SELECTING THE BEST: THE PERSISTENT EFFECTS OF LUCK*

Mikhail Drugov†

Margaret Meyer‡

Marc Möller§

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 analyze a dynamic model of organizational learning in which the organization maximizes the probability of selecting the most able agent ("selective efficiency") and agents compete to become selected by exerting costly efforts. The agents' coarsely-measured performance reflects unobservable ability, privately-chosen effort, and noise. We show that, even when noise swamps underlying ability differences, optimally designed talent-selection processes favor early strong performers. Making early-career luck persistent, e.g. through professional fast-tracks or high-potential programs, is thus rationalized as the outcome of "selecting the best". Moreover, under a mild condition on noise, the role of early luck dominates that of late luck. Organizational selection also affects the persistence of initial advantages ("societal luck") stemming from identity, e.g. race or gender, and 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 early-career luck and societal luck. Both types of luck are more persistent in careers where workers are better informed about relative abilities and where performance measurement is coarser. Stronger competition for selection amplifies the persistence of early-career luck, but dampens the persistence of identity-based advantages.

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.

†Universitat Autònoma de Barcelona and Barcelona School of Economics (BSE), New Economic School, and CEPR, mdrugov@nes.ru.

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

§University of Bern, marc.moeller@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 (Konow, 2000; Fong, 2001; Cappelen et al., 2007; 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).[1] In light of the above debate, it is thus important to understand to what extent 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

[1]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).

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 (Rosenbaum, 1979; Forbes, 1987; Baker et al., 1994) magnify the importance of early performance for final success. If initial success can be attributed at least in part to 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 inducing 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 educational 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 indicating 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).

In this paper, we explain why making luck persistent is not an unwanted consequence of too much or the wrong kind of bias being employed, but may constitute an integral part of an optimally designed organizational learning process. By rationalizing the persistence of luck as an inseparable feature of individual careers and by shedding light on its institutional determinants, our analysis informs 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". The organization attempts to select the more able agent on the basis of the ordinal ranking of performances at each stage.[2] 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 during the initial stage of his career, on the agent's probability of being selected in the final stage. Our analysis distinguishes between 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 is reserved for one agent, the one endowed with the advantageous identity. Conceptually, these two kinds of luck differ in that when luck is associated with an agent's identity, the organization's use of bias, as well as the agents' early actions, may condition it.

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

[2]Ordinal performance measurement arises naturally when performance is hard to quantify. This assumption is relaxed in Section 5.2, where we allow performance measurement to be less coarse than ordinal.

selection productivity may be very sensitive to which agent was selected.[3] We show that in this limit, even though the effort differential converges to zero, so 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, the limiting equilibrium bias depends on the rate at which the difference between the favorite's and the underdog's efforts vanishes. This highlights the fact that bias should be understood as 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, Murdock, 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

[3] 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.

heterogeneous in their identity. Societal luck endows one of the agents with an "advantageous" 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*, which we define 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; 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 they derive from their identity. However, such a conclusion would be premature as it ignores the agents' effort responses: A move from II-bias to ID-biases could have the opposite effects if it were to mute disadvantaged agents' incentives to partially compensate for their disadvantage with higher efforts. A key contribution of Section 4 is to demonstrate 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 even the utility difference between the advantaged and the disadvantaged agent is lower under ID than under II biases, 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 of our model by analyzing how the persistence of early-career luck and societal luck vary with important characteristics of jobs: 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 can be expected to be especially persistent in jobs involving teamwork and management duties. Given the growing prevalence of such jobs in most industries (Deming, 2017, 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; Song et al., 2019; Autor et al., 2020)—further amplifies the persistence of early-career luck but turns out to dampen the persistence of identity-based advantages.

Related literature The relevance of initial luck for individual long-term success has been documented across a variety of 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). Our paper's main contribution is to identify a mechanism rationalizing the persistence of luck as an institutional feature of a meritocratic society aiming to select the best.

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), Gibbons and Waldman (1999, 2006), and Altonji and

Pierret (2001) 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 or by a "fast-track effect" 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). In the seminal models, serial correlation of promotion rates arises from workers' time-invariant ability differences or human capital accumulation. Our analysis shows that, alternatively, serial correlation can be explained by the non-vanishing optimality of fast-tracking (bias).

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".[4] 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 size of 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.[5] 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 (Bartoš et al., 2016; Lepage, 2024), and to persistent discrimination for subgroups

[4]Using Armed Forces Qualification Test scores as measures of unobserved ability, Lange (2007) finds that it takes only 3 years for employers' expectation error about workers' productivity to decline by one half. Similarly, Lluis (2005) finds evidence that employer learning affects mobility between upper and executive levels of German firms but only for workers below 35 years of age.

[5]See Sections 3 and 4 of Onuchic (2025) for an excellent survey of this literature.

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.[6]

A distinguishing feature of our theory is its analysis of organizational learning as a *strategic interaction*—between a principal choosing bias to optimize selection and agents exerting efforts to become selected. More specifically, our analysis combines the "pure" organizational learning model of Meyer (1991) with a dynamic version of the labor-tournament model of Lazear and Rosen (1981). This combination generates the results in Section 5 characterizing the environments where we predict early-career luck and societal luck to be most persistent.

Finally, one component of our analysis in Section 4 of the persistence of societal luck is the recognition that good performance in the face of a disadvantage is particularly informative about an agent's high ability. This feature is also present in the theoretical analysis of Fryer (2007) and is documented in a field experiment by Bohren et al. (2019). It 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 that the effort responses to identity-based policies amplify their beneficial effects with respect to both learning and persistence.

[6] 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 $a_i \in \Re$. 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. The principal's goal is thus simple: to select the more able agent.

Agent i's performance at stage t, $x_{i,t} \in \Re$, 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 $\epsilon_{i,t} \in \Re$. 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, $q \equiv \mathbb{P}(\Delta a = h) \geq \frac{1}{2}$, 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 superior 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*.[7]

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 $\beta \in \Re$ 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 β and the selection rule to maximize *selective*

[7]Virtually all 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 $q = \frac{1}{2}$.

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. The value of being selected is the same for both agents and is normalized to 1. The cost-of-effort functions are assumed quadratic $C_t(e_{i,t}) = \frac{c_t}{2}e_{i,t}^2$ with $c_t > 0, t = 1, 2$.[8] This specification allows effort costs to 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) $\lim_{y \to z} L(y) = \infty$, where*

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

The symmetry of g coupled with identical costs across agents captures the idea that the only source of heterogeneity between agents is the difference in their abilities; it is a weaker assumption than individual shocks, $\epsilon_{i,t}$, being i.i.d. across agents. Log-concavity of g is equivalent to the monotone likelihood ratio property in our setting; it guarantees that, in either stage, a larger performance differential $\Delta x_t \equiv x_{A,t} - x_{B,t}$ implies a higher likelihood that A's ability exceeds B's. It also implies that L is increasing. Strict log-concavity makes all the implications strict. Together with the remaining two assumptions, it ensures that the principal's problem is well-behaved.

[8] 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) and Nalebuff and Stiglitz (1983).

Identity Our analysis distinguishes between two cases. While Section 3 restricts attention to the case where agents have the same identity (e.g. gender, ethnicity, socioeconomic background), in Section 4 we allow agents to differ in this dimension. We assume there that one agent—whose identity is common knowledge—has a relative advantage over the other during the first stage of their careers.[9] More specifically, our analysis in Section 4 assumes that the first-stage performance of the advantaged agent $i \in \{A, B\}$ is increased by a known amount $\alpha \geq 0$ to $x_{i,1} + \alpha$. Because in our model only relative performance matters, an alternative but 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 conditions the bias on the winner's identity and the case where she is required to set an identity-independent bias.

Timing In the beginning of 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. In the beginning of the second stage, the principal chooses the level of bias. Agents observe the bias and then exert 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 both agents and the principal are correct. It is without loss of generality to assume that the principal selects the winner of the second stage, because the selection of the first-stage winner can be implemented through the choice of an infinite bias.

[9] 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.

Interpretation of heterogeneity The key parameter of our model, $h > 0$, which captures the degree of agents' heterogeneity in abilities, also has a broader interpretation as the *ratio* of agents' heterogeneity to the scale of noise.[10] 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. 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 the effect of luck at later stages of individual careers? In this section, we analyze the *interplay* between agents aiming to become 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. Our analysis here abstracts from advantages or disadvantages agents derive from differences in their identities ($\alpha = 0$). To isolate the long-term effects of luck, we focus on the limiting case where noise swamps ability and effort differences in determining outcomes ($h \to 0$).

3.1 Equilibrium

We begin by considering the agents' effort choices. If, in the first stage, agents anticipate that the principal will, in the second stage, employ a bias in favor of the first-stage

[10] To see this, introduce a scaling transformation $\Delta\epsilon_t \to \sigma\Delta\epsilon_t$, with $\sigma > 1$, which makes the difference in the noise terms more dispersed: The cdf becomes $G(\frac{\Delta\epsilon_t}{\sigma})$, the pdf $\frac{1}{\sigma}g(\frac{\Delta\epsilon_t}{\sigma})$, and the support $[-\sigma z, \sigma z]$. If the underlying heterogeneity in abilities is H, then $G(\frac{H}{\sigma})$ 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 σ only through the *heterogeneity-to-noise ratio* $h \equiv \frac{H}{\sigma}$.

winner ($\beta > 0$), agents 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 his 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.*

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

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

The agents work equally hard in the second stage, because in that stage, the marginal benefit of effort is the same for the two agents, despite the asymmetry arising from the use of bias and the agents' updating of beliefs. 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 for simplicity that $q = 1$, so that 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, we consider the principal's choice of bias. Recall that the principal chooses bias in the second stage, without having observed the agents' first-stage efforts and without knowing which agent was the favorite. Let Δe_1 be the principal's conjecture about the first-stage effort differential. The first-stage effort differential matters to the principal because it influences the probability

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

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 β to maximize selective efficiency, defined as the probability that the more able agent is selected as the winner of the second stage. Selective efficiency is given by

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

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 first-order condition corresponding to (4) we obtain the following:

Lemma 2 (Bias) *For any conjectured first-stage effort differential* $\Delta e_1 \geq 0$*, the princi-*

pal's optimal bias $\beta^*(\Delta e_1; h)$ is strictly positive and given by the unique solution to

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

The first-order condition (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 left-hand side measures the informativeness of a first-stage win by the relative likelihood that a first-stage win is achieved by the more able agent rather than the less able agent. The ratio on the right-hand side measures the informativeness of an agent's achieving a hypothetical second-stage draw, when disadvantaged by bias β, 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.[11]

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

We now turn to our main interest, the persistence of early-career luck. For this purpose, we combine Lemmas 1 and 2 to examine the probability, in an equilibrium

[11] When $q = \frac{1}{2}$, so agents are as uninformed about their relative abilities as the principal, Lemma 1 shows that $\Delta e_1 = 0$. In this case, $\kappa(\Delta e_1; h)$ simplifies to $G(h)$, and (5) reduces to the first-order condition determining optimal bias in the pure learning model of Meyer (1991).

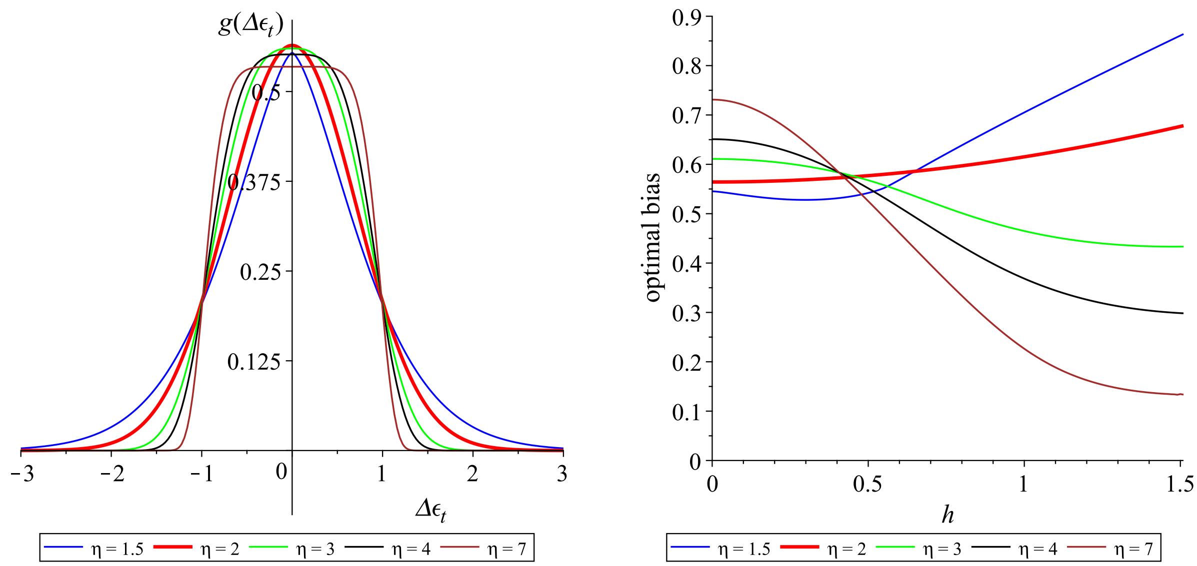

Figure 1: **Example distributions of noise and optimal bias.** The left panel depicts the density functions when noise follows an exponential power distribution $g(\Delta\epsilon_t;\eta) = \frac{\eta}{2\Gamma(\frac{1}{\eta})}\exp(-|\Delta\epsilon_t|^{\eta})$, with mean zero and shape parameter $\eta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

$(\Delta e_1^*(h), \beta^*(h))$, that the agent who wins the first stage is ultimately selected by the principal as the winner of the second stage. To ensure that we focus on the persistent effects of luck, as distinct from those of ability differences, we examine the limit as noise swamps ability differences ($h \to 0$).

In our setting, an equilibrium $(\Delta e_1^*(h), \beta^*(h))$ consists of a first-stage effort differential and a bias that constitute mutual "best responses" in the sense 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 h, the heterogeneity in agents' abilities, goes to 0, the first-stage effort differential vanishes, for any q and any anticipated bias, and hence from (3) the informativeness of a first-stage win converges to zero (the LHS of (5) converges to one). One might be tempted to conclude that, as $h \to 0$, the principal's choice of bias converges 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 $h \to 0$, the informativeness of a second-stage draw, against any level of bias, *also* converges to zero (the RHS of (5) also converges to one). Consequently, 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 in the limit as $h \to 0$.

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

$$\frac{\partial \Delta e_1^*(\beta;0)}{\partial h} \equiv \lim_{h\to 0} \frac{\partial \Delta e_1^*(\beta;h)}{\partial h} = -4g(0)g(\beta)g'(\beta)\frac{1}{c_1c_2}(2q-1) \geq 0. \tag{6}$$

The proposition below characterizes the limiting value of bias that arises as part of an equilibrium $(\Delta e_1^*(h), \beta^*(h))$ when $h \to 0$.

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

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

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-1 effort differential tends to 0 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 $h \to 0$ 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:

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

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 $h \to 0$ in making the following definition:

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

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

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' informational advantage 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 $\lim_{h\to 0} \Delta e_1^*(h) = 0$, 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

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

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

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

It is important to recognize that D_0^* depends on the size of the limiting equilibrium bias β_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 he is 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.

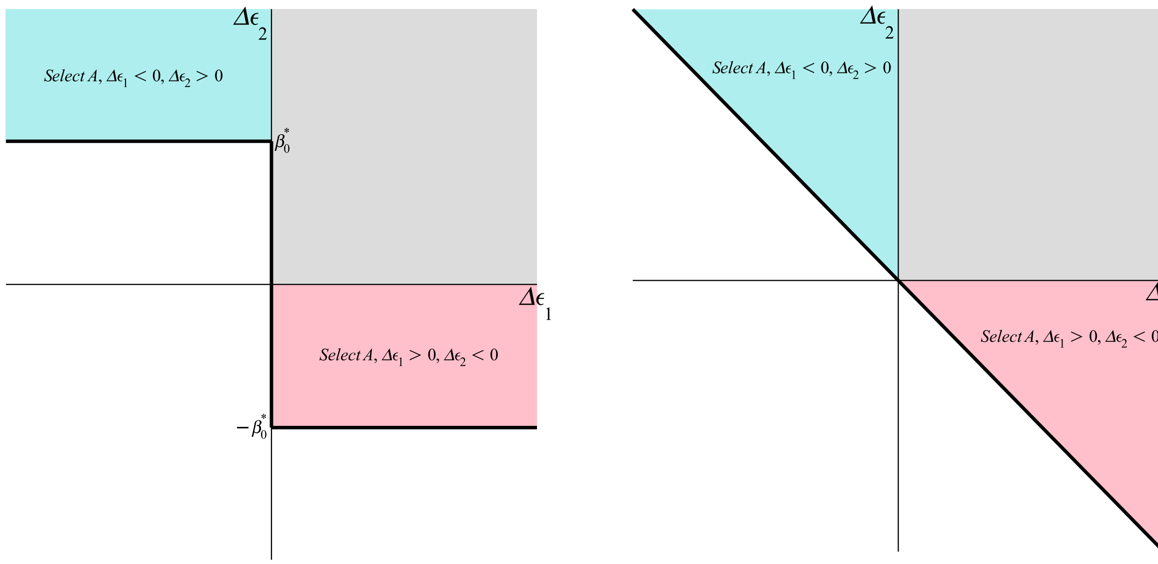

Figure 2: **Dominance.** Selection in the equilibrium described by Proposition 1 (left panel) and in the hypothetical selection process based on winning margins (right panel). The bold line separates the area where agent A is selected (colored) from the area where agent B is selected (white). Both processes 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 the figure, these probabilities are readily determined, and we get:

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

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 $h \to 0$, 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 Δx_1 and Δx_2 differ in sign, while selecting the agent who wins twice if Δx_1 and Δx_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, 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 β_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 $2g$, and $g(z) = 0$ by Assumption 1,

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

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 distribution, 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 restricted to observing rank-order information to implement a bias larger than what the average bias would be if bias could be based 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

Our analysis in Section 3 highlights the relevance of early-career luck for an individual's long-term success. The "luck" on which we have focused so far 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 exists 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 social and economic 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 allow agents to differ in their identities, by assuming 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 uncorrelated with 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 advantages individuals derive from their identities. For this purpose, we focus 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 impact on agents' behavior of the commonly observed societal advantage α from that of agents' information (as captured by $q > \frac{1}{2}$).[12] Hence there exists neither a favorite nor an underdog (in the sense used in Section 3), and we can assume, without loss of generality, that agent A receives the advantage α. We distinguish two scenarios. Under *identity-dependent* (ID) biases, the principal can condition her second-stage bias β_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.[13] Alternatively, there may be behavioral reasons why advantages are not accounted for, even when they are known to exist.[14] Our objective 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)

[12] Our result in Corollary 3 generalizes to arbitrary q, for small values of the exogenous advantage. See also Section 5.1 for the impact of q on the persistence of societal luck.

[13] 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."

[14] 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 α, optimal II bias becomes insensitive to α, so that our analysis of the II-case approximates the case where an advantage exists but is neglected by the principal.

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 Δe_1^{ID} and Δe_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 β_A and β_B, and net advantage $\tilde{\alpha}$, selective efficiency can be written as:

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

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 β_A^* and β_B^* solve the first-order conditions

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

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

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

Equations (15) and (16) are the analogues of (5), and their comparison shows 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 with a net advantage $\tilde{\alpha}$ or against it. From the log-concavity of g, it thus follows that for all $\tilde{\alpha} > 0$:

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

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 a characterization of the agents' 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) Independently of whether 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:

$$-\alpha < \Delta e_1^{II} < 0 \qquad \textit{and} \qquad -\alpha < \Delta e_1^{ID} < 0.$$

(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.

Similarly to Section 3, where for $q > \frac{1}{2}$ the first-stage competition was asymmetric due to agents' information about their relative abilities, with societal luck the agents' first-stage effort differential also arises exclusively from the impact on second-stage effort costs of which agent wins the first stage. 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 Δe_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 β_A makes the second-stage competition even more balanced following a win by A, and an increase in β_B makes the second-stage competition even less balanced following a win by B.

We are now ready to 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 she will do better in the equilibrium under ID biases, even though she is less constrained in this regime 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 here, however, is in the comparison across regimes of the persistence of societal luck, given by the probability P_α that the initially advantaged agent is ultimately selected:

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

Under II bias, societal luck is *always* made persistent in equilibrium, i.e. it always holds that $P_\alpha(\beta^{II}, \beta^{II}, \alpha + \Delta e_1^{II}) > \frac{1}{2}$. This is most easily seen by noting that, for $\beta_A = \beta_B = \beta$, persistence can be written more simply 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 in equilibrium, both the net advantage, $\tilde{\alpha}^{II} = \alpha + \Delta e_1^{II}$, and the principal's choice of bias, β^{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 striking contrast, allowing for ID biases may completely eliminate the persistence of societal luck. For example, we can show that, 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 $\alpha > 0$.[15]

The following corollary to Proposition 2 provides a general comparison of the persistence of societal luck between the equilibria in the II and ID regimes. It also compares these equilibria with respect to the expected utility difference ΔU between the advantaged and the disadvantaged agent:

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

Corollary 3 (Persistence of societal luck) *Under the assumptions of Proposition 2*

[15] The logistic distribution does not satisfy part (iv) of Assumption 1, yet all our results remain valid.

(ii) and for exogenous advantage $\alpha > 0$ sufficiently small, allowing bias to be identity-dependent

(i) reduces the persistence of societal luck in equilibrium:

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

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

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

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_α in (18) is increasing in β_A and decreasing in β_B. This reduction in persistence 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 α is sufficiently small, because for small α, a reduction in $\tilde{\alpha}$ is guaranteed to lower $P_\alpha(\beta^*(\tilde{\alpha}), \beta^*(\tilde{\alpha}), \tilde{\alpha})$.[16] In fact, part (ii) of Corollary 3 shows that this effort response reduces persistence by so much that even the utility difference between the advantaged and the disadvantaged agent is lower under ID than under II biases.

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

In summary, 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—for instance, 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 the selection into final positions, such as the licensing of surgeons. The first part of Corollary 3 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, the second part of Corollary 3 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 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 question, we examine the comparative statics of our model with respect to the parameter q. Recall that for $q = \frac{1}{2}$, agents are as uninformed as the principal about their relative abilities, whereas for $q = 1$, agents know perfectly who is more able. An increase in q augments the agents' informational advantage relative to the principal.

Throughout Section 5, for simplicity and to facilitate comparisons between the persistence of early-career luck and that of societal luck, we will focus on the setting where h is small when analyzing both forms of persistence.[17]

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, P_0^* is strictly increasing in q, P_α^{II} is strictly increasing in q for small h, and P_α^{ID} is strictly increasing in q for small q and h if $L(y)$ is convex for $y \geq 0$.*

Proposition 3 highlights that the persistence of luck is the result of an equilibrium inter-

[17] It can be shown that in equilibrium, both forms of persistence are locally insensitive to h as h rises from 0.

action: Luck is made more persistent in careers where organizations bias selection more strongly in favor of early winners to capitalize on agents' strategic responses to their knowledge of relative abilities. Agents' information about relative abilities is improved by direct interaction or collaboration and by the observation of factors that affect performance but are unobservable to the principal. Empirical studies suggest that agents' informational advantage might be substantial when workers interact in teams, where the observation of co-workers' abilities and efforts constitute the basis 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 involving higher degrees of teamwork, which is in line with the casual observation that in highly collaborative industries (e.g. science, management consulting), biases in the form of grants and fast tracks are abundant.[18]

To understand Proposition 3, first note that the persistence $P_0^* = G(\beta_0^*)$ of early-career luck is strictly increasing in β_0^*. The fixed-point equation (7), which determines equilibrium bias in the limit, shows that β_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 agent, and also, 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 for h small, and hence increase the limiting equilibrium bias.

For societal luck, Proposition 3 generalizes the analysis in Section 4 to allow for an arbitrary $q \geq \frac{1}{2}$. The agents' first-stage effort differential will now in general depend on whether the favorite or the underdog received the societal advantage α. Nevertheless, under II bias, as h gets small, both of these effort differentials converge to zero. Hence, from (18), the limiting persistence of societal luck under II bias is $\lim_{h\to 0} P_\alpha^{II} =$

[18] 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.

$G(\alpha)G(\beta_0^{II}) + G(-\alpha)G(-\beta_0^{II})$, and for any $\alpha > 0$, $\lim_{h\to 0} P_\alpha^{II}$ is increasing in the limiting equilibrium bias. This bias, β_0^{II}, solves a fixed-point equation (equation (A46) in the Appendix) analogous to (7), which shows that when agents are better informed about their relative abilities, β_0^{II} increases, because of the same two forces that cause β_0^* to increase in q.

Finally, for the case of ID bias, the limiting persistence of societal luck is $\lim_{h\to 0} P_\alpha^{ID} = G(\tilde{\alpha})G(\beta_{A0}^{ID}) + G(-\tilde{\alpha})G(-\beta_{B0}^{ID})$. For q sufficiently close to $\frac{1}{2}$, both β_{A0}^{ID} and β_{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 biases rises with q if the noise distribution is normal or has thinner tails (L convex), because this condition limits the rate at which β_{B0}^{ID} rises with q relative to β_{A0}^{ID}.

5.2 Coarseness of performance measurement

How does the principal's ability to measure agents' performance affect the persistence of luck? In this section we extend our model to allow performance measurement to be less coarse than ordinal. For simplicity, we focus on the case where agents are as uninformed as the principal ($q = \frac{1}{2}$) and on the regime of identity-dependent biases. For any strictly increasing series of cutoffs $d_n > 0$, let $I_N = \{[0, d_1), [d_1, d_2), \ldots, [d_{N-1}, d_N), [d_N, \infty)\}$ define an information partition. The principal observes not only the sign of Δx_1 but also which interval $[d_{n-1}, d_n)$ contains the winning margin $|\Delta x_1|$, and she can condition the bias on this interval. We say that I_{N-1} is a *coarsening* of I_N when I_{N-1} arises from I_N via the merging of two adjacent intervals. With a slight abuse of notation, we let I_{ord} and I_{card} denote the extreme cases of purely ordinal and fully cardinal performance measurement.

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

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

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

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

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 $h \to 0$, $|\Delta x_1|$ 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

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

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 the measurement of agents' relative performance is coarser. Formally, given any information structure 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 of the principal's performance measurement 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 involving a high degree of management duties. Moreover, an increase in automation including the recent rise of AI shifts task and job composition towards more managerial, team-oriented, and non-routine activities (Deming, 2017, 2021; Acemoglu et al., 2022). Our results thus imply that the increasing prevalence of such difficult-to-measure activities should lead to an increasing role of initial luck in labor markets.

Proposition 4 resonates well with the fact that fast tracks abound in the identification of leadership potential both in 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.[19] 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 ex ante expectation of $L(\beta_0)$ is a constant, $2g(0)$:

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

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.[20] 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, and hence any coarsening of the principal's information structure leads to an increase in the (expected) persistence of

[19]For a description of these programs see `https://graduateships.com/companies/unilever/` and `https://www.psd.gov.sg/leadership/public-service-leadership-careers`.

[20]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.

early-career luck.[21]

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, $\lim_{h\to 0} P_\alpha^{ID}(I_{ord}) = G(\tilde{\alpha})G(\beta_{A0}^{ID}) + G(-\tilde{\alpha})G(-\beta_{B0}^{ID})$, where $\beta_{B0}^{ID} > \beta_{A0}^{ID}$. Convexity of L, as noted in Section 5.1, limits the size of β_{B0}^{ID} relative to β_{A0}^{ID} and is here sufficient to ensure that $\lim_{h\to 0} P_\alpha^{ID}(I_{ord}) > \frac{1}{2}$. Hence, organizational selection makes societal advantages more persistent when performance measurement is ordinal rather than cardinal.[22]

5.3 Intensity of competition

Does initial luck have a stronger or a weaker effect on final success in careers that are more competitive? We show here that, in contrast to our findings in Sections 5.1 and 5.2, the direction in which the persistence of luck responds to more intense competition for selection depends on the kind of luck under consideration. We interpret an increase in competition as an increase in agents' incentives to exert effort, which can arise either from a reduction in their marginal costs of effort, or from an increase in their value of being selected. Because we have normalized their value of being selected, in what follows 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:*

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

[22]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.

(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 α, P_α^{ID} is strictly increasing in c_t if $L(y)$ is convex for $y \geq 0$.

To understand the first part of Proposition 5, remember from our analysis of early-career luck in Sections 3 and 5.1 that, when agents are better informed about their relative ability than the principal. i.e. $q > \frac{1}{2}$, their effort choices make their first-stage performance ranking more aligned with their ability ranking. More specifically, the proportionality of (6) with respect to $\frac{1}{c_1 c_2}$ shows that intensifying competition—by increasing the value of being selected relative to the costs of effort—makes the *positive* effort differential between the favorite and the underdog rise faster with heterogeneity h in the limit as $h \to 0$. Consequently, as competition becomes more intense, (7) shows that organizations will implement larger biases as noise swamps ability differences, thus making early-career luck more persistent.

For societal luck, greater competition between uninformed agents ($q = \frac{1}{2}$) has the opposite effect on persistence, because as we have shown in Section 4, the disadvantaged agent will partially compensate for his disadvantage by working harder than his rival, i.e. the effort differentials Δe_1^{ID} and Δe_1^{II} are *negative*. Intensifying competition will amplify this effect, further reducing the net advantage of the advantaged agent. For ID biases, P_α^{ID} is increasing in net advantage for small α, as long as h is small and L is convex, so under these conditions agents' effort response to greater competition reduces P_α^{ID}.[23]

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 in-

[23]For II bias, it can be shown that agents' effort response to increased competition disappears as h gets small, but for arbitrary h, P_α^{II} will fall as competition intensifies, as long as α is small enough to guarantee that P_α^{II} is increasing in net advantage. (As noted in footnote 16, as $\alpha \to \infty$, the principal will ignore the first-stage outcome, so the persistence of societal luck will eventually decrease to $\frac{1}{2}$.)

creasing 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; Song et al., 2019). 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.[24]

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 (Heckman, 2007; 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.

[24]Spain's selection mechanism for prosecutors and judges is an example where workers invest considerable and growing resources during the early stages of their careers. Between 2003 and 2007 more than 20,000 law graduates spent an average of three years in full-time preparation for a yearly, centralized, oral examination ("oposiciónes"), with fewer than 10% of candidates becoming selected (Bagues and Perez-Villadoniga, 2012).

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 the selection of the most gifted. Findings such as these 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 the impact of initial luck on final success can be expected to be most amplified. This happens when agents are well informed about their relative abilities and the organization is restricted to use ordinal rather than cardinal performance information. To the extent that the rise of teamwork (Deming, 2017) and the proliferation of management duties (Deming, 2021) have improved workers' information about co-workers' abilities while making the measurement of individual contributions and relative performance more difficult, our theory suggests an increasing persistence of luck in individual careers. Although more intense competition for top jobs and for jobs at superstar employers contributes to this trend by making early-career luck more persistent, the incentive effects of greater competition can be helpful in mitigating advantages associated with agents' identities.

Appendix

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 Δe_1.

We first show that agents exert identical effort in the second stage and that this holds *independently* of q and Δe_1. In case w, A's and B's second-stage efforts solve:

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

By the symmetry of g, 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 $\beta > 0$. Overall utility of agent A is:

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

A change in $e_{A,1}$ does not affect $e_{B,2}^w$, $e_{B,2}^l$, or β, because it is unobservable, and the local effect via the induced changes in $e^w_{A,2}$ and $e^l_{A,2}$ is zero by the envelope theorem. Using $\Delta e^w_2 = \Delta e^l_2 = 0$ and the symmetry of g, the first-order condition for $e_{A,1}$ is

$$\begin{aligned} C_1'(e_{A,1}) &= [qg\,(h+\Delta e_1) + (1-q)g\,(-h+\Delta e_1)] \\ &\quad \cdot \left\{[G\,(h+\beta) - G\,(h-\beta)] - \left[C_2\,(e^w_2) - C_2\left(e^l_2\right)\right]\right\}. \end{aligned} \tag{A1}$$

Analogously, for agent B the first-order condition for $e_{B,1}$ can be written as

$$\begin{aligned} C_1'(e_{B,1}) &= [(1-q)g\,(h-\Delta e_1) + qg\,(-h-\Delta e_1)] \\ &\quad \cdot \left\{[G\,(h+\beta) - G\,(h-\beta)] + \left[C_2\,(e^w_2) - C_2\left(e^l_2\right)\right]\right\}. \end{aligned} \tag{A2}$$

By symmetry of g, and noting that the marginal benefit stemming from the enhanced probability of selection is identical across agents, subtracting (A2) from (A1) gives

$$\frac{C_1'(e_{A,1}) - C_1'(e_{B,1})}{C_2(e^l_2) - C_2(e^w_2)} = 2\,[qg\,(h+\Delta e_1) + (1-q)g\,(-h+\Delta e_1)]\,. \tag{A3}$$

Given that costs are strictly increasing and strictly convex, we conclude that in equilibrium, $\Delta e_1 = e_{A,1} - e_{B,1}$ and $e^l_2 - e^w_2$ must have the same sign.

To determine the sign of $e^l_2 - e^w_2$, compare agent A's first-order conditions for second-stage effort, after a first-stage win by A vs. after a first-stage loss by A, respectively:

$$C_2'(e^w_2) = q^w(\Delta e_1, q)g(h+\beta) + (1 - q^w(\Delta e_1, q))g(-h+\beta), \tag{A4}$$

$$C_2'(e^l_2) = q^l(\Delta e_1, q)g(-h-\beta) + (1 - q^l(\Delta e_1, q))g(h-\beta). \tag{A5}$$

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

$$C_2'(e^w_2) - C_2'(e^l_2) = [q^w(\Delta e_1, q) - q^l(\Delta e_1, q)][g(h+\beta) - g(-h+\beta)]. \tag{A6}$$

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,

$$e_2^l - e_2^w \gtrless 0 \quad \Longleftrightarrow \quad q^w(\Delta e_1, q) - q^l(\Delta e_1, q) \gtrless 0. \tag{A7}$$

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

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

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

Observe that q^w and q^l are, respectively, strictly decreasing and strictly increasing in Δe_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 the 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 $q = \frac{1}{2}$, in an equilibrium $e_{A,1} = e_{B,1}$.

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

Eq. (5) is the first-order condition of (4). To check the second-order condition 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

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

Hence, the left-hand side of (5), $\frac{\kappa(\Delta e_1;h)}{1-\kappa(\Delta e_1;h)} > 1$. The right-hand side, $\frac{g(h-\beta)}{g(h+\beta)}$, equals 1 at $\beta = 0$, is strictly increasing in β by strict log-concavity of g, and goes to $+\infty$ for $\beta \to z - h$. 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

This proof characterizes the principal's equilibrium choice of bias in the limit $h \to 0$, denoted as $\beta_0^* = \lim_{h\to 0} \beta^*(h)$. We start by considering agents' efforts.

For $h \to 0$, (A6) implies that $(e_2^w - e_2^l) \to 0$, making the right-hand sides of (A1) and (A2) become equal. Hence, in the limit, first-stage efforts converge: $\lim_{h\to 0} \Delta e_1^*(h) = 0$.

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

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

Since $\lim_{h\to 0} \Delta e_1^* = 0$, $\lim_{h\to 0} S_\beta(\beta, \Delta e_1; h) = g(\beta) - g(-\beta) = 0$ for all β. Hence, to characterize β_0^* we need to differentiate $S_\beta(\beta^*(h), \Delta e_1^*(h); h) = 0$ totally with respect to h and take the limit $h \to 0$. Because $\lim_{h\to 0} S_\beta(\beta, \Delta e_1^*; h) = 0$ for all β, $\lim_{h\to 0} S_{\beta\beta}(\beta^*, \Delta e_1^*; h) = 0$ and we get

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

(A11) is equivalent to (7) and it thus remains to show that (A11) has a unique strictly positive solution. For this purpose, we now compute $\lim_{h\to 0}\frac{\partial \Delta e_1^*(\beta;h)}{\partial h}$, using that $\lim_{h\to 0}\frac{\partial \beta^*}{\partial h} = 0$ (as we show below). First, multiply (A3) with $C_2(e_2^l) - C_2(e_2^w)$, then differentiate both sides with respect to h and take the limit $h \to 0$ to obtain

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

where all efforts denote their optimal values in the limit and we used that in this limit $\Delta e_1^* = 0$ and $e_2^w = e_2^l$. Next, differentiate (A6) with respect to h and take the limit $h \to 0$:

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

Use $C_t''(e) = c_t$ and $\lim_{h\to 0} C_2'(e_2^w) = g(\beta_0^*)$ from (A4), to combine (A12) and (A13):

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

Substitution into (A11) leads to

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

Because L and the term in square brackets are strictly increasing and because $L(0) = 0$ and $\lim_{y\to z} L(y) = \infty$ by Assumption 1, there exists a unique $\beta_0^* > 0$ solving (A11).

Finally, to show that $\lim_{h\to 0}\frac{\partial \beta^*(h)}{\partial h} = 0$, suppose that 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

$$\lim_{h\to 0} \frac{\partial^2 \Delta e_1^*(h)}{\partial h^2} = 0. \tag{A16}$$

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,

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

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 $h \to 0$ and using (A16) yields $\lim_{h\to 0} S_{hh}(\beta, \Delta e_1^*(h); h) = 0$. Hence, $\lim_{h\to 0} S_{\beta hh}(\beta, \Delta e_1^*(h); h) = 0$ for any β and it follows that $\lim_{h\to 0} \frac{\partial \beta^*(h)}{\partial h} = 0$. ■

Proof of Corollary 2

Given equation (12), we need to show that convexity of $L(y)$ for $y \geq 0$ implies that $P_0^* = G(\beta_0^*) > \frac{3}{4}$. To simplify notation, let $k = |\Delta\epsilon_1|$, and note that k is distributed with density $2g(k)$ on the support $[0, z]$. 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,

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

where we have used that $g(z) = 0$ by Assumption 1(iv). The convexity of $L(y)$ for $y \geq 0$ implies that $L(\beta_0^*) \geq \mathbb{E}[L(k)] \geq L(\mathbb{E}[k])$, 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 first derive properties of the principal's optimal bias, given her belief (correct in equilibrium) about the agents' effort differential Δe_1 and the corresponding net advantage $\tilde{\alpha} = \alpha + \Delta e_1$. $\beta_A^*(\tilde{\alpha})$, $\beta_B^*(\tilde{\alpha})$, and $\beta^*(\tilde{\alpha})$ are strictly positive, because the left-hand sides of the first-order conditions (15) and (16) are strictly larger than one, while the right hand sides are equal to one when bias is zero, and strictly increasing in bias by the log-concavity of g. Moreover, (17) holds because for all $\tilde{\alpha} > 0$,

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

Because the first two ratios in (A18) are decreasing in $\tilde{\alpha}$ whereas the third term is increasing, $\beta_A^*(\tilde{\alpha})$ and $\beta^*(\tilde{\alpha})$ are strictly decreasing whereas $\beta_B^*(\tilde{\alpha})$ is strictly increasing. As the terms in (A18) become equal for $\tilde{\alpha} \to 0$, it holds that $\lim_{\alpha\to 0} \beta_A^*(\tilde{\alpha}) = \lim_{\alpha\to 0} \beta_B^*(\tilde{\alpha}) = \lim_{\alpha\to 0} \beta^*(\tilde{\alpha})$. Differentiating the left hand side 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 for $\alpha \to 0$, thus implying that $\lim_{\tilde{\alpha}\to 0} \frac{\partial \beta^*}{\partial \tilde{\alpha}} = 0$. Since the first-order conditions (15) are identical except for the sign of $\tilde{\alpha}$, it holds that $\lim_{\tilde{\alpha}\to 0} \frac{\partial \beta_A^*}{\partial \tilde{\alpha}} = -\lim_{\tilde{\alpha}\to 0} \frac{\partial \beta_B^*}{\partial \tilde{\alpha}}$. Finally, since the first ratio in (A18) is strictly decreasing in $\tilde{\alpha}$ even as $\tilde{\alpha} \to 0$, we have $\lim_{\tilde{\alpha}\to 0} \frac{\partial \beta_A^*}{\partial \tilde{\alpha}} < 0$.

Part (i) The proof of this claim treats jointly the cases of identity-dependent and identity-independent bias—for the latter case, simply impose $\beta_A = \beta_B = \beta$ throughout. In the second stage, agents exert identical efforts: The proof of this claim is analogous to the proof of Lemma 1 and is therefore omitted. Let e_2^w and e_2^l denote the agents' (identical) second-stage efforts after the advantaged agent A won or lost the first stage, respectively. From agent A's expected utility in stage one

$$-C_1(e_{A,1}) + \frac{1}{2} \sum_{\Delta a \in \{-h,h\}} \{G(\Delta a + \alpha + \Delta e_1)[G(\Delta a + \beta_A) - C_2(e_2^w)] \\ + G(-\Delta a - \alpha - \Delta e_1)[G(\Delta a - \beta_B) - C_2(e_2^l)]\},$$

and the analogue for agent B, we obtain the first-order conditions for $e_{A,1}$ and $e_{B,1}$:

$$2C_1'(e_{A,1}) = \sum_{\Delta a} g(\Delta a + \tilde{\alpha})[G(\Delta a + \beta_A) - G(\Delta a - \beta_B) + C_2(e_2^l) - C_2(e_2^w)], \tag{A19}$$

$$2C_1'(e_{B,1}) = \sum_{\Delta a} g(\Delta a + \tilde{\alpha})[G(-\Delta a + \beta_B) - G(-\Delta a - \beta_A) + C_2(e_2^w) - C_2(e_2^l)]. \tag{A20}$$

Comparing the marginal benefits of stage-one effort across agents, it follows from the symmetry of g about 0 that $G(-x) = 1 - G(x)$, and hence the marginal increase in the probability of selection is the same for A and B. Subtracting B's first-order condition from A's yields:

$$\frac{C_1'(e_{A,1}) - C_1'(e_{B,1})}{C_2(e_2^l) - C_2(e_2^w)} = \sum_{\Delta a \in \{-h,h\}} g(\Delta a + \alpha + \Delta e_1). \tag{A21}$$

Given that 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 the sign of $e_2^l - e_2^w$, consider the advantaged agent A's expected utility in the second stage, separately for the two cases where the advantaged agent won (w) or lost (l) the first stage, respectively:

$$q^w G(h+\beta_A) + (1-q^w)G(-h+\beta_A) - C_2(e_2^w),$$
$$q^l G(-h-\beta_B) + (1-q^l)G(h-\beta_B) - C_2(e_2^l).$$

Here we have introduced

$$q^w = \frac{G(h+\alpha+\Delta e_1)}{G(h+\alpha+\Delta e_1)+G(-h+\alpha+\Delta e_1)},$$
$$q^l = \frac{G(h-\alpha-\Delta e_1)}{G(h-\alpha-\Delta e_1)+G(-h-\alpha-\Delta e_1)}$$

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

$$C_2'(e_2^w) = q^w g(h+\beta_A) + (1-q^w)g(-h+\beta_A), \tag{A22}$$
$$C_2'(e_2^l) = q^l g(-h-\beta_B) + (1-q^l)g(h-\beta_B). \tag{A23}$$

Note that q^w (resp. q^l) is a decreasing (resp. increasing) function of the net advantage and that $q^l > q^w \Leftrightarrow \alpha + \Delta e_1 > 0$. Moreover, in equilibrium with ID biases, $\beta_A - \beta_B$ and $q^w - q^l$ have the same sign, since the principal chooses a larger bias the more confident she is that the first-stage winner is the more able agent.

We now argue, by contradiction, that $-\alpha < \Delta e_1 < 0$. Suppose first that, instead, $\Delta e_1 \le -\alpha$. Then $\alpha + \Delta e_1 \le 0$ implies that $q^l \le q^w$ and thus $\beta_A \ge \beta_B$. (For identity-independent bias, this condition holds trivially.) We have, for all $\beta \in (0, \beta_A]$,

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

where the first equality is the principal's first-order condition for β_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

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

with strict inequality if $\beta_B < \beta_A$. Since $\beta_B > 0$ and, under the hypothesis, $q^l \leq q^w$, the right-hand side 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 that $C_2'(e_2^w) \leq C_2'(e_2^l)$, and by the convexity of C_2 it follows that $e_2^w \leq e_2^l$. Since in equilibrium, Δe_1 must have the same sign as $e_2^l - e_2^w \geq 0$, we obtain 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 right-hand side 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, Δe_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.[25] 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, $\lim_{h\to 0} \beta_A^{ID} > 0$. We now show that

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

[25] Recall that we assumed, in footnote 8, that the slopes of the marginal cost of effort curves in each period are large enough to ensure the existence and uniqueness of equilibria.

For contradiction, assume that $\Delta e_1^{ID} \geq \Delta e_1^{II}$. Starting from $(\beta_A^{ID}, \beta_B^{ID}, \Delta e_1^{ID})$ suppose the principal's optimal response to Δe_1^{ID} is now restricted to be identity-independent, resulting in $\beta_A = \beta_B = \hat{\beta} \equiv \beta^*(\alpha + \Delta e_1^{ID})$. Consider the agents' corresponding effort response $\Delta e_1^*(\hat{\beta}; \hat{\beta})$. As costs are quadratic, it follows from (A21) that $\Delta e_1^*(\hat{\beta}; \hat{\beta})$ satisfies the implicit equation

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

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 above, 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 left-hand side of (A26) for any fixed Δe_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 left-hand side 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 Δe_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, let $\gamma = (\beta^*)^{-1} - \alpha$. Then $\gamma(\beta)$ gives the conjectured effort differential Δe_1 that makes β the principal's optimal II-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 $\beta \to 0$ and for $\beta \to \infty$, 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 $\gamma(\beta) > \Delta e_1^*(\beta, \beta)$. Note that $\hat{\beta} = \beta^*(\alpha + \Delta e_1^{ID}) \leq \beta^*(\alpha + \Delta e_1^{II}) = \beta^{II}$ because β^* 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 $\Delta e_1^{ID} \geq \Delta e_1^*(\hat{\beta}, \hat{\beta})$, contradicting our earlier result. ■

Proof of Corollary 3

This proof assumes that α 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 α values exists because i) (19) shows that $P_\alpha(\beta, \beta, \tilde{\alpha})$ is increasing in $\tilde{\alpha}$, ii) $\lim_{\alpha\to 0} \Delta e_1^{II} = 0$, so $\lim_{\alpha\to 0} \tilde{\alpha} = 0$, and iii) as shown in the proof of Proposition 2, $\lim_{\tilde{\alpha}\to 0} \frac{\partial \beta^*}{\partial \tilde{\alpha}} = 0$.

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_α is increasing in β_A but decreasing in β_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 α by showing that

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

For each bias regime (evaluating at the appropriate equilibrium values of the biases and effort differential), we can write

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

Here we used that costs are quadratic and that in the limit as $\alpha \to 0$, the agents exert the same first-stage effort. This common limiting value, e_1^*, is the same in both bias regimes and solves (A19) and (A20) for $\alpha \to 0$:

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

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

The proof of Proposition 2 showed that $\lim_{\alpha\to 0} \frac{\partial \beta_A^*}{\partial \alpha} = -\lim_{\alpha\to 0} \frac{\partial \beta_B^*}{\partial \alpha} < 0$ and that $\lim_{\alpha\to 0} \frac{\partial \beta^*}{\partial \alpha} = 0$. 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

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

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

Proof of Proposition 3

For early-career luck, $P_0^* = G(\beta_0^*)$ by Corollary 1. So P_0^* is strictly increasing in β_0^*, and since the right-hand side of (A15) is strictly decreasing in q and strictly increasing in β_0^*, we have that β_0^* and hence P_0^* are strictly increasing in q. For societal luck, 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 q and h, both β_A^{ID} and β_B^{ID} are strictly increasing in q. To simplify notation, we omit the superscript ID in β_A and β_B. Following the convention used in Section 4, suppose that agent A is socially advantaged. Since ability and societal advantage are independent, with probability $\frac{1}{2}$, A is the ability-favorite, and with probability $\frac{1}{2}$, A is the ability-underdog. Denote the corresponding first-stage effort differentials by Δe_1^f and Δe_1^u, respectively. Selective efficiency is

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

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, differentiate (A30) and (A31) with respect to h and let $h \to 0$. In doing so, we can ignore the effect of h on β_A and β_B since, by an argument similar to the one in the proof of Proposition 1, $\lim_{h\to 0}\frac{\partial \beta_A^{ID}}{\partial h} = \lim_{h\to 0}\frac{\partial \beta_B^{ID}}{\partial h} = 0$. Moreover, because for $h \to 0$ it becomes irrelevant whether the advantaged agent A is the favorite or the underdog, both Δe_1^f and Δe_1^u converge to the same value Δe_1^{ID}. Starting with (A30) we get

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

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

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

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

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

where we have used the assumption that $C_1(e) = \frac{1}{2}c_1e^2$. To compute the first term on the right-hand side of (A34), consider the analogues to (A4 and A5):

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

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

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, $\lim_{h\to 0} C_2'(e_2^{fw}) = g(\beta_A)$ and $\lim_{h\to 0} C_2'(e_2^{fl}) = g(\beta_B)$. Now differentiate (A35-A36) with respect to h and take the limit as $h \to 0$:

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

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

Finally, use $C_2(e) = \frac{1}{2}c_2e^2$ and substitute everything into (A34) to get

$$\frac{\partial \Delta e_1^f}{\partial h}\bigg|_{h\to 0} = (2q-1)\Psi(\beta_A, \beta_B), \tag{A39}$$

where we have defined

$$\Psi(\beta_A,\beta_B)=\frac{g'(\alpha+\Delta e_1^{ID})\left[g(\beta_B)^2-g(\beta_A)^2\right]-2g(\alpha+\Delta e_1^{ID})[g(\beta_B)g'(\beta_B)+g(\beta_A)g'(\beta_A)]}{c_1c_2-g'(\alpha+\Delta e_1^{ID})\left[g(\beta_B)^2-g(\beta_A)^2\right]}$$

To see that $\Psi(\beta_A,\beta_B)>0$, note first that the denominator of Ψ is positive because it has the same sign as the sum of the second-order conditions for the agents' first-stage effort choice problems. The numerator is also 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 $g'(\beta_A)<0$.

By following analogous steps, we can show that $\frac{\partial\Delta e_1^u}{\partial h}\big|_{h\to 0}=-\frac{\partial\Delta e_1^f}{\partial h}\big|_{h\to 0}$. We can then rewrite (A32), the limiting first-order condition for β_A, as

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

Performing the analogous steps for (A31) yields

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

Finally, differentiate (A40) and (A41) with respect to q, using the fact that in the limit as $h\to 0$, Δe_1^{ID} is independent of q. For q close to $\frac{1}{2}$, we can ignore terms proportional to $(2q-1)^2$, and we obtain

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

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

This proves that $\frac{\partial\beta_A^{ID}}{\partial q}>0$ and $\frac{\partial\beta_B^{ID}}{\partial q}>0$ for q and h sufficiently small.

Step 2. The next step is to extend our definition of the persistence of societal luck to the case where $q\geq\frac{1}{2}$. In the ID case, the persistence of societal luck, defined by (18) for the special case $q = \frac{1}{2}$, generalizes to

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

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

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

Dividing (A42) by (A43) we get $\frac{L'(\beta_A)}{L'(\beta_B)}\frac{\partial\beta_A/\partial q}{\partial\beta_B/\partial q} = \frac{G(-\alpha-\Delta e_1^{ID})}{G(\alpha+\Delta e_1^{ID})}$ and then

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

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

It remains to consider the II-case. 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 $h \to 0$, so under II bias, (A35) and (A36) become the same. Hence, $\lim_{h\to 0}\Delta e_1^f = 0$. Similarly, when A is the ability-underdog, $\lim_{h\to 0}\Delta e_1^u = 0$. Persistence (A44) is then

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

where $\beta_0^{II} \equiv \lim_{h\to 0}\beta^{II}$. Thus, $\lim_{h\to 0} P_\alpha^{II}$ is strictly increasing in β_0^{II}.

Adding (A32) to the analogous condition for $\lim_{h\to 0}\frac{\partial^2 S}{\partial\beta_B\partial h} = 0$, using (A39) and the fact that $\lim_{h\to 0}\frac{\partial\Delta e_1^u}{\partial h} = -\lim_{h\to 0}\frac{\partial\Delta e_1^f}{\partial h}$, and setting $\beta_A = \beta_B = \beta_0^{II}$ yields a fixed-point equation for β_0^{II}:

$$2g(\alpha) = L(\beta_0^{II})\left[1 - \frac{8(2q-1)^2}{c_1c_2}g(\alpha)^2 g(\beta_0^{II})^2\right]. \tag{A46}$$

Since the right-hand side of (A46) is decreasing in q and increasing in β_0^{II}, β_0^{II} is increasing in q. Thus $\lim_{h\to 0} P_\alpha^{II}$ also increases in q and, by continuity, the same is true for small h.

■

Proof of Proposition 4

To keep notation simpler this proof focuses on information structures as described in Section 5.2, but all arguments extend to cases where some (or all) realizations of $|\Delta x_1|$ can be observed precisely. Note that information structures are assumed symmetric with respect to the agents' "names".

Consider first the persistence of early-career luck. Here $\alpha = 0$ (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 $h \to 0$, $|\Delta x_1|$ falls in $[d_{n-1}, d_n)$ with probability $2[G(d_n) - G(d_{n-1})]$, we can use the first order conditions (22) determining the limiting biases β_0^n to obtain

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

where $d_0 = 0$ and $d_{N+1} = \infty$. Hence, regardless of the details of the information structure, the ex ante expectation of $L(\beta)$ 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 β_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 β_0^{lm} and β_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 (Assumption 1) 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 β_0 are strictly positive imply that $G(L^{-1}(.))$ is strictly concave on the support 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, and assume that bias can condition on agents' identities. 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 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 $h \to 0$, of A's net advantage by $\tilde{\alpha}$. The same arguments used to prove part (i) of Proposition 2 show that $\tilde{\alpha} > 0$. With $q = \frac{1}{2}$, the limiting biases, as $h \to 0$, 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)). Define $H(y) \equiv G(L^{-1}(y)) - \frac{1}{2}$. Then (A44) can be expressed as

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

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

$\tilde{\alpha} > 0$. Hence $P_{\alpha}^{ID}(I_{ord}) > \frac{1}{2} = P_{\alpha}^{ID}(I_{card})$ for h small. ■

Proof of Proposition 5

For part (i), we have $P_0^* = G(\beta_0^*)$ (from Corollary 1), so P_0^* is strictly increasing in β_0^*. As the right-hand side of (A15) is strictly increasing in c_t for $q > \frac{1}{2}$, β_0^* is strictly decreasing in c_t. Hence P_0^* is strictly decreasing in c_t.

For part (ii), recall that for $h \to 0$, P_{α}^{ID} is given by (A44), with the biases β_A and β_B solving (A40) and (A41). For $q = \frac{1}{2}$ and $h \to 0$, the first-stage effort differential Δe_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

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

We now show that the solution satisfies $\frac{\partial \Delta e_1^{ID}}{\partial c_t} > 0$ for $t = 1, 2$. For the remainder of the proof, we omit the superscript ID in Δe_1^{ID}. Setting $q = \frac{1}{2}$ and differentiating (A40), (A41), and (A49) with respect to c_t, we obtain

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

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

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

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

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

On the left-hand side, 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 $L'(\beta_A) > 0$. As $\left[\frac{g(-\alpha-\Delta e_1)}{G(-\alpha-\Delta e_1)}\right]'$ 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 large parentheses) is positive follows by adding the agents' second-order conditions for stage-one efforts. The right-hand side is positive since $\beta_B > \beta_A > 0$ and g is decreasing in the positive domain. Hence, we can conclude that $\frac{\partial \Delta e_1^{ID}}{\partial c_t} > 0$ 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 $\tilde{\alpha}$:

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

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

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

where the inequality holds since $H(0) = 0$ and $H(y)$ is strictly concave for $y > 0$, given the convexity of L on the positive domain. For small α, Δe_1^{ID} and $\tilde{\alpha}$ are also small, since for $\alpha = 0$, $\Delta e_1^{ID} = 0$ (Lemma 1). Hence for small α and h, P_α^{ID} is strictly increasing in c_t for $t = 1, 2$. ■

References

Acemoglu, Daron and David Autor, "Skills, Tasks and Technologies: Implications for Employment and Earnings.," in David Card and Orley Ashenfelter, eds., *Handbook of Labor Economics*, Vol. 4, Elsevier, 2011, pp. 1043–1171.

_ , _ , Jonathon Hazell, and Pascual Restrepo, "Artificial Intelligence and Jobs: Evidence from Online Vacancies," *Journal of Labor Economics*, 2022, *40* (S1), S293–S340.

Alesina, Alberto and George Marios Angeletos, "Fairness and redistribution," *American Economic Review*, 2005, *95* (4), 960–980.

_ , Stefanie Stantcheva, and Edoardo Teso, "Intergenerational Mobility and Preferences for Redistribution," *American Economic Review*, 2018, *108* (2), 521–554.

Almås, Ingvild, Alexander W. Cappelen, and Bertil Tungodden, "Cutthroat capitalism versus cuddly socialism: Are americans more meritocratic and efficiency-seeking than Scandinavians?," *Journal of Political Economy*, 2020, *128* (5), 1753–1788.

Altonji, Joseph G. and Charles R. Pierret, "Employer learning and statistical discrimination," *Quarterly Journal of Economics*, 2001, *116* (1), 313–350.

Ariga, Kenn, Yasushi Ohkusa, and Giorgio Brunello, "Fast track: Is it in the genes? The promotion policy of a large Japanese firm," *Journal of Economic Behavior and Organization*, 1999, *38* (4), 385–402.

Asch, Solomon E., "Forming impressions of personality," *Journal of Abnormal and Social Psychology*, 1946, *41* (3), 258–290.

Autor, David, David Dorn, Lawrence F Katz, Christina Patterson, and John Van Reenen, "The Fall of the Labor Share and the Rise of Superstar Firms," *The Quarterly Journal of Economics*, 02 2020, *135* (2), 645–709.

Bagues, Manuel F. and Berta Esteve-Volart, "Can Gender Parity Break the Glass Ceiling? Evidence from a Repeated Randomized Experiment," *The Review of Economic Studies*, 2010, *77* (4), 1301–1328.

_ and Maria J. Perez-Villadoniga, "Do recruiters prefer applicants with similar skills? Evidence from a randomized natural experiment," *Journal of Economic Behavior & Organization*, 2012, *82* (1), 12–20.

Baker, George, Michael Gibbs, and Bengt Holmström, "The Internal Economics of the Firm: Evidence from Personnel Data," *Quarterly Journal of Economics*, 1994, *109* (4), 881–919.

Bao, Renjie, Jan De Loecker, and Jan Eeckhout, "Manager Pay Inequality and Market Power," *Review of Economic Studies*, 2026, *forthcoming.*

Bardhi, Arjada, Yingni Guo, and Bruno Strulovici, "Early-Career Discrimination: Spiraling or Self-Correcting?," 2025. https://www.arjadabardhi.com/s/ReStud_ECD_draft45.pdf.

Bartoš, Vojtěch, Michal Bauer, Julie Chytilová, and Filip Matějka, "Attention discrimination: Theory and field experiments with monitoring information acquisition," *American Economic Review*, 2016, *106* (6), 1437–1475.

Bedard, Kelly and Elizabeth Dhuey, "The Persistence of Early Childhood Maturity: International Evidence of Long-Run Age Effects," *The Quarterly Journal of Economics*, 2006, *121* (4), 1437–1472.

Belzil, Christian and Michael Bognanno, "Promotions, demotions, halo effects, and the earnings dynamics of American executives," *Journal of Labor Economics*, 2008, *26* (2), 287–310.

Benabou, Roland and Jean Tirole, "Belief in a Just World and Redistributive Politics," *Quarterly Journal of Economics*, 2006, *121* (2), 699–746.

Benson, Alan, Danielle Li, and Kelly Shue, ""Potential" and the Gender Promotion Gap," *American Economic Review*, February 2026, *116* (2), 375–417.

Blank, Rebecca M., "Tracing the Economic Impact of Cumulative Discrimination," *American Economic Review*, 2005, *95* (2), 99–103.

Bohren, J. Aislinn, Alex Imas, and Michael Rosenberg, "The dynamics of discrimination: Theory and evidence," *American Economic Review*, 2019, *109* (10), 3395–3436.

_ **, Peter Hull, and Alex Imas**, "Systemic Discrimination: Theory and Measurement," *Quarterly Journal of Economics*, 2025, *140* (3), 1743–1799.

Cappelen, Alexander W., Astri Drange Hole, Erik Ø. Sørensen, and Bertil Tungodden, "The Pluralism of Fairness Ideals: An Experimental Approach," *American Economic Review*, 2007, *97* (3), 818–827.

_ **, James Konow, Erik Ø. Sørensen, and Bertil Tungodden**, "Just luck: An experimental study of risk-taking and fairness," *American Economic Review*, 2013, *103* (4), 1398–1413.

Card, David, Jörg Heining, and Patrick Kline, "Workplace Heterogeneity and the Rise of West German Wage Inequality," *Quarterly Journal of Economics*, 05 2013, *128* (3), 967–1015.

Cattan, Sarah, Kjell G. Salvanes, and Emma Tominey, "First-Generation Elite: The Role of School Social Networks," *American Economic Review*, December 2025, *115* (12), 4369–4403.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez, "Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States," *Quarterly Journal of Economics*, 09 2014, *129* (4), 1553–1623.

Ciocca-Eller, Christina, "What Makes a Quality College? Reexamining the Equalizing Potential of Higher Education in the United States," *American Journal of Sociology*, 2023, *129* (3), 637–714.

Cole, Stephen, Jonathan R. Cole, and Gary A. Simon, "Chance and Consensus in Peer Review," *Science*, 1981, *214* (4523), 881–886.

Cong, Lin William and Yizhou Xiao, "Persistent Blessings of Luck: Theory and an Application to Venture Capital," *Review of Financial Studies*, 2022, *35* (3), 1183–1221.

Corak, Miles, "Income Inequality, Equality of Opportunity, and Intergenerational Mobility," *Journal of Economic Perspectives*, September 2013, *27* (3), 79–102.

Currie, Janet and Douglas Almond, "Human capital development before age five," in David Card and Orley Ashenfelter, eds., *Handbook of Labor Economics*, Vol. 4, Part B, ch. 15, Elsevier, 2011, pp. 1315–1486.

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

Deming, David J., "The Growing Importance of Social Skills in the Labor Market," *Quarterly Journal of Economics*, 2017, *132* (4), 1593–1640.

_ , "The Growing Importance of Decision-Making on the Job," *NBER Working Paper 28733*, April 2021.

DiPrete, Thomas A. and Gregory M. Eirich, "Cumulative advantage as a mechanism for inequality: A review of theoretical and empirical developments," *Annual Review of Sociology*, 2006, *32*, 271–297.

Du, Qianqian, Huasheng Gao, and Maurice D. Levi, "The relative-age effect and career success: Evidence from corporate CEOs," *Economics Letters*, 2012, *117* (3), 660–662.

Durandard, Théo, "Dynamic delegation in promotion contests," 2023. https://arxiv.org/abs/2308.05668.

Exley, Christine and Kirby Nielsen, "The Gender Gap in Confidence: Expected But Not Accounted For," *American Economic Review*, 2024, *114* (3), 851–885.

Fabbri, Allessandro and Michele Pellizzari, "First-Generation Graduates in OECD Countries," *CEPR Discussion Paper 20796*, 2025.

Farber, Henry and Robert Gibbons, "Learning and Wage Dynamics," *Quarterly Journal of Economics*, 1996, *111* (4), 1007–1047.

Fong, Christina, "Social preferences, self-interest, and the demand for redistribution," *Journal of Public Economics*, 2001, *82*, 225–246.

Forbes, J. Benjamin, "Early Intraorganizational Mobility: Patterns and Influences," *Academy of Management Journal*, 1987, *30* (1), 110–125.

Frank, Robert H., *Success and Luck: Good Fortune and the Myth of Meritocracy*, Princeton University Press, 2016.

Fryer, Roland G., "Belief flipping in a dynamic model of statistical discrimination," *Journal of Public Economics*, 2007, *91* (5-6), 1151–1166.

Gabaix, Xavier and Augustin Landier, "Why has CEO Pay Increased So Much?," *Quarterly Journal of Economics*, 02 2008, *123* (1), 49–100.

Gamoran, Adam and Robert Mare, "Secondary school tracking and educational inequality: compensation, reinforcement or neutrality?," *American Journal of Sociology*, 1989, *94* (5), 1146–1183.

Gibbons, Robert and Michael Waldman, "A Theory of Wage and Promotion Dynamics inside Firms," *Quarterly Journal of Economics*, 1999, *114* (4), 1321–1358.

_ **and** _ , "Enriching a theory of wage and promotion dynamics inside firms," *Journal of Labor Economics*, 2006, *24* (1), 59–107.

Gill, David and Victoria Prowse, "Gender differences and dynamics in competition: The role of luck," *Quantitative Economics*, 2014, *5* (2), 351–376.

Hamilton, Barton H., Jack A. Nickerson, and Hideo Owan, "Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation," *Journal of Political Economy*, 2003, *111* (3), 465–497.

Hansen, Stephen, Tejas Ramdas, Raffaella Sadun, and Joe Fuller, "The Demand for Executive Skills," 2021. NBER Working Paper, http://www.nber.org/papers/w28959.

Heckman, James J., "The economics, technology, and neuroscience of human capability formation," *Proceedings of the National Academy of Sciences*, 2007, *104* (33), 13250–13255.

Konow, James, "Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions," *American Economic Review*, 2000, *90* (4), 1072–1092.

Lange, Fabian, "The speed of employer learning," *Journal of Labor Economics*, 2007, *25* (1), 1–35.

Lavy, Victor and Rigissa Megalokonomou, "The Short- and the Long-Run Impact of Gender-Biased Teachers," *American Economic Journal: Applied Economics*, 2024, *16* (2), 176–218.

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

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

Lefgren, Lars J., David P. Sims, and Olga B. Stoddard, "Effort, luck, and voting for redistribution," *Journal of Public Economics*, 2016, *143*, 89–97.

Lepage, Louis-Pierre, "Experience-based Discrimination," *American Economic Journal: Applied Economics*, 2024.

Li, Danielle, Lindsey Raymond, and Peter Bergman, "Hiring as Exploration," *Review of Economic Studies*, 2025.

Lluis, Stéphanie, "The Role of Comparative Advantage and Learning in Wage Dynamics and Intrafirm Mobility: Evidence from Germany," *Journal of Labor Economics*, 2005, *23* (4), 725–767.

Mas, Alexandre and Enrico Moretti, "Peers at Work," *American Economic Review*, March 2009, *99* (1), 112–45.

Merton, Robert K., "The Matthew effect in science," *Science*, 1968, *159* (3810), 56–62.

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

Minni, Virginia, "Making the Invisible Hand Visible: Managers and the Allocation of Workers to Jobs," 2025. https://virginiaminni.github.io/assets/files/Minni_JMP.pdf.

Murdock, Bennet B., "The serial position effect of free recall," *Journal of Experimental Psychology*, 1962, *64* (5), 482–488.

Nalebuff, Barry J. and Joseph E. Stiglitz, "Prizes and incentives: towards a general theory of compensation and competition," *Bell Journal of Economics*, 1983, pp. 21–43.

Nanda, Ramana, Sampsa Samila, and Olav Sorenson, "The persistent effect of initial success: Evidence from venture capital," *Journal of Financial Economics*, 2020, *137* (1), 231–248.

Onuchic, Paula, "Recent Contributions to Theories of Discrimination," 2025. https://paulaonuchic.com/s/ThDisc2025.pdf.

Oreopoulos, Philip, Till von Wachter, and Andrew Heisz, "The short- and long-term career effects of graduating in a recession," *American Economic Journal: Applied Economics*, 2012, *4* (1), 1–29.

Pastorino, Elena, "Careers in Firms: The Role of Learning about Ability and Human Capital Acquisition," *Journal of Political Economy*, 2024, *132* (6), 1994–2073.

Piketty, Thomas, *Capital in the Twenty-First Century*, Cambridge, MA: The Belknap Press of Harvard University Press, 2014.

Rabin, Matthew and Joel L. Schrag, "First impressions matter: A model of confirmatory bias," *Quarterly Journal of Economics*, 1999, *114* (1), 37–82.

Roemer, John E., "Equality of Opportunity," in Kenneth Arrow, Samuel Bowles, and Steven Durlauf, eds., *Meritocracy and Economic Inequality*, Princeton (NJ): Princeton University Press, 2000, pp. 17–32.

Rosen, Sherwin, "The Economics of Superstars," *American Economic Review*, 1981, *71* (5), 845–858.

Rosenbaum, James E., "Tournament Mobility: Career Patterns in a Corporation," *Administrative Science Quarterly*, 1979, *24* (2), 220–241.

Sandel, Michael, *The Tyranny of Merit*, New York: Farrar, Straus and Giroux, 2020.

Sethi, Rajiv and Rohini Somanathan, "Meritocracy and Representation," *Journal of Economic Literature*, 2023, *61*, 941–957.

Song, Jae, David J Price, Fatih Guvenen, Nicholas Bloom, and Till von Wachter, "Firming Up Inequality," *Quarterly Journal of Economics*, 2019, *134* (1), 1–50.

Zhang, Sam, K. Hunter Wapman, Daniel B. Larremore, and Aaron Clauset, "Labor advantages drive the greater productivity of faculty at elite universities," *Science Advances*, 2022, *8* (46), 1–22.