in-

- equality: compensation, reinforcement or neutrality?" *American Journal of Sociology*, 94, 1146–1183.
- GIBBONS, R. AND M. WALDMAN (1999): "A Theory of Wage and Promotion Dynamics inside Firms," *Q. J. Econ.*, 114, 1321–1358.
- GILL, D. AND V. PROWSE (2014): "Gender differences and dynamics in competition: The role of luck," *Quantitative Economics*, 5, 351–376.
- GREEN, J. R. AND N. L. STOKEY (1983): "A Comparison of Tournaments and Contracts," *J. Polit. Econ.*, 91, 349–364.
- HAMILTON, B., J. NICKERSON, AND H. OWAN (2003): "Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation," *J. Polit. Econ.*, 111, 465–497.
- HANSEN, S., T. RAMDAS, R. SADUN, AND J. FULLER (2021): "The Demand for Executive Skills," NBER Working Paper, <http://www.nber.org/papers/w28959>.
- LANGE, F. (2007): "The speed of employer learning," *J. Labor Econ.*, 25, 1–35.
- LAVY, V. AND R. MEGALOKONOMOU (2024): "The Short- and the Long-Run Impact of Gender-Biased Teachers," *American Econ. J.: Applied Economics*, 16, 176–218.
- LAZEAR, E. P. (2000): "Performance pay and productivity," *Am. Econ. Rev.*, 90, 1346–1361.
- LAZEAR, E. P. AND S. ROSEN (1981): "Rank-order tournaments as optimum labor contracts," *J. Polit. Econ.*, 89, 841–864.
- LEFGREN, L. J., D. P. SIMS, AND O. B. STODDARD (2016): "Effort, luck, and voting for redistribution," *J. Public Econ.*, 143, 89–97.
- LEPAGE, L.-P. (2024): "Experience-based Discrimination," *American Econ. J.: Applied Economics*.
- LI, D., L. RAYMOND, AND P. BERGMAN (2025): "Hiring as Exploration," *Rev. Econ. Stud.*
- MAS, A. AND E. MORETTI (2009): "Peers at Work," *Am. Econ. Rev.*, 99, 112–45.
- MERTON, R. K. (1968): "The Matthew effect in science," *Science*, 159, 56–62.
- MEYER, M. A. (1991): "Learning from coarse information: Biased contests and career profiles," *Rev. Econ. Stud.*, 58, 15–41.
- MINNI, V. (2025): "Making the Invisible Hand Visible: Managers and the Allocation of Workers to Jobs," [https://virginiaminni.github.io/assets/files/Minni\\_JMP.pdf](https://virginiaminni.github.io/assets/files/Minni_JMP.pdf).
- MURDOCK, B. B. (1962): "The serial position effect of free recall," *Journal of Experimental Psychology*, 64, 482–488.
- NANDA, R., S. SAMILA, AND O. SORENSON (2020): "The persistent effect of initial success: Evidence from venture capital," *J. Financ. Econ.*, 137, 231–248.
- ONUCHIC, P. (2025): "Recent Contributions to Theories of Discrimination," <https://paulaonuchic.com/s/ThDisc2025.pdf>.
- OREOPOULOS, P., T. VON WACHTER, AND A. HEISZ (2012): "The short- and long-term career effects of graduating in a recession," *American Econ. J.: Applied Economics*, 4, 1–29.
- PASTORINO, E. (2024): "Careers in Firms: The Role of Learning about Ability and

- Human Capital Acquisition,” *J. Polit. Econ.*, 132, 1994–2073.
- PIKETTY, T. (2014): *Capital in the Twenty-First Century*, Cambridge, MA: The Belknap Press of Harvard University Press.
- RABIN, M. AND J. L. SCHRAG (1999): “First impressions matter: A model of confirmatory bias,” *Q. J. Econ.*, 114, 37–82.
- ROEMER, J. E. (2000): “Equality of Opportunity,” in *Meritocracy and Economic Inequality*, ed. by K. Arrow, S. Bowles, and S. Durlauf, Princeton (NJ): Princeton University Press, 17–32.
- ROSEN, S. (1981): “The Economics of Superstars,” *Am. Econ. Rev.*, 71, 845–858.
- SANDEL, M. (2020): *The Tyranny of Merit*, New York: Farrar, Straus and Giroux.
- SETHI, R. AND R. SOMANATHAN (2023): “Meritocracy and Representation,” *J. Econ. Lit.*, 61, 941–957.
- ZHANG, S., K. H. WAPMAN, D. B. LARREMORE, AND A. CLAUSET (2022): “Labor advantages drive the greater productivity of faculty at elite universities,” *Science Advances*, 8, 1–22.



**Proof of Proposition 3**

For early-career luck,  $P_0^* = G(\beta_0^*)$  by Corollary 1. So  $P_0^*$  is strictly increasing in  $\beta_0^*$ , and since for  $q > \frac{1}{2}$ , the RHS of (A15) is strictly decreasing in  $q$  and strictly increasing in  $\beta_0^*$ , we have that  $\beta_0^*$  and hence  $P_0^*$  are strictly increasing in  $q$ .

For societal luck, we need to generalize the analysis in Section 4 to arbitrary  $q \geq \frac{1}{2}$ . We first consider the ID case and then use some of its steps for the II case.

**Step 1.** We show first that for small  $h$  and small  $2q - 1 > 0$ , both  $\beta_A^{ID}$  and  $\beta_B^{ID}$  are strictly increasing in  $q$ . To simplify notation, we omit the superscript  $ID$  in  $\beta_A$  and  $\beta_B$ . Following the convention used in Section 4 (as distinct from that used in Section 3), suppose that agent  $A$  is socially advantaged. Since ability is distributed independently of societal advantage, there is a probability  $\frac{1}{2}$  that  $A$  is the ability-favorite and a probability  $\frac{1}{2}$  that  $A$  is the ability-underdog. Denote the first-stage effort differentials in these two cases by  $\Delta e_1^f = e_{A,1}^f - e_{B,1}^f$  and  $\Delta e_1^u = e_{A,1}^u - e_{B,1}^u$ , respectively. Selective efficiency is

$$\begin{aligned} S = & \frac{1}{2}q \left[ G(h + \alpha + \Delta e_1^f)G(h + \beta_A) + (1 - G(h + \alpha + \Delta e_1^f))G(h - \beta_B) \right. \\ & \left. + G(h - \alpha - \Delta e_1^u)G(h + \beta_B) + (1 - G(h - \alpha - \Delta e_1^u))G(h - \beta_A) \right] \\ & + \frac{1}{2}(1 - q) \left[ G(h + \alpha + \Delta e_1^u)G(h + \beta_A) + (1 - G(h + \alpha + \Delta e_1^u))G(h - \beta_B) \right. \\ & \left. + G(h - \alpha - \Delta e_1^f)G(h + \beta_B) + (1 - G(h - \alpha - \Delta e_1^f))G(h - \beta_A) \right]. \end{aligned}$$

Each line here gives the probability of selecting the better agent for one of the four possible cases:  $A$  is the favorite and better than  $B$ ;  $A$  is the underdog and worse than  $B$ ;  $A$  is the underdog but better than  $B$ ;  $A$  is the favorite but worse than  $B$ . Optimal biases solve

$$\begin{aligned} 2 \frac{\partial S}{\partial \beta_A} = & \left[ qG(h + \alpha + \Delta e_1^f) + (1 - q)G(h + \alpha + \Delta e_1^u) \right] g(h + \beta_A) \\ & - \left[ qG(-h + \alpha + \Delta e_1^u) + (1 - q)G(-h + \alpha + \Delta e_1^f) \right] g(h - \beta_A) = 0 \end{aligned} \quad (\text{A30})$$

$$\begin{aligned} 2 \frac{\partial S}{\partial \beta_B} = & \left[ qG(h - \alpha - \Delta e_1^u) + (1 - q)G(h - \alpha - \Delta e_1^f) \right] g(h + \beta_B) \\ & - \left[ qG(-h - \alpha - \Delta e_1^f) + (1 - q)G(-h - \alpha - \Delta e_1^u) \right] g(h - \beta_B) = 0. \end{aligned} \quad (\text{A31})$$

To analyze optimal biases as  $h$  gets small, totally differentiate (A30) and (A31) with respect to  $h$  and let  $h \rightarrow 0$ . We can ignore the effect of  $h$  on  $\beta_A$  and  $\beta_B$  since, similarly to the proof of Proposition 1,  $\lim_{h \rightarrow 0} \frac{\partial \beta_A^{ID}}{\partial h} = \lim_{h \rightarrow 0} \frac{\partial \beta_B^{ID}}{\partial h} = 0$ . Moreover, as  $h \rightarrow 0$ ,  $\Delta e_1^f$  and  $\Delta e_1^u$  converge to the same value  $\Delta e_1^{ID}$ , since it becomes irrelevant whether the advantaged



agent  $A$  is the favorite or the underdog. Starting with (A30), we get

$$\begin{aligned} 2 \lim_{h \rightarrow 0} \frac{\partial^2 S}{\partial \beta_A \partial h} &= g(\alpha + \Delta e_1^{ID}) g(\beta_A) \left( 2 + (2q - 1) \left( \frac{\partial \Delta e_1^f}{\partial h} \Big|_{h \rightarrow 0} - \frac{\partial \Delta e_1^u}{\partial h} \Big|_{h \rightarrow 0} \right) \right) \\ &\quad + 2G(\alpha + \Delta e_1^{ID}) g'(\beta_A) = 0. \end{aligned} \quad (\text{A32})$$

To compute  $\frac{\partial \Delta e_1^f}{\partial h} \Big|_{h \rightarrow 0}$ , consider the analogue of (A3):

$$\frac{C_1'(e_{A,1}^f) - C_1'(e_{B,1}^f)}{C_2(e_2^{fl}) - C_2(e_2^{fw})} = 2 \left[ qg(h + \alpha + \Delta e_1^f) + (1 - q)g(-h + \alpha + \Delta e_1^f) \right], \quad (\text{A33})$$

where  $e_2^{fw}$  and  $e_2^{fl}$  denote the second-stage efforts when the advantaged agent,  $A$ , is the ability-favorite, and won or lost the first stage, respectively. Multiply both sides by  $C_2(e_2^{fl}) - C_2(e_2^{fw})$  and then differentiate with respect to  $h$  to obtain

$$\begin{aligned} c_1 \frac{\partial \Delta e_1^f}{\partial h} &= 2g(\alpha + \Delta e_1^f) \left[ C_2'(e_2^{fl}) \frac{\partial e_2^{fl}}{\partial h} - C_2'(e_2^{fw}) \frac{\partial e_2^{fw}}{\partial h} \right] \\ &\quad + 2 \left[ (2q - 1)g'(\alpha + \Delta e_1^f) + g'(\alpha + \Delta e_1^f) \frac{\partial \Delta e_1^f}{\partial h} \right] [C_2(e_2^{fl}) - C_2(e_2^{fw})], \end{aligned} \quad (\text{A34})$$

where we have used the assumption that  $C_1(e) = \frac{1}{2}c_1 e^2$ . To compute the first term on the RHS of (A34), consider the analogues to (A4) and (A5):

$$C_2'(e_2^{fw}) = q^{fw}g(h + \beta_A) + (1 - q^{fw})g(-h + \beta_A), \quad (\text{A35})$$

$$C_2'(e_2^{fl}) = q^{fl}g(-h - \beta_B) + (1 - q^{fl})g(h - \beta_B), \quad (\text{A36})$$

where  $q^{fw} = q^w(\Delta e_1^f, q)$  and  $q^{fl} = q^l(\Delta e_1^f, q)$  are as defined in (A8) and (A9). Thus,  $\lim_{h \rightarrow 0} C_2'(e_2^{fw}) = g(\beta_A)$  and  $\lim_{h \rightarrow 0} C_2'(e_2^{fl}) = g(\beta_B)$ . Now differentiate (A35-A36) with respect to  $h$  and take the limit as  $h \rightarrow 0$ , recalling that  $\lim_{h \rightarrow 0} \frac{\partial \beta_A^{ID}}{\partial h} = \lim_{h \rightarrow 0} \frac{\partial \beta_B^{ID}}{\partial h} = 0$ :

$$C_2''(e_2^{fw}) \frac{\partial e_2^{fw}}{\partial h} = (2q - 1)g'(\beta_A), \quad (\text{A37})$$

$$C_2''(e_2^{fl}) \frac{\partial e_2^{fl}}{\partial h} = -(2q - 1)g'(\beta_B). \quad (\text{A38})$$

Finally, use  $C_2(e) = \frac{1}{2}c_2 e^2$  and substitute everything into (A34) to get

$$\frac{\partial \Delta e_1^f}{\partial h} \Big|_{h \rightarrow 0} = (2q - 1)\Psi(\beta_A, \beta_B), \quad \text{where} \quad (\text{A39})$$

$$\Psi(\beta_A, \beta_B) = \frac{g'(\alpha + \Delta e_1^{ID}) [g(\beta_B)^2 - g(\beta_A)^2] - 2g(\alpha + \Delta e_1^{ID}) [g(\beta_B)g'(\beta_B) + g(\beta_A)g'(\beta_A)]}{c_1 c_2 - g'(\alpha + \Delta e_1^{ID}) [g(\beta_B)^2 - g(\beta_A)^2]}$$

We now argue that  $\Psi(\beta_A, \beta_B) > 0$ . First, the denominator is strictly positive; this follows from adding the SOC's for the agents' first-stage effort-choice problems. The numerator is also strictly positive because, by Proposition 2,  $\alpha + \Delta e_1^{ID} > 0$ , so  $g'(\alpha + \Delta e_1^{ID}) < 0$ ; and  $\beta_B > \beta_A > 0$ , so  $g^2(\beta_B) - g^2(\beta_A) < 0$ ,  $g'(\beta_B) < 0$ , and  $g'(\beta_A) < 0$ . Analogous steps show that  $\frac{\partial \Delta e_1^u}{\partial h} \big|_{h \rightarrow 0} = -\frac{\partial \Delta e_1^f}{\partial h} \big|_{h \rightarrow 0}$ .

We can then rewrite (A32), the limiting first-order condition for  $\beta_A$ , as

$$\frac{g(\alpha + \Delta e_1^{ID})}{G(\alpha + \Delta e_1^{ID})} [1 + (2q - 1)^2 \Psi(\beta_A, \beta_B)] = L(\beta_A). \quad (\text{A40})$$

For  $\beta_B$ , perform the analogous steps for (A31) to those we used for (A30), giving

$$\frac{g(-\alpha - \Delta e_1^{ID})}{G(-\alpha - \Delta e_1^{ID})} [1 + (2q - 1)^2 \Psi(\beta_A, \beta_B)] = L(\beta_B). \quad (\text{A41})$$

Finally, differentiate (A40) and (A41) with respect to  $q$ , noting that as  $h \rightarrow 0$ ,  $\Delta e_1^{ID}$  is independent of  $q$ . For  $2q - 1$  small, we can ignore terms proportional to  $(2q - 1)^2$ , yielding

$$2 \frac{g(\alpha + \Delta e_1^{ID})}{G(\alpha + \Delta e_1^{ID})} (2q - 1) \Psi(\beta_A, \beta_B) = L'(\beta_A) \frac{\partial \beta_A}{\partial q}, \quad (\text{A42})$$

$$2 \frac{g(-\alpha - \Delta e_1^{ID})}{G(-\alpha - \Delta e_1^{ID})} (2q - 1) \Psi(\beta_A, \beta_B) = L'(\beta_B) \frac{\partial \beta_B}{\partial q}. \quad (\text{A43})$$

Hence  $\frac{\partial \beta_A^{ID}}{\partial q} > 0$  and  $\frac{\partial \beta_B^{ID}}{\partial q} > 0$  for both  $h$  and  $2q - 1 > 0$  sufficiently small.

**Step 2.** We now extend our definition of the persistence of societal luck to  $q \geq \frac{1}{2}$ . In the ID case, the persistence of societal luck, defined by (18) for  $q = \frac{1}{2}$ , generalizes to

$$\begin{aligned} P_\alpha^{ID} = & \frac{1}{2}q \left[ G(h + \alpha + \Delta e_1^f)G(h + \beta_A) + G(-h - \alpha - \Delta e_1^f)G(h - \beta_B) \right. \\ & + G(-h + \alpha + \Delta e_1^u)G(-h + \beta_A) + G(h - \alpha - \Delta e_1^u)G(-h - \beta_B) \\ & + \frac{1}{2}(1 - q) [G(h + \alpha + \Delta e_1^u)G(h + \beta_A) + G(-h - \alpha - \Delta e_1^u)G(h - \beta_B) \\ & \left. + G(-h + \alpha + \Delta e_1^f)G(-h + \beta_A) + G(h - \alpha - \Delta e_1^f)G(-h - \beta_B) \right]. \end{aligned}$$

Using the fact that for  $h \rightarrow 0$ ,  $\Delta e_1^f = \Delta e_1^u = \Delta e_1^{ID}$ , we obtain

$$\lim_{h \rightarrow 0} P_\alpha^{ID} = G(\alpha + \Delta e_1^{ID})G(\beta_A) + G(-\alpha - \Delta e_1^{ID})G(-\beta_B). \quad (\text{A44})$$

Dividing (A42) by (A43) yields  $\frac{L'(\beta_A)}{L'(\beta_B)} \frac{\partial \beta_A / \partial q}{\partial \beta_B / \partial q} = \frac{G(-\alpha - \Delta e_1^{ID})}{G(\alpha + \Delta e_1^{ID})}$ , and we use this, along with the fact that as  $h \rightarrow 0$ ,  $\Delta e_1^{ID}$  is independent of  $q$ , to derive

$$\begin{aligned} \frac{\partial}{\partial q} \lim_{h \rightarrow 0} P_\alpha^{ID} &= G(\alpha + \Delta e_1^{ID}) g(\beta_A) \frac{\partial \beta_A}{\partial q} - G(-\alpha - \Delta e_1^{ID}) g(\beta_B) \frac{\partial \beta_B}{\partial q} \\ &= G(\alpha + \Delta e_1^{ID}) \frac{\partial \beta_A}{\partial q} \left( g(\beta_A) - g(\beta_B) \frac{L'(\beta_A)}{L'(\beta_B)} \right) \propto \frac{\partial \beta_A}{\partial q} \left( \frac{L'(\beta_B)}{g(\beta_B)} - \frac{L'(\beta_A)}{g(\beta_A)} \right). \end{aligned}$$

In the positive domain, when  $L$  is convex,  $\frac{L'(y)}{g(y)}$  is a strictly increasing function. Hence, since  $\frac{\partial \beta_A}{\partial q} > 0$ , as shown above, and  $\beta_B > \beta_A$ , we have  $\frac{\partial}{\partial q} \lim_{h \rightarrow 0} P_\alpha^{ID} > 0$ .

It remains to consider the case of II bias. When the socially advantaged agent  $A$  is the ability-favorite, the posteriors  $q^{fw}$  and  $q^{fl}$  in (A35) and (A36) converge to each other as  $h \rightarrow 0$ , so under II bias, (A35) and (A36) become the same. Hence,  $\lim_{h \rightarrow 0} \Delta e_1^f = 0$ . Similarly, when  $A$  is the ability-underdog,  $\lim_{h \rightarrow 0} \Delta e_1^u = 0$ . Persistence (A44) is then

$$\lim_{h \rightarrow 0} P_\alpha^{II} = G(\alpha) G(\beta_0^{II}) + G(-\alpha) G(-\beta_0^{II}), \quad (\text{A45})$$

where  $\beta_0^{II} \equiv \lim_{h \rightarrow 0} \beta^{II}$ . Thus, for all  $\alpha > 0$ ,  $\lim_{h \rightarrow 0} P_\alpha^{II}$  is strictly increasing in  $\beta_0^{II}$ .

Adding (A32) to the analogous condition for  $\lim_{h \rightarrow 0} \frac{\partial^2 S}{\partial \beta_B \partial h} = 0$ , using (A39) and the fact that  $\lim_{h \rightarrow 0} \frac{\partial \Delta e_1^u}{\partial h} = -\lim_{h \rightarrow 0} \frac{\partial \Delta e_1^f}{\partial h}$ , and setting  $\beta_A = \beta_B = \beta_0^{II}$  yields a fixed-point equation for  $\beta_0^{II}$ :

$$2g(\alpha) = L(\beta_0^{II}) \left[ 1 - \frac{8(2q-1)^2}{c_1 c_2} g(\alpha)^2 g(\beta_0^{II})^2 \right]. \quad (\text{A46})$$

Since the RHS of (A46) decreases in  $q$  for  $q > \frac{1}{2}$  and increases in  $\beta_0^{II}$ ,  $\beta_0^{II}$  increases in  $q$ . Thus  $\lim_{h \rightarrow 0} P_\alpha^{II}$  also increases in  $q$  and, by continuity, the same is true for small  $h$ . ■

## Proof of Proposition 4

Consider first the persistence of early-career luck. The proof focuses on information structures  $I_N$  as described in Section 5.2, but all arguments extend to cases where some (or all) realizations of  $|\Delta x_1|$  are observed precisely. Note that information structures are assumed symmetric with respect to the agents' "names".

In focusing here on early-career luck, we are taking  $\alpha = 0$  (homogeneous identities). Given  $q = \frac{1}{2}$ , consider the equilibrium in which  $\Delta e_1^* = 0$ . Such an equilibrium always exists because if the principal and both agents expect  $\Delta e_1 = 0$ , then the principal's choice of bias and the agents' stage-two best responses will depend only on the "win-classification"  $|\Delta x_1| \in [d_{n-1}, d_n]$  and not on which agent achieved the win. As a result, the stage-one marginal benefit of effort will be the same for the two agents.

Since for  $h \rightarrow 0$ ,  $|\Delta x_1|$  falls in  $[d_{n-1}, d_n]$  with probability  $2[G(d_n) - G(d_{n-1})]$ , we can

use the FOCs (22) determining the limiting biases  $\beta_0^n$  to obtain

$$\mathbb{E}[L(\beta_0)] = \sum_{n=1}^{N+1} 2[G(d_n) - G(d_{n-1})]L(\beta_0^n) = \sum_{n=1}^{N+1} 2[g(d_{n-1}) - g(d_n)] = 2g(0), \quad (\text{A47})$$

where  $d_0 = 0$  and  $d_{N+1} = \infty$ . Hence, the ex ante expectation of  $L(\beta_0)$  always equals  $2g(0)$ . We now show that starting from an arbitrary information structure  $I_N$ , any coarsening  $I_{N-1}$  generates a mean-preserving contraction of the distribution of  $L(\beta_0)$ .

Suppose  $I_{N-1}$  pools two adjacent intervals  $[d_l, d_m)$  and  $[d_m, d_h)$  contained in  $I_N$ . When, under  $I_{N-1}$ , the principal learns that  $|\Delta x_1| \in [d_l, d_h)$ , her optimal bias  $\beta_0^{lh}$  satisfies  $L(\beta_0^{lh}) = \frac{g(d_l) - g(d_h)}{G(d_h) - G(d_l)}$ . Under  $I_N$ , she can distinguish between  $|\Delta x_1| \in [d_l, d_m)$  and  $|\Delta x_1| \in [d_m, d_h)$ , and the corresponding optimal biases  $\beta_0^{lm}$  and  $\beta_0^{mh}$  satisfy  $L(\beta_0^{lm}) = \frac{g(d_l) - g(d_m)}{G(d_m) - G(d_l)}$  and  $L(\beta_0^{mh}) = \frac{g(d_m) - g(d_h)}{G(d_h) - G(d_m)}$ , respectively. Strict log-concavity of  $g$  implies that

$$L(\beta_0^{lm}) < L(\beta_0^{lh}) < L(\beta_0^{mh}).$$

In combination with (A47), these inequalities prove that the distribution of  $L(\beta_0)$  under  $I_{N-1}$  is a mean-preserving contraction of the distribution of  $L(\beta_0)$  under  $I_N$ .

Finally, the persistence of early-career luck is  $P_0^* = \mathbb{E}[G(\beta_0)] = \mathbb{E}[G(L^{-1}(L(\beta_0)))]$ . The assumption that  $L(y)$  is convex for  $y \geq 0$  and the result that all of the values of  $\beta_0$  are strictly positive imply that  $G(L^{-1}(\cdot))$  is strictly concave at all realizations of  $L(\beta_0)$ . Therefore, the mean-preserving contraction in the distribution of  $L(\beta_0)$  resulting from a coarsening of the information structure leads to a strict increase in  $P_0^*$ .

Now, for  $\alpha > 0$ , consider the persistence of societal luck in the regime in which bias can be conditioned on agents' identities. Recall that we use  $I_{ord}$  and  $I_{card}$  to denote the cases of purely ordinal and fully cardinal performance measurement. Under fully cardinal information, for any  $h$ , the principal will in equilibrium perfectly filter out the net advantage of agent  $A$ ,  $\tilde{\alpha}^{card} = \alpha + \Delta e_1^{card}$ , by setting  $\beta_A^{ID}(|\Delta x_1|) = |\Delta x_1| - \tilde{\alpha}^{card}$  and  $\beta_B^{ID}(|\Delta x_1|) = |\Delta x_1| + \tilde{\alpha}^{card}$ . It follows that, under fully cardinal information and for any  $h$ , societal luck has no impact on final selection, that is,  $P_\alpha^{ID}(I_{card}) = \frac{1}{2}$ .

Under purely ordinal information, denote the limiting equilibrium value, as  $h \rightarrow 0$ , of  $A$ 's net advantage by  $\tilde{\alpha}$ . With  $q = \frac{1}{2}$ , the limiting ID biases, as  $h \rightarrow 0$ , are strictly positive and given by  $\beta_A^{ID} = L^{-1}\left(\frac{g(\tilde{\alpha})}{G(\tilde{\alpha})}\right)$  and  $\beta_B^{ID} = L^{-1}\left(\frac{g(-\tilde{\alpha})}{G(-\tilde{\alpha})}\right)$  (see (A40) and (A41)). The same arguments used to prove part (i) of Proposition 2 imply that  $\tilde{\alpha} > 0$ . Define

$H(y) \equiv G(L^{-1}(y)) - \frac{1}{2}$ . Then (A44) can be expressed as

$$\begin{aligned} \lim_{h \rightarrow 0} P_\alpha^{ID}(I_{ord}) &= \frac{1}{2} + G(\tilde{\alpha})H\left(\frac{g(\tilde{\alpha})}{G(\tilde{\alpha})}\right) - G(-\tilde{\alpha})H\left(\frac{g(-\tilde{\alpha})}{G(-\tilde{\alpha})}\right) \\ &= \frac{1}{2} + g(\tilde{\alpha}) \left\{ \frac{H\left(\frac{g(\tilde{\alpha})}{G(\tilde{\alpha})}\right)}{\frac{g(\tilde{\alpha})}{G(\tilde{\alpha})}} - \frac{H\left(\frac{g(-\tilde{\alpha})}{G(-\tilde{\alpha})}\right)}{\frac{g(-\tilde{\alpha})}{G(-\tilde{\alpha})}} \right\}. \end{aligned} \quad (\text{A48})$$

The term in curly brackets is strictly positive because  $H(0) = 0$ ,  $H(y)$  is strictly concave for  $y > 0$  given the convexity of  $L$  on the positive domain, and  $\frac{g(\tilde{\alpha})}{G(\tilde{\alpha})} < \frac{g(-\tilde{\alpha})}{G(-\tilde{\alpha})}$  given that  $g$  is strictly log-concave and  $\tilde{\alpha} > 0$ . Hence for  $h$  small,  $P_\alpha^{ID}(I_{ord}) > \frac{1}{2} = P_\alpha^{ID}(I_{card})$ . ■

### Proof of Proposition 5

For part (i), we have  $P_0^* = G(\beta_0^*)$  (from Corollary 1), so  $P_0^*$  strictly increases in  $\beta_0^*$ . As the RHS of (A15) strictly increases in  $c_t$  for  $q > \frac{1}{2}$ ,  $\beta_0^*$  and  $P_0^*$  strictly decrease in  $c_t$ .

For part (ii), recall that for  $h \rightarrow 0$ ,  $P_\alpha^{ID}$  is given by (A44), with  $\beta_A$  and  $\beta_B$  solving (A40) and (A41). For  $q = \frac{1}{2}$  and  $h \rightarrow 0$ ,  $\Delta e_1^{ID}$  can be obtained from (A33), using the same steps as in the proof of Proposition 3, as the solution to the fixed-point equation

$$\Delta e_1 = \frac{g(\alpha + \Delta e_1)}{c_1 c_2} [g(\beta_B)^2 - g(\beta_A)^2]. \quad (\text{A49})$$

We now show that the solution satisfies  $\frac{\partial \Delta e_1^{ID}}{\partial c_t} > 0$  for  $t = 1, 2$ . Henceforth, write  $\Delta e_1$  for  $\Delta e_1^{ID}$ . Setting  $q = \frac{1}{2}$  and differentiating (A40), (A41), and (A49) with respect to  $c_t$  yields

$$\left[ \frac{g(\alpha + \Delta e_1)}{G(\alpha + \Delta e_1)} \right]' \frac{\partial \Delta e_1}{\partial c_t} = L'(\beta_A) \frac{\partial \beta_A}{\partial c_t}, \quad (\text{A50})$$

$$\left[ \frac{g(-\alpha - \Delta e_1)}{G(-\alpha - \Delta e_1)} \right]' \frac{\partial \Delta e_1}{\partial c_t} = L'(\beta_B) \frac{\partial \beta_B}{\partial c_t}, \quad (\text{A51})$$

$$\begin{aligned} \frac{\partial \Delta e_1}{\partial c_t} &= \left[ -\frac{g(\alpha + \Delta e_1)}{c_t^2 c_{-t}} + \frac{g'(\alpha + \Delta e_1)}{c_t c_{-t}} \frac{\partial \Delta e_1}{\partial c_t} \right] [g(\beta_B)^2 - g(\beta_A)^2] \\ &\quad + \frac{g(\alpha + \Delta e_1)}{c_t c_{-t}} 2 \left[ g(\beta_B) g'(\beta_B) \frac{\partial \beta_B}{\partial c_t} - g(\beta_A) g'(\beta_A) \frac{\partial \beta_A}{\partial c_t} \right]. \end{aligned} \quad (\text{A52})$$

Using (A50) and (A51) to substitute for  $\frac{\partial \beta_A}{\partial c_t}$  and  $\frac{\partial \beta_B}{\partial c_t}$  in (A52), and rearranging, yields

$$\begin{aligned} \frac{\partial \Delta e_1}{\partial c_t} \left\{ 2g(\beta_A)g'(\beta_A) \frac{\left[ \frac{g(\alpha + \Delta e_1)}{G(\alpha + \Delta e_1)} \right]'}{L'(\beta_A)} - 2g(\beta_B)g'(\beta_B) \frac{\left[ \frac{g(-\alpha - \Delta e_1)}{G(-\alpha - \Delta e_1)} \right]'}{L'(\beta_B)} \right. \\ \left. + \left( 1 - \frac{g'(\alpha + \Delta e_1)}{c_t c_{-t}} [g(\beta_B)^2 - g(\beta_A)^2] \right) \right\} = -\frac{g(\alpha + \Delta e_1)}{c_t^2 c_{-t}} [g(\beta_B)^2 - g(\beta_A)^2]. \end{aligned}$$

On the LHS, the first term in the curly brackets is positive since  $\left[ \frac{g(\alpha + \Delta e_1)}{G(\alpha + \Delta e_1)} \right]' < 0$ ,  $g'(\beta_A) < 0$ , and  $L'(\beta_A) > 0$ . Since  $\left[ \frac{g(-\alpha - \Delta e_1)}{G(-\alpha - \Delta e_1)} \right]'$  and  $\left[ \frac{g(\alpha + \Delta e_1)}{G(\alpha + \Delta e_1)} \right]'$  have opposite signs, the second term is also positive. Finally, that the last term in the curly brackets (the one in large parentheses) is positive follows by adding the agents' SOC's for stage-one efforts. The RHS is positive since  $\beta_B > \beta_A > 0$ , so  $g(\beta_B)^2 - g(\beta_A)^2 < 0$ . Hence,  $\frac{\partial \Delta e_1^{ID}}{\partial c_t} > 0$  for  $t = 1, 2$  and  $\alpha > 0$ .

The final step uses the function  $H(y) \equiv G(L^{-1}(y)) - \frac{1}{2}$  and expression (A48) from the proof of Proposition 4. Noting that  $\frac{\partial \tilde{\alpha}}{\partial c_t} = \frac{\Delta e_1^{ID}}{\partial c_t} > 0$ , differentiate (A48) with respect to  $\tilde{\alpha}$ :

$$\begin{aligned} \frac{\partial}{\partial \tilde{\alpha}} \lim_{h \rightarrow 0} P_\alpha^{ID} &= g(\tilde{\alpha})H\left(\frac{g(\tilde{\alpha})}{G(\tilde{\alpha})}\right) + g(-\tilde{\alpha})H\left(\frac{g(-\tilde{\alpha})}{G(-\tilde{\alpha})}\right) \\ &+ G(\tilde{\alpha})H'\left(\frac{g(\tilde{\alpha})}{G(\tilde{\alpha})}\right) \left[ \frac{g(\tilde{\alpha})}{G(\tilde{\alpha})} \right]' + G(-\tilde{\alpha})H'\left(\frac{g(-\tilde{\alpha})}{G(-\tilde{\alpha})}\right) \left[ \frac{g(-\tilde{\alpha})}{G(-\tilde{\alpha})} \right]'. \end{aligned} \quad (\text{A53})$$

Evaluating (A53) at  $\tilde{\alpha} = 0$  yields

$$\begin{aligned} \frac{\partial}{\partial \tilde{\alpha}} \lim_{h \rightarrow 0} P_\alpha^{ID} \Big|_{\tilde{\alpha}=0} &= 2g(0)H\left(\frac{g(0)}{G(0)}\right) - H'\left(\frac{g(0)}{G(0)}\right) 4g(0)^2 \\ &= 2g(0) [H(2g(0)) - H'(2g(0))2g(0)] > 0. \end{aligned}$$

The inequality holds since  $H(0) = 0$  and  $H(y)$  is strictly concave for  $y > 0$ , given the convexity of  $L(y)$  on  $[0, \infty)$ . For small  $\alpha$ ,  $\Delta e_1^{ID}$  and  $\tilde{\alpha}$  are also small, since for  $\alpha = 0$ ,  $\Delta e_1^{ID} = 0$  (Lemma 1). Hence for small  $\alpha$  and  $h$ ,  $P_\alpha^{ID}$  strictly increases in  $c_t$  for  $t = 1, 2$ . ■