

# On the Future of Education in the Age of AI

Nianzu Xiong

Department of Economics, University of Oxford

October 2026

## Abstract

For outsiders, the main route into elite universities, research labs and professional firms has been technical skill: work that anyone can check, whatever the candidate's background. Generative AI now does much of this work. We model how institutions choose people once the skill that matters most is judgment, which is harder to observe. Letters from referees at other institutions carry no weight, because those referees want their students placed regardless of quality. Sponsors inside the institution share in its reputation, so they are believed, but they favor their own candidates. As AI makes checkable skills less valuable, four things follow: outsiders are under-selected even by fair evaluators, favoritism matters more, the apprenticeships that once trained and revealed talented outsiders disappear, and winning a patron gets easier. AI can push the other way if it is used to verify ideas. A simple pass-fail test that the outsider's home institution commits to in advance helps outsiders more than full disclosure.

**JEL:** D82, D83, I23, J24, J71, O33. **Keywords:** artificial intelligence, Blackwell informativeness, cheap talk, Bayesian persuasion, influence activities, apprenticeship, collective reputation.

# 1 Introduction

Elite intellectual institutions, such as leading research departments, top law firms, strategy consultancies and research laboratories, perform two functions. They produce ideas, and they certify those who produce them. The second function rests on a technology of selection, and the openness of such institutions to outsiders depends on what that technology can observe and on whose reports it can believe.

Historically, mobility into these institutions has depended mainly on *verifiable* human capital. Technical competence, meaning the capacity to execute an empirical design, solve a model or implement an estimator correctly, has three properties that make it an unusually democratic signal. It can be demonstrated within a short relationship; its assessment consumes little evaluative attention; and it leaves little room for discretion, since its quality is, to a first approximation, objectively checkable. The apprenticeship structure of elite training can be understood as an institution organized around these properties. It converts a candidate's execution capital into a credible signal irrespective of provenance, and, once the candidate is inside, it exposes her to the tacit knowledge and judgment of the institution.

Generative AI is, to a first approximation, a technology for automating execution. As its capability rises, the return to execution capital falls, and the margin on which institutions select shifts toward *conceptual capital*: the capacity to identify questions worth asking, to frame them productively, and to recognize which among many feasible projects matter. Conceptual capital differs from execution capital in the dimensions that governed mobility. It is assessed subjectively and through sustained interaction, and the people who hold the relevant information (advisers, letter writers, colleagues) have interests of their own.

This paper develops a theory of how this shift reshapes selection into elite institutions. Our starting point is that the information an institution holds about a candidate is determined in equilibrium. Information about conceptual capital is soft: it is held privately by referees and transmitted through cheap talk. Whether it is transmitted at all depends

on the referee's incentives. A referee at an outside institution cares about placing her student and has no stake in the elite's reputation; her preferences over the elite's decision are state-independent, and her soft information is, in every equilibrium, irrelevant to the outcome. A sponsor inside the institution shares in its reputation; her preferences are state-dependent, and she transmits information credibly, albeit with a bias. The insiders' informational advantage therefore comes from incentive alignment as well as proximity, and it takes the form of a Blackwell ranking: the outsider's experiment is a garbling of the insider's.

We show that this ranking matters for selection once the verifiable component of value declines. It works through four mechanisms.

1. **Informational exclusion.** A Blackwell-superior experiment generates a posterior-mean distribution that is a mean-preserving spread of the inferior one. At selective thresholds, the better-observed group is over-represented, even when ability distributions coincide and evaluators are unbiased. Verifiable execution capital, observed equally for all, dilutes this asymmetry; AI removes the dilution (Theorem 1, Proposition 2).
2. **Amplified favoritism.** Sponsors' bias operates on the soft component. Measured relative to the residual dispersion of the selection index, a given premium has a larger effect after AI. Under full automation, a premium worth one-tenth of a standard deviation of the pre-AI index is worth a third of one (Proposition 3).
3. **Loss of apprenticeship.** Apprenticeship once performed a dual function: it *formed* outsiders' conceptual capital, and it *revealed* it by converting outsiders into informational insiders. AI reduces demand for apprentices. The resulting group difference in conceptual capital is real, endogenous and self-confirming in the sense of Coate and Loury (1993) (Propositions 4–5).
4. **Cheaper patronage.** Candidates with proximity to sponsors can purchase favor through influence activities (Milgrom, 1988). Because AI compresses the dispersion of the se-

lection index, the efficiency of influence required to bridge a given rank gap falls. The composition of entrants shifts: before AI, an elite admits conceptual talent, execution talent, and a small number of exceptionally skilled cultivators of patrons; after AI, it admits conceptual talent and a broader class of moderately skilled cultivators (Propositions 6–7).

AI can also work in the other direction, as a *verifier* of ideas, which is what outsiders lack. When an AI-based evaluation is publicly observed, it adds a common experiment to both groups, and the insiders’ advantage in value terms vanishes as its precision grows (Proposition 8). In addition, if an outsider’s home institution can design a verifiable test of ideas with commitment, the logic of Bayesian persuasion (Kamenica and Gentzkow, 2011) implies that the optimal test is a coarse pass–fail standard that places outsiders strictly more often than full disclosure, whereas soft information of any quantity places none of them (Proposition 9). Verifiability therefore matters more than precision. Whether AI opens or closes elite institutions depends on whether it is used to replace execution or to verify ideas, and on whether the growing volume of AI-assisted output congests the public channels on which outsiders depend.

Finally, we embed selection in a model of collective reputation (Tirole, 1996). Each member bears only a fraction of the reputational cost of a weak appointment, so the equilibrium premium exceeds the institutional optimum and increases with institutional size. Reputation responds with a lag, granting incumbents a reputational overhang, and the formation channel partially masks decline: insiders are in fact better on average, even as the institution is selecting less well than it could. In the long run, the decline is corrected by the entry of institutions that commit to verifiable evaluation.

**Related literature.** The static logic of differential precision builds on the statistical-discrimination literature (Phelps, 1972; Aigner and Cain, 1977), and in particular on Cornell and Welch (1996), who show that screening with group-specific information generates

under-selection at selective thresholds. Self-confirming group differences arising from investment incentives originate with [Coate and Loury \(1993\)](#). Our contribution relative to these works is to endogenize the precision ranking through communication incentives, to cast it in terms of Blackwell experiments ([Blackwell, 1953](#)), and to show how a technological shock to the composition of valued skills activates, amplifies and entrenches it.

The communication analysis draws on cheap talk ([Crawford and Sobel, 1982](#)) and on delegation versus communication within organizations ([Dessein, 2002](#)). The analysis of verifiable tests draws on Bayesian persuasion ([Kamenica and Gentzkow, 2011](#)), its applications to grading and certification ([Ostrovsky and Schwarz, 2010](#); [Boleslavsky and Cotton, 2015](#)), competition among senders ([Gentzkow and Kamenica, 2017](#)), and the comparison of distributions via rotations ([Johnson and Myatt, 2006](#)). Our treatment of manipulation of AI evaluation relates to [Frankel and Kartik \(2019\)](#). Influence activities follow [Milgrom \(1988\)](#) and [Milgrom and Roberts \(1988\)](#); signal jamming follows [Holmström \(1999\)](#); signaling follows [Spence \(1973\)](#).

On the economics of AI, we adopt the task-based perspective of [Acemoglu and Restrepo \(2018\)](#) and the distinction between prediction and judgment in [Agrawal et al. \(2019\)](#); [Korinek \(2023\)](#) discusses AI in economic research specifically. On gatekeeping in science, the theory is consistent with evidence on favoritism and information in editorial decisions ([Laband and Piette, 1994](#); [Brogaard et al., 2014](#)), social ties in economics ([Colussi, 2018](#)), expertise and bias in peer review ([Li, 2017](#)), the concentration of publication and promotion ([Heckman and Moktan, 2020](#)), and the entrenchment of scientific elites ([Azoulay et al., 2019](#)).

**Outline.** Section 2 presents the environment. Section 3 studies communication and derives the Blackwell ranking. Section 4 establishes informational exclusion and the favoritism multiplier. Sections 5 and 6 develop the apprenticeship and influence channels. Section 7 studies AI as a verifier, including the persuasion problem. Section 8 analyzes

reputation. Section 9 derives empirical implications, and Section 10 concludes. Proofs are collected in the Appendix.

## 2 Environment

### 2.1 Candidates and technology

An elite institution fills a mass  $K$  of positions from a population of candidates. Each candidate belongs to a group  $g \in \{I, O\}$ : *insiders*, of mass  $m_I$ , are affiliated with a member of the institution; *outsiders*, of mass  $m_O$ , are not. A candidate is characterized by a state  $\omega = (\theta, a)$ , where  $\theta$  denotes conceptual capital and  $a$  execution capital. To isolate informational mechanisms, we initially take ability distributions to be identical across groups; Section 5 relaxes this. Unless stated otherwise,

$$\theta \sim N(0, \tau^2), \quad a \sim N(0, \sigma_a^2), \quad \theta \perp a. \quad (1)$$

Theorem 1 and Proposition 9 hold for general distributions.

A candidate's value to the institution is

$$v = \theta + \alpha(A) a, \quad \alpha'(A) < 0, \quad (2)$$

where  $A \in [0, 1]$  indexes AI capability. AI substitutes for execution, so the marginal value of execution capital declines with  $A$ . We use  $\alpha(A) = \alpha_0(1 - A)$  in numerical illustrations; all results require only monotonicity.

### 2.2 Evidence: hard and soft

The institution learns about candidates through *evidence*. We distinguish two kinds.

*Hard evidence* is publicly observed and cannot be manipulated by its provider. Execution

capital generates hard evidence: work products reveal  $a$  perfectly and at negligible cost, for both groups. Conceptual capital generates hard evidence only through public output,

$$y = \theta + \eta, \quad \eta \sim N(0, \rho_y^{-1}), \quad (3)$$

where  $\rho_y$  is the precision with which the institution's evaluators can assess a candidate's public work. Section 7 adds an AI-based evaluation  $T$  as a further source of hard evidence.

*Soft evidence* is observed privately by a referee and transmitted through unverifiable messages. Each candidate has one referee who observes

$$s = \theta + \varepsilon, \quad \varepsilon \sim N(0, \rho_s^{-1}), \quad (4)$$

independently of  $y$ . The referee of an insider is her *sponsor*, a member of the institution. The referee of an outsider is a faculty member at an outside institution. We give both referees *identical* information: the asymmetry we derive arises from incentives, not from access.<sup>1</sup>

## 2.3 The institution

The institution comprises  $N$  members. It appoints the  $K$  candidates with the highest expected value given all credible information, subject to the procedural norm described in Section 3. Its members share a common reputation, which is increasing in the average quality of appointees; Section 8 models reputation explicitly. For now, it suffices that a member's payoff from an appointment of expected value  $w$ , relative to the marginal appointee at cutoff  $c$ , is

$$u_j = \pi \mathbf{1}\{\text{own candidate}\} + \frac{\gamma}{N} (w - c), \quad (5)$$

---

<sup>1</sup>Allowing insiders' referees to be better informed, as is plausible, reinforces every result.

where  $\pi > 0$  is the private benefit of placing one's own candidate (collaboration, the continuation of a research program, intellectual legacy),  $\gamma$  is the value of reputation, and  $1/N$  is the share of reputational gains and losses borne by each member.

An outside referee obtains a payoff  $\pi_O > 0$  if her candidate is appointed and zero otherwise. She has no stake in the elite's reputation.

### 3 Verifiability and communication

#### 3.1 Outside referees

The communication game between an outside referee and the institution has a simple structure. The referee observes  $s$ , sends a message  $m \in M$ , and the institution decides whether to appoint. Because the referee's payoff depends only on the decision, not on the state, she prefers whichever message maximizes the probability of appointment, whatever she has observed.

**Lemma 1** (Irrelevance of outside soft information). *In every perfect Bayesian equilibrium of the communication game between an outside referee and the institution, the probability that an outsider is appointed, conditional on her hard evidence, is independent of the referee's soft signal  $s$ . The outcome is therefore identical to that of an equilibrium in which the institution observes only hard evidence.*

Lemma 1 formalizes a familiar intuition, that letters from outside are discounted. Here the discount is complete and arises from incentives alone. An outsider is evaluated exclusively on hard evidence.

#### 3.2 Sponsors

A sponsor's preferences are state-dependent: by (5), she gains from appointing her candidate if and only if  $\pi + (\gamma/N)(w - c) \geq 0$ , where  $w$  is her posterior expectation of the

candidate's value. Her interests are partially aligned with the institution's, and partial alignment supports informative communication.

We assume that the institution operates under a norm of *reciprocal deference*: each member's recommendation regarding her own candidate is followed. Such norms are sustained in repeated interaction among members who would each lose more from the breakdown of deference than they would gain from overriding a colleague (cf. [Dessein, 2002](#), on delegation within organizations). Under reciprocal deference, the sponsor effectively decides.

**Proposition 1** (Sponsorship). *Under reciprocal deference, a sponsor recommends her candidate if and only if*

$$W \geq c - b^*, \quad b^* = \frac{N\pi}{\gamma}, \quad (6)$$

where  $W = \mathbb{E}[v \mid a, y, s]$ . Consequently, the institution's information about insiders is the experiment  $\mathcal{E}_I = (a, y, s)$ , while its information about outsiders is  $\mathcal{E}_O = (a, y)$ . In particular,  $\mathcal{E}_O$  is a garbling of  $\mathcal{E}_I$ , and the posterior precisions about  $\theta$  are

$$\rho_I = \rho_y + \rho_s, \quad \rho_O = \rho_y. \quad (7)$$

Two points follow. First, the Blackwell ranking  $\mathcal{E}_I \succeq \mathcal{E}_O$  is derived rather than assumed: the experiments are generated by the same physical evidence, but only one referee can credibly transmit what she knows. Second, the sponsor's information arrives bundled with a bias  $b^*$ . Information and favoritism are joint products of sponsorship.

*Remark 1* (Communication without deference). Absent deference, a sponsor's recommendation is followed only if the institution's posterior conditional on a recommendation exceeds the cutoff. With a normal posterior increment of standard deviation  $\Delta$ , this credibility constraint requires the standardized bias  $b^*/\Delta$  to lie below a bound that is decreasing in the institution's selectivity. Highly selective institutions can thus extract information from sponsors only if they defer, which provides a rationale for the norm. See [Dessein \(2002\)](#) for

the general trade-off between delegation and communication.

Let  $\beta_g \equiv \tau^2 \rho_g / (1 + \tau^2 \rho_g)$  denote the weight on the group's evidence about  $\theta$ , so that  $\beta_I > \beta_O$  whenever  $\rho_s > 0$ . The posterior mean of value is

$$W_g = \alpha a + \mathbb{E}[\theta \mid \mathcal{E}_g] \sim N(0, V_g), \quad V_g(\alpha) = \alpha^2 \sigma_a^2 + \beta_g \tau^2. \quad (8)$$

We refer to  $V_g$  as the group's *visibility*. The institution appoints the  $K$  candidates with the highest score  $W + b^* \mathbf{1}\{g = I\}$ , so that the cutoff  $c$  satisfies

$$m_I \left[ 1 - \Phi \left( \frac{c - b^*}{\sqrt{V_I}} \right) \right] + m_O \left[ 1 - \Phi \left( \frac{c}{\sqrt{V_O}} \right) \right] = K. \quad (9)$$

**Assumption 1** (Selectivity).  $K < (m_I + m_O)/2$ .

Let  $P_g$  denote group  $g$ 's appointment probability and  $S_O \equiv m_O P_O / K$  the outsider share of appointments.

## 4 Informational exclusion and the favoritism multiplier

### 4.1 A Blackwell foundation

We first show that the insiders' informational advantage generates over-representation at selective thresholds under weak conditions. Let  $F_g$  denote the distribution of the posterior mean  $W_g = \mathbb{E}[v \mid \mathcal{E}_g]$  for an arbitrary prior over  $\omega$  and arbitrary experiments.

**Definition 1** (Rotation).  $F_I$  is a *rotation* of  $F_O$  at  $t^*$  if  $F_I(t) \geq F_O(t)$  for  $t < t^*$  and  $F_I(t) \leq F_O(t)$  for  $t > t^*$ .

**Theorem 1** (Blackwell exclusion). Suppose  $\mathcal{E}_O$  is a garbling of  $\mathcal{E}_I$  with respect to  $\omega$ . Then for every  $\alpha \geq 0$ :

- (i)  $W_I$  dominates  $W_O$  in the convex order,  $W_O = \mathbb{E}[W_I \mid \mathcal{E}_O]$ ;

- (ii) for every cutoff  $c$ , the institution's expected option value from screening an insider exceeds that from screening an outsider:  $\mathbb{E}[(W_I - c)^+] \geq \mathbb{E}[(W_O - c)^+]$ ;
- (iii) if  $F_I$  is a rotation of  $F_O$  at  $t^*$ , then  $P_I \geq P_O$  for every common cutoff  $c \geq t^*$ ; under Gaussian evidence  $t^* = 0$ , and the inequality is strict whenever  $\rho_s > 0$  and  $c > 0$ ;
- (iv) if  $a$  has a continuous distribution and  $\mathbb{E}[\theta^2] < \infty$ , then as  $\alpha \rightarrow \infty$  the difference in appointment probabilities at the market-clearing cutoff vanishes.

Parts (i)–(iii) say that any Blackwell advantage, however it arises, manifests as an advantage in the upper tail. Part (iv) says that a large verifiable component, observed equally for all, renders the advantage immaterial. AI moves the economy from the regime of part (iv) toward that of part (iii).

The theorem has a decision-theoretic reading. The institution's problem is an option: it appoints only candidates whose posterior exceeds  $c$ , so its payoff is convex in the posterior mean. Blackwell's theorem states that more informative experiments are more valuable for every convex payoff; the theorem translates this into a statement about who is appointed. In this sense, a group that is better understood is a more valuable option for a selective institution.

## 4.2 Comparative statics under Gaussian evidence

Two statistics summarize the allocation under Gaussian evidence: the visibility ratio and the standardized premium,

$$k(\alpha) \equiv \frac{V_I(\alpha)}{V_O(\alpha)} = \frac{\alpha^2 \sigma_a^2 + \beta_I \tau^2}{\alpha^2 \sigma_a^2 + \beta_O \tau^2} \geq 1, \quad \tilde{b}(\alpha) \equiv \frac{b^*}{\sqrt{V_O(\alpha)}}. \quad (10)$$

Dividing all scores by  $\sqrt{V_O}$  preserves rankings: outsiders' standardized scores are  $N(0, 1)$ , and insiders' are  $N(0, k)$  shifted by  $\tilde{b}$ . Thus  $(k, \tilde{b})$  is sufficient for the allocation.

**Lemma 2.**  *$k$  is strictly decreasing in  $\alpha$ , with  $k(0) = \beta_I/\beta_O$  and  $k \rightarrow 1$  as  $\alpha \rightarrow \infty$ . Hence  $k$  is strictly increasing in  $A$ .*

**Proposition 2** (Informational exclusion). *Let  $b^* = 0$  and suppose Assumption 1 holds. Then  $P_O < P_I$ , and the outsider share  $S_O$  is strictly decreasing in  $A$ .*

The intuition concerns the upper tail. Bayesian shrinkage compresses noisier evidence toward the prior mean, so outsiders' posterior distribution has less mass above a selective threshold. Execution capital, equally visible for both groups, dilutes this asymmetry; as its return vanishes, the asymmetry in conceptual visibility governs selection. The result requires neither prejudice nor any difference in ability. An institution staffed by unbiased Bayesians becomes more insular solely because the residual object of selection is less legible from outside, and by Lemma 1 it is hard to read because outside reports cannot be believed.

**Proposition 3** (Favoritism multiplier). *For any  $b^* > 0$ , the allocation under capability  $A$  coincides with that under visibility ratio  $k(\alpha(A))$  and premium*

$$b^{\text{eq}}(A) = b^* \sqrt{\frac{V_O(\alpha(0))}{V_O(\alpha(A))}} > b^* \quad \text{for } A > 0, \quad (11)$$

*evaluated at pre-AI visibility. Moreover,  $S_O$  is strictly decreasing in  $\tilde{b}$ . Hence AI amplifies the exclusionary effect of any fixed premium.*

Table 1 illustrates magnitudes for  $\tau = \sigma_a = \alpha_0 = 1$ ,  $\rho_y^{-1/2} = 1.5$ ,  $(\rho_y + \rho_s)^{-1/2} = 0.5$ , outsiders constituting 80% of applicants, and an appointment rate of 5%. Absent any premium, the outsider share falls by twenty percentage points. A premium of  $b^* = 0.3$  costs outsiders ten points before AI and nineteen points under full automation.

Table 1: Outsider share of appointments,  $S_O$ : baseline model

| AI capability $A$   | $b^* = 0$ | $b^* = 0.3$ | Standardized premium $\tilde{b}$ |
|---------------------|-----------|-------------|----------------------------------|
| 0 (pre-AI)          | 70.7%     | 60.6%       | 0.26                             |
| 0.5                 | 60.8%     | 45.4%       | 0.40                             |
| 0.8                 | 52.7%     | 34.2%       | 0.51                             |
| 1 (full automation) | 50.3%     | 31.1%       | 0.54                             |

Notes: Parameters are illustrative, not calibrated:  $\tau = \sigma_a = \alpha_0 = 1$ , outsider posterior noise 1.5, insider posterior noise 0.5,  $m_O / (m_I + m_O) = 0.8$ , appointment rate 5%.

## 5 Apprenticeship: formation and revelation

The baseline treats the outsider pool as fixed. Historically, however, the most important route by which outsiders became insiders was apprenticeship: execution work for members of the institution. Apprenticeship performed two distinct functions. It *formed* conceptual capital, because judgment is acquired by working alongside those who exercise it. And it *revealed* conceptual capital, because an apprentice acquires a sponsor who observes her soft signal and can credibly transmit it. We now show that AI, by reducing demand for execution labor, removes both functions.

### 5.1 Setup

Before selection, members hire a mass  $D(A)$  of outsiders as apprentices, with  $D'(A) < 0$  because AI substitutes for the execution tasks apprentices perform. Apprentices are selected on hard evidence of execution capital: an outsider is hired if  $a \geq \bar{a}(A)$ , where  $m_O[1 - \Phi(\bar{a}/\sigma_a)] = D(A)$ . Apprenticeship has two effects.

- *Formation*: the apprentice’s conceptual capital rises from  $\theta$  to  $\theta + \lambda$ , with  $\lambda \geq 0$ . Insiders, trained within the institution, receive the same increment.
- *Revelation*: the apprentice’s employer becomes her sponsor, so her evidence becomes  $(a, y, s)$  and she receives premium  $b^*$ .

Outsiders thus partition into apprentices, of mass  $D(A)$ , who are informationally indis-

tinguishable from insiders, and non-apprentices, of mass  $m_O - D(A)$ , who are evaluated on hard evidence alone. Let  $P_{ap}$  and  $P_{na}$  denote their appointment probabilities.

**Proposition 4** (Decomposition). *The outsider share satisfies*

$$\frac{dS_O}{dA} = \underbrace{\frac{D'(A)(P_{ap} - P_{na})}{K}}_{\text{extensive margin}} + \underbrace{\frac{D(A)\partial_A P_{ap} + (m_O - D(A))\partial_A P_{na}}{K}}_{\text{intensive margin}}. \quad (12)$$

*Whenever apprentices are more likely to be appointed than non-apprentices, the collapse of apprenticeship reduces the outsider share, independently of the visibility effect of Proposition 2. Moreover, the mean conceptual capital of outsiders relative to insiders is*

$$\mu_I - \mu_O(A) = \lambda \left( 1 - \frac{D(A)}{m_O} \right), \quad (13)$$

*which is increasing in  $A$ .*

The second part of the proposition is the more important one. The decline of apprenticeship creates a *real* difference in conceptual capital between groups. An observer who compares insiders and outsiders ex post will find that insiders are better on average, and this will be accurate. The institution's exclusion of outsiders is thereby rendered statistically justified, even though it originates in the disappearance of a training channel rather than in any primitive difference between groups.

## 5.2 Discouraged investment

The formation channel interacts with candidates' own investment incentives. Suppose that, before evaluation, each candidate chooses an investment  $x_\theta \geq 0$  in conceptual capital and  $x_a \geq 0$  in execution capital at convex cost  $C(x_\theta + x_a)$ , raising  $\theta$  and  $a$  one-for-one. The institution anticipates equilibrium investment and adjusts its priors, as in signal-jamming models (Holmström, 1999). Let  $U$  denote the value of appointment and  $f_g$  the density of group  $g$ 's score at its cutoff.

**Proposition 5** (Discouraged investment). *In equilibrium, the marginal return to investment in conceptual capital for a candidate in group  $g$  is  $U\beta_g f_g$ , and the marginal return to investment in execution capital is  $U\alpha f_g$ . Suppose  $b^* = 0$  and  $c^2 \geq V_I$ . Then:*

- (i) *outsiders' marginal return to conceptual investment is strictly below insiders';*
- (ii) *the ratio of outsiders' to insiders' maximal marginal return to human-capital investment is  $f_O / f_I$  when  $\alpha \geq \beta_I$ , and  $(\beta_O f_O) / (\beta_I f_I)$  when  $\alpha \leq \beta_O$ ; the latter is strictly smaller.*

Before AI, an outsider could invest in a dimension whose returns were fully visible regardless of her origin. After AI, the only remunerative dimension is one in which her returns are attenuated by the shrinkage factor  $\beta_O$ . The outsider's incentive to invest falls relative to the insider's, outsiders invest less, and the institution's expectation that outsiders are weaker is confirmed. This is the logic of [Coate and Loury \(1993\)](#), except that here the trigger is a change in the verifiability of the skill that matters.

## 6 Influence activities

If a sponsor's recommendation determines appointment, candidates have an incentive to influence it, and the means of influence need not be intellectual. A candidate may supply services of personal value to her sponsor: administrative labor, accommodation of the sponsor's schedule, assistance in matters unrelated to research. Following [Milgrom \(1988\)](#), we call these *influence activities*. They require proximity, so they are available to insiders and apprentices but not to other outsiders.

### 6.1 Setup

Each proximate candidate has an influence efficiency  $e \geq 0$ , distributed according to  $G_e$  independently of  $(\theta, a)$ . Providing services of value  $t$  to her sponsor costs her  $t/e$ . The

sponsor's private benefit from appointing the candidate becomes  $\pi + t$ . Let  $U$  denote the candidate's value of appointment.

**Proposition 6** (Influence as a random premium). *A proximate candidate with posterior value  $W$  is appointed if and only if*

$$W + b^* + B \geq c, \quad B \equiv \frac{NU}{\gamma} e. \quad (14)$$

*Equivalently, a candidate with  $W < c - b^*$  is appointed if and only if her efficiency exceeds*

$$e^*(W) = \frac{\gamma(c - b^* - W)}{NU}. \quad (15)$$

*In standardized units  $\omega = W / \sqrt{V_I}$  and  $\zeta = (c - b^*) / \sqrt{V_I}$ ,*

$$e^*(\omega) = \frac{\gamma}{NU} \sqrt{V_I(\alpha(A))} (\zeta - \omega), \quad (16)$$

*so that, holding the standardized cutoff fixed, the efficiency required to bridge any given rank gap falls with  $A$  by the factor  $\sqrt{V_I(\alpha(A)) / V_I(\alpha(0))}$ . The value of services required,  $t^* = \gamma(c - b^* - W) / N$ , falls by the same factor.*

Influence thus acts as a random, privately purchased premium, uncorrelated with merit. Its effect is subject to the same multiplier as in Proposition 3: the compression of the selection index makes favor cheaper. In our parametrization, full automation lowers the required efficiency at every standardized rank by one-third.

## 6.2 The composition of entrants

Define three types of appointee: *conceptual talents*, whose posterior on  $\theta$  alone clears the cutoff; *execution talents*, who clear it by virtue of  $\alpha a$ ; and *cultivators*, who clear it only through  $B$ .

Table 2: Outsider share and quality with all channels

| A   | Outsider share $S_O$ |             |                  | All channels | Quality shortfall |
|-----|----------------------|-------------|------------------|--------------|-------------------|
|     | Baseline             | + Influence | + Apprenticeship |              |                   |
| 0   | 70.7%                | 63.5%       | 70.4%            | 69.0%        | 0.22              |
| 0.5 | 60.8%                | 49.5%       | 52.3%            | 45.8%        | 0.36              |
| 0.8 | 52.7%                | 39.0%       | 37.4%            | 27.0%        | 0.41              |
| 1   | 50.3%                | 36.0%       | 31.4%            | 18.6%        | 0.42              |

Notes: Monte Carlo with  $3 \times 10^6$  candidates;  $b^* = 0$  to isolate the new channels. Quality shortfall is the difference in mean  $\theta$  of appointees between the symmetric-information, no-influence benchmark and the all-channels economy, in units of  $\tau$ .

**Proposition 7** (Composition). *Holding the standardized cutoff fixed:*

- (i) *as  $A \rightarrow 1$ , the mass of execution talents vanishes;*
- (ii) *at every standardized rank below the cutoff, the minimum efficiency of a cultivator falls with  $A$ , so the set of cultivators expands from those with exceptionally high  $e$  to all those with  $e$  above a strictly lower threshold;*
- (iii) *the expected conceptual capital of cultivators is below that of other appointees, so their rising share lowers the average quality of appointees relative to the symmetric-information benchmark.*

Proposition 7 predicts who enters. Before AI, the institution admitted conceptual talents, execution talents, and a small number of exceptionally skilled cultivators. After AI, it admits conceptual talents and a broader class of moderately skilled cultivators. The execution talents, historically the main route for able outsiders, are the ones displaced.

Table 2 reports simulations combining all channels, with  $\lambda = 0.3$ , a pre-AI apprenticeship mass equal to 8% of the population declining linearly to zero, and  $B$  exponentially distributed with mean 0.2 among proximate candidates. With every channel active, the outsider share falls from 69% to 19%, the share of appointees who enter through influence rises from 21% to 24%, and the quality shortfall relative to the symmetric-information benchmark nearly doubles.

## 7 AI as a verifier

The preceding sections treat AI purely as a substitute for execution. AI is also a potential technology of *evaluation*. This section asks under what conditions it reverses the forces identified above.

### 7.1 A common public test

Suppose AI provides a publicly observed evaluation of conceptual capital,  $T = \theta + \xi$  with  $\xi \sim N(0, \rho_T(A)^{-1})$  and  $\rho'_T(A) \geq 0$ . Being hard evidence,  $T$  is available for both groups. Public output is subject to two opposing forces: AI enables outsiders to produce and disseminate work, but it also multiplies the volume of submissions competing for finite evaluative attention  $H$ . We write

$$\rho_y(A)^{-1} = \sigma_0^2 - d(A) + g(n(A)/H), \quad (17)$$

with  $d' > 0$  (production),  $n' > 0$  and  $g' > 0$  (congestion). Insiders' soft channel is insulated from congestion. Precisions become  $\rho_O = \rho_y + \rho_T$  and  $\rho_I = \rho_y + \rho_T + \rho_s$ .

**Proposition 8** (Opening versus closure). *Let  $b^* = 0$ .*

- (i) *The insiders' experiment remains Blackwell-superior for every  $\rho_T$ , yet  $k \rightarrow 1$  and  $S_O \rightarrow m_O / (m_I + m_O)$  as  $\rho_T \rightarrow \infty$ .*
- (ii) *Writing  $x = \alpha^2 \sigma_a^2$ , the institution becomes more closed at the margin,  $dS_O/dA < 0$ , if and only if*

$$\frac{d}{dA} \log(\beta_I - \beta_O) > \frac{d}{dA} \log(x + \beta_O \tau^2), \quad (18)$$

*where  $\beta_I - \beta_O = \tau^4 \rho_s / [(1 + \tau^2 \rho_I)(1 + \tau^2 \rho_O)]$ .*

Part (i) distinguishes two notions that are easily confused. A common verifiable test does not eliminate the insiders' informational advantage in the Blackwell sense; it renders

that advantage immaterial in value terms. Part (ii) says that AI closes institutions when it erodes the verifiable component of value faster than it builds verifiable channels for ideas, or when congestion of public channels outpaces both.

## 7.2 Designable tests and persuasion

A public test may also be designable. The outsider's home institution may choose the protocol by which its students' ideas are assessed (which projects to submit, which benchmarks to apply, how results are reported) and commit to it, for instance by delegating evaluation to an AI system whose specification is public. The home institution maximizes placement; the elite institution is Bayesian. This is a problem of Bayesian persuasion (Kamenica and Gentzkow, 2011).

To isolate the persuasion logic, take  $A = 1$ , so  $v = \theta$ , let the outsider's conceptual capital have an arbitrary distribution  $G$  with continuous density and mean  $\mu_O$ , and suppose the elite appoints an outsider if and only if her posterior mean is at least  $c > \mu_O$ , taking  $c$  as given.

**Proposition 9** (Optimal verifiable test). *The placement-maximizing verifiable test is a pass-fail standard that passes the candidate if and only if  $\theta \geq \hat{\theta}$ , where*

$$\mathbb{E}[\theta \mid \theta \geq \hat{\theta}] = c. \quad (19)$$

*It places a fraction  $q^* = 1 - G(\hat{\theta})$  of outsiders. Moreover,*

$$q^* > 1 - G(c) > 0, \quad (20)$$

*where  $1 - G(c)$  is placement under full disclosure of  $\theta$  and 0 is placement under any soft channel (Lemma 1).*

A perfectly precise but soft assessment of an outsider's ideas is of no use to her, while

Table 3: Outsider placement under alternative evaluation regimes

| Cutoff $c$ (s.d. above mean) | Soft channel | Full disclosure | Optimal verifiable test |
|------------------------------|--------------|-----------------|-------------------------|
| 0.5                          | 0            | 30.9%           | 69.8%                   |
| 1.0                          | 0            | 15.9%           | 38.1%                   |
| 1.5                          | 0            | 6.7%            | 16.6%                   |
| 2.0                          | 0            | 2.3%            | 5.8%                    |

*Notes:* Standard normal prior over outsiders' conceptual capital; elite appoints if posterior mean  $\geq c$ , taken as fixed.

a coarse but verifiable one is worth more than full disclosure. Table 3 reports placement rates for a standard normal prior: for a cutoff one standard deviation above the outsider mean, optimal design more than doubles placement relative to full disclosure, from 15.9% to 38.1%.

*Remark 2* (Equilibrium and competition). With a fixed capacity, persuasion by outsiders' institutions raises the elite's cutoff, which partially offsets placement gains; the elite is Bayesian, so its expected surplus from a persuaded marginal appointee is zero. When many outside institutions design tests competitively, the familiar forces of competition in persuasion push toward more disclosure ([Gentzkow and Kamenica, 2017](#)), and the problem acquires the structure of the grading and certification models of [Ostrovsky and Schwarz \(2010\)](#) and [Boleslavsky and Cotton \(2015\)](#).

*Remark 3* (Manipulation). An AI evaluator that candidates can game generates evidence whose informativeness depends on heterogeneity in gaming ability ([Frankel and Kartik, 2019](#)). If gaming ability is correlated with proximity, as influence efficiency is, manipulation reintroduces the insiders' advantage through the verifiable channel. The design of evaluation that is robust to manipulation is therefore part of the problem of mobility.

*Remark 4* (AI as a generator of ideas). If AI also generates ideas, decompose conceptual capital as  $\theta = \iota + j$ , where  $\iota$  is idea generation and  $j$  is judgment about which ideas to pursue. Substitution of  $\iota$  lowers its value weight in the same way as substitution of  $a$  does. If, as seems plausible, judgment is less verifiable than generation, the analysis applies with

more force, because what remains to be selected on is even harder to observe.

## 8 Reputation

### 8.1 The reputational commons

The premium  $b^* = N\pi/\gamma$  of Proposition 1 reflects an externality. Each member bears only a share  $1/N$  of the reputational cost of a weak appointment.

**Proposition 10** (Reputational commons). *An institution that fully internalized reputational effects would set the premium  $b^{FB} = \pi/\gamma$ , and  $b^{FB} = 0$  if private benefits carry no institutional weight. The equilibrium premium exceeds  $b^{FB}$  by the factor  $N$ . With influence activities, the effective premium  $b^* + B$  is increasing in  $N$  in the sense of first-order stochastic dominance.*

Larger institutions are more susceptible to favoritism and influence because each member internalizes less of the reputational consequence, even if members are no more partial.

### 8.2 Dynamics

External audiences, such as prospective students, funders, journals and employers, update reputation from the realized quality of appointees. Let  $Q_t$  be the expected conceptual capital of the cohort appointed at  $t$ , and

$$R_{t+1} = (1 - \delta)R_t + \delta Q_t. \quad (21)$$

**Proposition 11** (Delayed decline). *Let  $Q^{\text{sym}}(A)$  denote appointee quality under symmetric information and no premium, and  $L(A) \equiv Q^{\text{sym}}(A) - Q(A) \geq 0$  the selection loss. Following a permanent increase in  $A$ , reputation converges geometrically, at rate  $1 - \delta$ , to  $R^* = Q^{\text{sym}}(A) -$*

$L(A)$ . The selection loss  $L$  is increasing in the standardized premium and in the mass of influence entrants.

Two features of the transition matter. First, there is a *reputational overhang*: for a period of order  $1/\delta$ , incumbents enjoy a reputation accumulated under a more open regime, which weakens pressure for reform. Second, the formation channel partially *masks* decline. Because insiders do acquire more conceptual capital (Proposition 4), the institution's appointees remain strong in absolute terms even as the selection loss  $L$  grows. Audiences that benchmark the institution against outsiders, rather than against what the institution could achieve, will detect the decline late.

**Proposition 12** (Entry). *Suppose a challenger commits to evaluating outsiders through verifiable channels, raising its outsider precision to  $\rho_O^C > \rho_O$ , and admits no influence. If its steady-state quality exceeds the incumbent's,  $Q^C > Q^E$ , its reputation eventually exceeds the incumbent's, regardless of initial reputations.*

The theory therefore does not predict permanent entrenchment. Elite status slowly shifts toward institutions that keep channels of entry legible. The welfare cost is concentrated in the transition, when talent is misallocated.

## 9 Empirical implications

The model yields predictions that distinguish it from accounts based solely on rising competition.

- P1.** *Declining outsider share.* The share of entrants to elite doctoral programs and research positions from non-elite institutions falls after the diffusion of generative AI, more strongly in fields where AI substitutes more heavily for execution.
- P2.** *Collapse of apprenticeship.* Entry-level execution positions for outsiders contract before changes in appointment outcomes become visible, and the outsider share among

appointees falls most where apprenticeship was the dominant route.

- P3.** *Heterogeneity by evaluation format.* The decline is smaller where selection relies on verifiable work products than on letters and interviews (Lemma 1, Proposition 9).
- P4.** *Size and insularity.* Insider shares and the dispersion of appointee quality rise more in larger institutions (Proposition 10).
- P5.** *Outcome test.* Under pure informational exclusion, the expected performance of *marginal* insider and outsider appointees is equal, even when formation makes *average* insiders better. Under favoritism or influence, marginal outsiders outperform. This Becker-type test discriminates between mechanisms and is robust to the formation channel.
- P6.** *Composition.* Post-AI, the lower tail of insider appointees thickens, and lower-tail quality correlates with measures of service-oriented proximity to sponsors (Proposition 7).

## 10 Conclusion

Access to elite institutions depends largely on verifiability. Generative AI devalues the most verifiable form of human capital and leaves selection resting on soft information that only insiders' sponsors can credibly transmit. Four consequences follow: outsiders are under-selected even by unbiased evaluators; any given favoritism has a larger effect; the dissolution of apprenticeship creates real and self-confirming group differences; and patronage becomes cheaper to purchase. Because reputation is a commons, these forces erode quality slowly and are partly masked.

The analysis also suggests remedies. Making evaluation more *verifiable* matters more than making it more precise: Proposition 9 shows that a coarse verifiable standard can outperform perfect precision delivered through soft channels. Institutions can develop verifiable channels for conceptual capital, such as structured proposals assessed against

explicit criteria or audited AI-based evaluation; reconstitute apprenticeship around conceptual rather than executional tasks; limit discretion through blind initial review; protect public channels from congestion; disclose appointee outcomes by entry route to internalize reputational costs; and randomize among candidates near the threshold, where the expected cost in quality is small. Whether AI opens or closes elite institutions will depend largely on whether it is used to verify ideas or only to carry out execution.

## A Proofs

**Proof of Lemma 1.** Fix hard evidence  $h$  and an equilibrium. Let  $p(m)$  be the probability of appointment following message  $m$ . The referee's payoff from  $m$  is  $\pi_O p(m)$ , independent of  $s$ . Hence every message sent with positive probability by any type must attain  $\bar{p} = \sup_m p(m)$ , and appointment occurs with probability  $\bar{p}$  regardless of  $s$ . The outcome coincides with that of the equilibrium in which the institution ignores messages and appoints with probability  $\bar{p}$  given  $h$ .  $\square$

**Proof of Proposition 1.** Under reciprocal deference the sponsor's recommendation determines appointment, so she recommends if and only if  $\pi + (\gamma/N)(W - c) \geq 0$ , i.e.  $W \geq c - N\pi/\gamma$ . Her recommendation is a function of  $W = \mathbb{E}[v \mid a, y, s]$ , and appointments of insiders are therefore determined by  $(a, y, s)$ ; for outsiders, Lemma 1 implies they are determined by  $(a, y)$ . Since  $(a, y)$  is a coordinate projection of  $(a, y, s)$ ,  $\mathcal{E}_O$  is a garbling of  $\mathcal{E}_I$ . Gaussian updating with independent signals gives  $\rho_I = \rho_y + \rho_s$  and  $\rho_O = \rho_y$ .  $\square$

**Proof of Theorem 1.** (i) Garbling means  $\mathcal{E}_O$  is conditionally independent of  $\omega$  given  $\mathcal{E}_I$ ; hence  $\mathbb{E}[v \mid \mathcal{E}_I, \mathcal{E}_O] = W_I$  and, by the law of iterated expectations,  $W_O = \mathbb{E}[W_I \mid \mathcal{E}_O]$ . Conditional Jensen yields  $\mathbb{E}\phi(W_I) \geq \mathbb{E}\phi(W_O)$  for every convex  $\phi$ . (ii) Apply (i) with  $\phi(x) = (x - c)^+$ . (iii) For  $c \geq t^*$ ,  $P_I = 1 - F_I(c) \geq 1 - F_O(c) = P_O$  by definition of a rotation. Under Gaussian evidence,  $F_g = N(0, V_g)$  with  $V_I > V_O$  when  $\rho_s > 0$ , so the distributions cross exactly once, at 0, and the inequality is strict for  $c > 0$ . (iv) Write  $W_g/\alpha = a + M_g/\alpha$  with  $M_g = \mathbb{E}[\theta \mid \mathcal{E}_g]$ . Since  $\mathbb{E}[M_g^2] \leq \mathbb{E}[\theta^2] < \infty$ ,  $M_g/\alpha \rightarrow 0$  in  $L^2$ , so  $W_g/\alpha \Rightarrow a$  for both groups. The market-clearing cutoff, rescaled by  $\alpha$ , converges to the corresponding quantile of  $a$ ; continuity of the distribution of  $a$  implies that both groups' appointment probabilities converge to the same limit.  $\square$

**Proof of Lemma 2.** Let  $x = \alpha^2 \sigma_a^2$ ,  $p = \beta_I \tau^2$ ,  $q = \beta_O \tau^2$ ,  $p > q$ . Then  $k = (x + p)/(x + q)$  and  $dk/dx = (q - p)/(x + q)^2 < 0$ . Since  $x$  is increasing in  $\alpha$  and  $\alpha'(A) < 0$ ,  $k$  is increasing in  $A$ .  $\square$

**Proof of Proposition 2.** At  $c = 0$  the institution would appoint  $(m_I + m_O)/2 > K$ , so  $c > 0$ . As  $V_I > V_O$ ,  $P_I > P_O$ . Let  $z = c/\sqrt{V_I} > 0$ , so  $P_I = 1 - \Phi(z)$  and  $P_O = 1 - \Phi(\sqrt{k}z)$ . Differentiating (9),

$$\frac{dz}{dk} = -\frac{m_O \phi(\sqrt{k}z) z / (2\sqrt{k})}{m_I \phi(z) + m_O \sqrt{k} \phi(\sqrt{k}z)} < 0.$$

Hence  $P_I$  increases in  $k$ ; since  $m_I P_I + m_O P_O = K$ ,  $m_O P_O$  and  $S_O$  decrease in  $k$ , and thus in  $A$  by Lemma 2.  $\square$

**Proof of Proposition 3.** Dividing scores by  $\sqrt{V_O}$  preserves rankings; standardized outsider scores are  $N(0, 1)$  and insider scores  $N(0, k) + \tilde{b}$ , so the allocation depends only on  $(k, \tilde{b})$ . Fixing the outsider cutoff, an increase in  $\tilde{b}$  raises insider appointments above  $K$ ; the cutoff must rise, and  $P_O$  falls. Since  $V_O$  is increasing in  $\alpha$ ,  $\tilde{b}$  increases in  $A$ ; equating standardized premia gives  $b^{\text{eq}}$ .  $\square$

**Proof of Proposition 4.** The share identity  $KS_O = DP_{ap} + (m_O - D)P_{na}$  follows from the partition of outsiders. Differentiating gives the decomposition, whose first term is negative when  $D' < 0$  and  $P_{ap} > P_{na}$ . Mean conceptual capital among outsiders is  $\lambda D(A)/m_O$  and among insiders is  $\lambda$ , which yields the gap, increasing in  $A$  since  $D' < 0$ .  $\square$

**Proof of Proposition 5.** A deviation to investment  $x_\theta$  shifts the candidate's posterior mean by  $\beta_g(x_\theta - \hat{x}_\theta)$ , where  $\hat{x}_\theta$  is anticipated investment, and leaves its distribution otherwise unchanged; a deviation in  $x_a$  shifts it by  $\alpha(x_a - \hat{x}_a)$ , since  $a$  is hard. The marginal effects on appointment probability are therefore  $\beta_g f_g$  and  $\alpha f_g$ . (i) With  $b^* = 0$ ,  $f_g(c) = \phi(c/\sqrt{V_g})/\sqrt{V_g}$ , and  $\partial \log f_g / \partial V_g = (c^2 - V_g)/(2V_g^2)$ , which is non-negative for  $V_g \leq V_I \leq c^2$ . Hence  $f_I \geq f_O$  and, with  $\beta_I > \beta_O$ ,  $\beta_I f_I > \beta_O f_O$ . (ii) Each candidate's maximal

marginal return is  $Uf_g \max\{\alpha, \beta_g\}$ . If  $\alpha \geq \beta_I$ , the ratio is  $f_O/f_I$ ; if  $\alpha \leq \beta_O$ , it is  $\beta_O f_O/(\beta_I f_I)$ , strictly smaller since  $\beta_O < \beta_I$ .  $\square$

**Proof of Proposition 6.** By Proposition 1 with private benefit  $\pi + t$ , the sponsor recommends if and only if  $W \geq c - N(\pi + t)/\gamma$ . The candidate's minimal transfer securing appointment is  $t^* = \max\{0, \gamma(c - W)/N - \pi\}$ , worthwhile if and only if  $t^*/e \leq U$ . Hence she is appointed iff  $W + b^* + NUe/\gamma \geq c$ . Solving for  $e$  gives  $e^*(W)$ , and substituting  $W = \omega\sqrt{V_I}$ ,  $c - b^* = \zeta\sqrt{V_I}$  gives the standardized expression, which scales with  $\sqrt{V_I}$ , decreasing in  $A$ .  $\square$

**Proof of Proposition 7.** (i) Execution talents satisfy  $\alpha a \geq c - \mathbb{E}[\theta \mid \mathcal{E}_g]$  while  $\mathbb{E}[\theta \mid \mathcal{E}_g] < c$ ; as  $\alpha \rightarrow 0$  with the standardized cutoff fixed, their mass converges to zero. (ii) Follows from Proposition 6. (iii) Cultivators are appointed with  $W < c - b^*$ , while other appointees have  $W \geq c - b^*$ ; since  $B$  is independent of  $\omega$ ,  $\mathbb{E}[\theta \mid \text{cultivator}] = \mathbb{E}[\mathbb{E}[\theta \mid W] \mid W < c - b^*, W + B \geq c - b^*] < \mathbb{E}[\mathbb{E}[\theta \mid W] \mid W \geq c - b^*]$ . Displacement of higher- $W$  candidates by cultivators lowers mean appointee quality relative to the benchmark.  $\square$

**Proof of Proposition 8.** (i)  $(a, y, T)$  is a projection of  $(a, y, T, s)$ , so the Blackwell ranking persists. As  $\rho_T \rightarrow \infty$ ,  $\beta_I, \beta_O \rightarrow 1$ ,  $k \rightarrow 1$ , and by Theorem 1(iii) and Proposition 2 the appointment probabilities equalize. (ii) Since  $k - 1 = (\beta_I - \beta_O)\tau^2/(x + \beta_O\tau^2)$ ,  $k$  increases in  $A$  iff the stated log-derivative condition holds; by Proposition 2,  $S_O$  moves inversely with  $k$ . The expression for  $\beta_I - \beta_O$  follows from  $\beta_g = \tau^2\rho_g/(1 + \tau^2\rho_g)$ .  $\square$

**Proof of Proposition 9.** Fix any verifiable test and let  $\mathcal{A}$  be the event that the posterior mean is at least  $c$ , with  $q = \mathbb{P}(\mathcal{A})$ . Then  $\mathbb{E}[\theta \mathbf{1}_{\mathcal{A}}] = \mathbb{E}[\mathbb{E}[\theta \mid \text{test}] \mathbf{1}_{\mathcal{A}}] \geq cq$ . Among all events of probability  $q$ , the upper tail  $\{\theta \geq G^{-1}(1 - q)\}$  maximizes  $\mathbb{E}[\theta \mathbf{1}_{\cdot}]$ , so  $\mathbb{E}[\theta \mid \theta \geq G^{-1}(1 - q)] \geq c$ . The left side is continuous and strictly decreasing in  $q$  and equals  $\mu_O < c$  at  $q = 1$ ; hence  $q \leq q^*$ , where  $q^*$  solves the equation with equality. The pass-

fail test with threshold  $\hat{\theta} = G^{-1}(1 - q^*)$  attains  $q^*$ . Full disclosure places  $1 - G(c)$ ; since  $\mathbb{E}[\theta \mid \theta \geq c] > c$ , the conditional-mean equation implies  $\hat{\theta} < c$ , so  $q^* > 1 - G(c)$ . Placement under soft channels is zero because, by Lemma 1, the elite's posterior equals the prior mean  $\mu_O < c$ .  $\square$

**Proof of Proposition 10.** An institution internalizing reputation replaces  $\gamma/N$  by  $\gamma$  in (5), giving  $b^{FB} = \pi/\gamma$ . Since  $B = NUe/\gamma$  is increasing in  $N$  for each  $e$ ,  $b^* + B$  increases in  $N$  in the first-order stochastic sense.  $\square$

**Proof of Proposition 11.** For  $\delta \in (0, 1)$  the law of motion is a stable linear difference equation with fixed point  $R^* = Q = Q^{\text{sym}} - L$ ; deviations decay geometrically at rate  $1 - \delta$ .  $L$  increases in  $\tilde{b}$  by Proposition 3 and in the mass of influence entrants by Proposition 7(iii).  $\square$

**Proof of Proposition 12.** Each institution's reputation converges to its own steady state independently of initial conditions; the ranking of limits is the ranking of steady-state qualities.  $\square$

## References

- Acemoglu, D., and P. Restrepo (2018). The race between man and machine: Implications of technology for growth, factor shares, and employment. *American Economic Review*.
- Agrawal, A., J. Gans, and A. Goldfarb (2019). Exploring the impact of artificial intelligence: Prediction versus judgment. *Information Economics and Policy*.
- Aigner, D., and G. Cain (1977). Statistical theories of discrimination in labor markets. *Industrial and Labor Relations Review*.
- Azoulay, P., C. Fons-Rosen, and J. Graff Zivin (2019). Does science advance one funeral at a time? *American Economic Review*.
- Blackwell, D. (1953). Equivalent comparisons of experiments. *Annals of Mathematical Statistics*.
- Boleslavsky, R., and C. Cotton (2015). Grading standards and education quality. *American Economic Journal: Microeconomics*.
- Brogaard, J., J. Engelberg, and C. Parsons (2014). Networks and productivity: Causal evidence from editor rotations. *Journal of Financial Economics*.
- Coate, S., and G. Loury (1993). Will affirmative-action policies eliminate negative stereotypes? *American Economic Review*.
- Colussi, T. (2018). Social ties in academia: A friend is a treasure. *Review of Economics and Statistics*.
- Cornell, B., and I. Welch (1996). Culture, information, and screening discrimination. *Journal of Political Economy*.
- Crawford, V., and J. Sobel (1982). Strategic information transmission. *Econometrica*.
- Dessein, W. (2002). Authority and communication in organizations. *Review of Economic Studies*.
- Frankel, A., and N. Kartik (2019). Muddled information. *Journal of Political Economy*.
- Gentzkow, M., and E. Kamenica (2017). Competition in persuasion. *Review of Economic Studies*.
- Heckman, J., and S. Moktan (2020). Publishing and promotion in economics: The tyranny of the top five. *Journal of Economic Literature*.
- Holmström, B. (1999). Managerial incentive problems: A dynamic perspective. *Review of Economic Studies*.
- Johnson, J., and D. Myatt (2006). On the simple economics of advertising, marketing, and product design. *American Economic Review*.

- Kamenica, E., and M. Gentzkow (2011). Bayesian persuasion. *American Economic Review*.
- Korinek, A. (2023). Generative AI for economic research: Use cases and implications for economists. *Journal of Economic Literature*.
- Laband, D., and M. Piette (1994). Favoritism versus search for good papers: Empirical evidence regarding the behavior of journal editors. *Journal of Political Economy*.
- Li, D. (2017). Expertise versus bias in evaluation: Evidence from the NIH. *American Economic Journal: Applied Economics*.
- Milgrom, P. (1988). Employment contracts, influence activities, and efficient organization design. *Journal of Political Economy*.
- Milgrom, P., and J. Roberts (1988). An economic approach to influence activities in organizations. *American Journal of Sociology*.
- Ostrovsky, M., and M. Schwarz (2010). Information disclosure and unraveling in matching markets. *American Economic Journal: Microeconomics*.
- Phelps, E. (1972). The statistical theory of racism and sexism. *American Economic Review*.
- Spence, M. (1973). Job market signaling. *Quarterly Journal of Economics*.
- Tirole, J. (1996). A theory of collective reputations (with applications to the persistence of corruption and to firm quality). *Review of Economic Studies*.