Titles as Signals of Expertise: Information Aggregation and Hierarchy Depth

Zijian Qi

March 22, 2026

Abstract

This paper studies how organizations use hierarchical titles as signals of expertise in collective decision-making. In the model, workers receive noisy signals about a common state, and the organization evaluates expertise through imperfect tests whose outcomes are summarized by a finite set of observable titles. Titles affect decisions by shaping how individual signals are weighted in information aggregation. We show that titles act as coarse signals of expertise, allowing organizations to concentrate decision authority on more reliable contributors and thereby improve decision accuracy. The analysis highlights a tradeoff in organizational design. More intensive evaluation and finer rank differentiation improve information aggregation, but the gains from both margins exhibit strong diminishing returns. As a result, when evaluation and hierarchy are costly, the optimal organization employs limited screening and a relatively shallow hierarchy. Relatively simple title structures capture most of the attainable gains from information aggregation.

Key words: Organizational design, Hierarchy depth, Information aggregation, Decision authority, Expertise, weighted majority

JEL codes: D23, D83, D71

1 Introduction

Modern organizations increasingly make important decisions in teams whose members have limited firsthand knowledge of one another's expertise. Employee–employer relationships are often short-lived—average job tenure in the U.S. private sector is under four years (U.S. Bureau of Labor Statistics, 2024)—and many firms rely on decentralized and team-based forms of collaboration that draw on specialized knowledge from different units (Bloom and Van Reenen, 2010; Bloom, Sadun, and Van Reenen, 2012b). At the same time, organizations frequently adjust their internal hierarchical structures and reporting relationships in response to technological change, market conditions, and competitive pressures, including changes in spans of control and the number of managerial layers (Rajan and Wulf, 2006; Bloom, Sadun, and Reenen, 2012a). Together, these features limit opportunities for long-term learning about coworkers' capabilities and make it common for individuals to collaborate with colleagues they have not worked with before.

Decision-making in such environments is particularly challenging. When collaborators lack reliable, experience-based information about one another's competence, they cannot easily assess whose judgments are more informative. Yet many organizational decisions are complex and high-stakes, so treating all opinions as equally reliable is rarely optimal. As a result, organizations face a fundamental design problem: how to structure influence within teams when expertise is heterogeneous but not directly observable.

In the absence of direct knowledge of expertise, organizations rely on institutional substitutes to guide collective decision-making. Formal titles—such as manager, director, or vice president—are especially prominent because they are publicly observable and organizationally endorsed. Although titles also govern reporting relationships and career incentives, they can be viewed as coarse summaries of prior evaluations of performance and ability. In practice, titles shape how information is aggregated: teams rarely treat all contributions symmetrically, and higher-ranked individuals tend to receive greater influence when opinions conflict or evidence is noisy. This creates an implicit aggregation rule in which rank determines decision authority, raising a natural design question: how finely should organizations differentiate titles, and how much information should they invest in doing so, in order to improve decision quality without incurring unnecessary organizational cost.

The study of hierarchy has deep roots in organizational economics. Early work emphasizes hierarchy as a response to bounded rationality and coordination constraints, allowing complex decision problems to be structured across organizational levels (Simon, 1947). Subsequent formal models characterize hierarchy as an information-processing architecture that economizes on communication and coordination costs in large organizations (Radner, 1993). More recent work focuses on three related roles of hierarchy: as a mechanism for allocating authority under communication frictions (Garicano and Prat, 2011), as an organizational response to incentive and monitoring problems (Garicano and Rossi-Hansberg, 2006), and as a determinant of managerial structure and spans of control in response to technological and competitive forces (Rajan and Wulf, 2006; Bloom et al., 2012a). Across these contributions, hierarchy serves a variety of organizational functions, but its role as a device for structuring information aggregation remains comparatively underexplored.

As a result, there is limited guidance on how the depth and granularity of hierarchy should be designed when titles primarily act as signals of expertise. Existing work provides little insight into how finely organizations should differentiate ranks, or how the informational gains from more detailed evaluations trade off against the organizational costs of maintaining a more complex hierarchy. When titles summarize imperfect assessments of ability, additional hierarchical layers can sharpen distinctions in expected expertise and improve decision quality, but they may also increase administrative burden and organizational complexity. Understanding this tradeoff re quires treating hierarchy depth and evaluation intensity as explicit design choices in an information aggregation problem.

This paper studies a collective decision-making environment in which an organization aggregates assessments from a large number of individuals whose expertise is heterogeneous and not directly observable. Individuals observe noisy signals about a common underlying state, and the organization can invest in evaluations—such as tests, performance reviews, or certifications—that imperfectly reveal individual expertise. Evaluation outcomes are summarized by a finite set of observable titles, which serve as coarse indicators of expected expertise. Titles play no role in incentives, reporting lines, or communication; instead, they affect decisions only by shaping how individual assessments are weighted in the collective choice.

Building on standard results in the information aggregation literature (Nitzan and Paroush, 1982; Shapley and Grofman, 1984), we take the aggregation rule as given and focus on the organizational design problem. Using this framework, we study how decision accuracy responds to two design margins: the intensity of evaluation, which governs how informative titles are about expertise, and the depth of hierarchy, which determines how finely evaluation outcomes are translated into rank categories. We show that both margins improve decision quality by concentrating influence on more reliable individuals, but that their informational benefits exhibit sharply diminishing returns. Consequently, the optimal organizational design features a finite level of screening and a finite depth of hierarchy: relatively simple title structures capture most of the attainable gains from information aggregation, while further differentiation adds complexity with little additional benefit.

Related literature: This paper contributes to a large literature on organizational hierarchy by isolating and formalizing one specific function of hierarchical structure: its role in aggregating dispersed information when individual expertise is heterogeneous and imperfectly observed. A broad range of existing work studies hierarchy through other lenses, including bounded rationality, incentives and monitoring, authority allocation, communication, and problem solving. While these contributions provide deep insights into why hierarchies arise and how they shape organizational behavior, they typically abstract from the informational design problem that is central to our analysis.

As mentioned earlier, the study of hierarchy has deep roots in organizational economics.

Early work emphasizes hierarchy as an efficient response to bounded rationality and coordination constraints in complex organizations. Simon (1947) argues that hierarchical decomposition allows organizations to cope with cognitive limits by structuring complex decisions across multiple levels. Radner (1993) formalizes this perspective by modeling hierarchy as an information-processing architecture that economizes on communication and coordination costs in large organizations, with related contributions emphasizing the role of hierarchy in managing complexity and coordinating large systems (Marschak and Radner, 1972; Van Zandt, 1997). In this foundational literature, hierarchical structure is typically treated as an organizational primitive: the depth and granularity of hierarchy are taken as given, rather than as design variables chosen to improve decision accuracy.

A large body of work studies hierarchy as a mechanism for allocating authority, providing incentives, and mitigating monitoring problems within organizations. Classic models analyze how layered hierarchical structures address moral hazard, free riding, and effort provision in teams (Holmstrom, 1982; Mookherjee, 1984; McAfee and McMillan, 1991), while subsequent contributions examine how authority, ownership, and control interact with incentives under incomplete contracts (Baker, Gibbons, and Murphy, 1999, 2002). Related work emphasizes hierarchy as an allocation of decision rights under communication frictions, analyzing how authority and delegation shape organizational performance when information is costly to transmit or acquire (Bolton and Dewatripont, 1994; Aghion and Tirole, 1997). More recent theoretical and empirical contributions study how delegation, centralization, and managerial structure respond to communication constraints, technological change, and competitive pressures (Dessein, 2002; Rantakari, 2008; Alonso and Matouschek, 2008; Alonso, Dessein, and Matouschek, 2008; Bloom et al., 2012a). Across this literature, hierarchy primarily governs incentives, authority, and control. By contrast, we abstract entirely from incentives, monitoring, strategic communication, and delegation, and instead focus on hierarchy as an informational device that structures how dispersed assessments are aggregated into a collective decision.

A closely related strand of the literature studies how collective decisions aggregate heterogeneous information. Early contributions by Sah and Stiglitz (1986, 1988) compare hierarchical and polyarchical decision structures in terms of their error properties. Work by Nitzan and Paroush (1982) and Shapley and Grofman (1984) characterizes optimal weighting schemes for aggregating signals of unequal reliability, showing that differential influence can improve decision accuracy when information quality differs. Related results appear in models of voting and information aggregation with heterogeneous expertise (Condorcet, 1785; Austen-Smith and Banks, 1996). While these papers establish the value of unequal weighting, they typically take the aggregation structure as given and do not study how hierarchical categories or titles should be designed endogenously, nor how the depth of hierarchy trades off against organizational costs.

A related strand of the literature studies hierarchy as a mechanism for allocating problem-solving tasks and specialized knowledge. In influential work, Garicano (2000) models organizations as knowledge hierarchies in which lower-level agents handle routine problems while rare or complex cases are escalated to more knowledgeable specialists. Subsequent research examines how communication technology and cognitive constraints shape optimal hierarchical depth and expertise specialization (Garicano and Rossi-Hansberg, 2006; Garicano and Prat, 2011), as well as related models of problem routing and knowledge allocation (Beggs, 2001). In this literature, hierarchy governs the assignment and escalation of problems across levels, whereas we study how hierarchy shapes the aggregation of many simultaneous assessments into a single collective decision.

Our paper complements these literatures by studying hierarchy exclusively as a mechanism for structuring information aggregation when individual expertise is heterogeneous and imperfectly observed. We model formal titles as coarse summaries of noisy evaluations of expertise and allow organizations to choose both how much information to generate about expertise and how finely to translate that information into hierarchical categories. Unlike existing work, hierarchy in our model does not govern reporting lines, delegation, incentives, or monitoring, and communication is costless conditional on titles. By endogenizing the depth of hierarchy and the assignment of evaluations to titles, we characterize when additional layers of hierarchy improve decision accuracy and when their marginal informational value is outweighed by organizational costs. This perspective highlights an additional and distinct function of hierarchy—information filtering—that organizations must consider alongside incentives, authority, and coordination when designing hierarchical structures.

The remainder of the paper is organized as follows. Section 2 presents the model: a binary decision problem with heterogeneous and unobservable expertise, a noisy evaluation technology, and the induced title structure. Section 3 analyzes decision-making under a fixed evaluation-and-title system, characterizes the optimal title-based aggregation rule, and derives a closed-form expression for the asymptotic error exponent that summarizes decision performance. Section 4 studies the organization's design problem, examining how performance changes with evaluation intensity and hierarchy depth and characterizing optimal choices under cost constraints. Section 5 discusses the implications of the results for organizational design and hierarchy. Section 6 concludes.

2 Model Setup

We develop a model of organizational decision-making under uncertainty in which an organization aggregates information from multiple agents whose expertise is heterogeneous and not directly observable. Agents provide assessments that differ in reliability, and the organization uses observable titles—generated through imperfect evaluation—to structure decision authority when forming a collective choice. The model specifies the state space, agents' information, the evaluation and title-assignment process, and the aggregation rule used to reach a decision.

State and Decision

The organization faces a binary decision regarding a project or action. The underlying state of the world is denoted by $\omega \in \{-1, 1\}$, where the labels -1 and 1 are purely notational. The organization holds a symmetric prior,

$$\Pr(\omega = 1) = \Pr(\omega = -1) = \frac{1}{2}.$$

The organization chooses an action

$$d \in \{-1, 1\},$$

and the organization's payoff depends on whether the chosen action matches the state. Specifically,

$$\pi(d, \omega) = \begin{cases} 1, & \text{if } d = \omega, \\ \\ 1 - \kappa, & \text{if } d \neq \omega, \end{cases}$$

where $\kappa > 0$ measures the loss from an incorrect decision.

To capture that decision stakes may rise with project scale and may vary across projects, we let

$$\kappa = \lambda^n,$$

where n is the number of workers that involved in this decision and $\lambda > 1$ is a random variable capturing the severity of losses conditional on failure. We assume that λ is drawn from a uniform distribution on $[1, 2]$, independently of all other primitives. The realization of λ reflects project-specific factors (e.g., technological risk, regulatory exposure, or market conditions) and is observed by the organization before the final action is taken.

Finally, the organization may choose not to proceed with a project when its expected payoff is negative. Throughout, expected performance is evaluated relative to an outside option normalized to zero.

Workers and Information

To improve the quality of the decision, the organization relies on signals produced by its workers. There are n workers indexed by $i = 1, 2, \ldots, n$, each of whom generates a binary signal about the underlying state. Workers differ in their expertise, which affects the reliability of the information they produce.

Worker i has an expertise level $\theta_i \in (\frac{1}{2}, 1)$, and it is drawn independently from a common distribution F with density f, supported on $(\frac{1}{2}, 1)$. Expertise level θ_i is not directly observable to the organization or to other workers, but the distribution F is known to the organization.

Conditional on the underlying state ω, worker i produces a signal $s_i \in \{-1, 1\}$, where

$$\Pr(s_i = \omega \mid \omega, \theta_i) = \theta_i.$$

Each signal is therefore informative about the underlying state, though its precision varies across workers. Conditional on the state, the signals $\{s_i\}_{i=1}^{n}$ are independent. Each signal represents primary information generated through task execution, analysis, or judgment, rather than being filtered through organizational hierarchy.

The organization relies on signals produced by its workers to improve the quality of the decision. To isolate the efficiency of informationl aggregation, we abstract from incentive provision and strategic communication, and focus on how organizations aggregate heterogeneous information when individual expertise is unobservable.

Evaluation and Titles

Before aggregating signals, the organization may invest in evaluating workers' expertise. The evaluation technology is a test consisting of m independent items. Conditional on expertise level θ_i, worker i answers each item correctly with probability θ_i, independently across items.

Let $J_i \in \{0, 1, \ldots, m\}$ denote worker i's number of correct answers. Then

$$J_i \sim \text{Binomial}(m, \theta_i).$$

The organization observes J_i but not θ_i.

Based on the evaluation result J_i, the organization assigns each worker a title from L possible titles. The number of titles L corresponds to the depth of the organizational hierarchy. An assignment rule is a map

$$r : \{0, 1, \ldots, m\} \to \{1, 2, \ldots, L\},$$

and worker i's title is

$$K_i := r(J_i) \in \{1, 2, \ldots, L\}.$$

Titles are publicly observable and are interpreted as coarse indicators of expected expertise.

The assignment rule r maps evaluation outcomes into titles and thereby determines the ex-ante distribution of expertise across titles. For each title $k \in \{1, \ldots, L\}$, this yields a probability p_k, a posterior expertise density f_k, and an associated mean expertise ξ_k, defined as follows:

$$p_k := \Pr(K_i = k), \qquad f_k(\theta_i) := f(\theta_i \mid K_i = k), \qquad \xi_k := \mathbb{E}[\theta_i \mid K_i = k].$$

These summarize how the evaluation-and-title system compresses information about expertise.

Implementing the evaluation-and-title system entails organizational costs, summarized by a cost function

$$C = C(m, L),$$

which depends on the intensity of evaluation m and the number of titles L. The cost reflects resources devoted to administering evaluations and maintaining the title structure. We assume that the cost is weakly increasing in both arguments, so that $C_m(\cdot) \geq 0$ and $C_L(\cdot) \geq 0$, reflecting higher costs associated with more intensive evaluations and more complex title systems.

Both the payoff and the organizational cost can be interpreted on a per-worker basis. The payoff from a correct decision is normalized to one and does not scale with organizational size, while the costs reflect resources incurred for each worker—each is subject to evaluation testing and must be trained to operate within the title system. Thus, the model captures a tradeoff between improved aggregation efficiency and organization structural costs.

Aggregation Rule

The organization aggregates the information generated by its workers to form a collective decision. Individual signal s_i provides noisy information about the underlying state, and observable titles affect how these signals are aggregated into the final choice.

Formally, the organization's decision rule maps the profile of signals into an action. An aggregation rule is therefore written as

$$d = \delta(s_1, \ldots, s_n;\, K_1, \ldots, K_n),$$

where δ is a function from assessment profiles to a binary action, conditional on the realized titles.

The aggregation rule captures how the organization combines information when forming a collective decision. Titles enter as parameters that shape the relative influence of individual signals, while the signals themselves are the primary carriers of information.

Organizational Design Problem

The organization chooses its internal decision structure to maximize expected payoffs. An organizational design (m, L, r, δ) specifies the intensity of evaluation m, the hierarchy depth L, the title assignment rule r, and the signal aggregation rule δ. The components (m, L, r) define the evaluation-and-title system: they determine how much information about expertise is elicited and how it is coarse-grained into observable ranks. Given these realized titles, the aggregation rule δ governs how individual signals are combined into a collective decision.

Given a design (m, L, r, δ) and a realization of λ, the organization's expected payoff from undertaking the project is $\mathbb{E}[\pi(d, \omega) \mid \lambda]$. The organization may choose to discontinue the project if its expected value is negative. As a result, the ex ante value of a design is given by $\mathbb{E}_\lambda[\max\{\mathbb{E}[\pi(d, \omega) \mid \lambda], 0\}]$.

The organizational design problem can be written as:

$$\max_{m,L,r,\delta} \mathbb{E}_\lambda \Big[\max \big\{ \mathbb{E}[\pi(d, \omega) \mid \lambda],\, 0 \big\} \Big] \;-\; C(m, L).$$

3 Analysis

This section analyzes the organization's decision-making problem given an evaluation-and-title system. We first characterize the aggregation rule δ that maximizes the probability of a correct collective decision. We then derive a tractable measure of decision performance and use it to study the organization's choice of (m, L, r). Finally, we introduce benchmark cases that bound achievable performance and serve as reference points for evaluating the effectiveness of any evaluation-and-title system.

3.1 Optimal Aggregation Rule

Fix an evaluation-and-title system (m, L, r). The evaluation and title assignment induce the ex-ante title probability $\{p_k\}_{k=1}^L$ and the corresponding mean expertise levels $\{\xi_k\}_{k=1}^L$, where

$\xi_k = \mathbb{E}[\theta_i \mid K_i = k]$. Conditional on the state, workers' signals are independent and satisfy

$$\Pr(s_i = \omega \mid \omega, K_i = k) = \xi_k.$$

Thus, conditional on titles, workers holding the same title have identical signal distributions and are statistically indistinguishable from the organization's perspective. Once titles are realized, the mean expertise level ξ_k fully summarizes the informativeness of signals produced by workers with title k.

The aggregation problem is therefore to determine how signals should be combined given the realized title profile. As shown in Lemma 1, the optimal rule takes the form of a weighted majority rule with title-dependent log-likelihood weights.

Lemma 1 (Optimal Signal Aggregation Rule)**.** *Fix an evaluation-and-title system* (m, L, r) *and consider the choice of aggregation rule* δ*. An aggregation rule* δ^* *maximizes the probability of a correct decision* $\Pr(d = \omega)$ *when it takes the weighted majority form*

$$d = \delta^*(s_1, \ldots, s_n; K_1, \ldots, K_n) = \operatorname{sgn}\left(\sum_{i=1}^{n} w_{K_i}\, s_i\right),$$

where the weight assigned to title k *is the log-likelihood ratio*

$$w_k = \log\left(\frac{\xi_k}{1 - \xi_k}\right),$$

and $\xi_k = \mathbb{E}[\theta_i \mid K_i = k]$ *is the mean expertise of workers with title* k *induced by* (m, L, r)*. If the weighted sum equals zero, the organization randomizes uniformly over* $\{-1, 1\}$*.*

Lemma 1 is a standard likelihood-ratio result in weighted voting with heterogeneous signal precision (see, e.g., Nitzan and Paroush, 1982; Shapley and Grofman, 1984). Under conditional independence, each signal contributes additively to the log-likelihood ratio of the state, and titles associated with higher expertise receive larger weights.

3.2 Decision Performance and Organizational Design

With the aggregation rule pinned down by Lemma 1, decision performance depends only on the evaluation-and-title system (m, L, r). The purpose of this subsection is to derive a tractable representation of performance that allows the organizational design problem to be written directly as a function of (m, L, r).

In large organizations, decision accuracy is governed by the exponential rate at which the probability of an incorrect collective decision declines with team size. This motivates the following measure of performance.

Fix (m, L, r) and consider the weighted majority rule in Lemma 1. Let

$$\Pr_n(m, L, r) := \Pr(d \neq \omega \mid m, L, r)$$

denote the probability of an incorrect decision with n workers under this rule. The probability is taken over the joint distribution of $(\omega, \{\theta_i\}_{i=1}^n, \{J_i\}_{i=1}^n, \{s_i\}_{i=1}^n)$, where titles are given by $K_i = r(J_i)$ and the decision is $d = \delta(s_1, \ldots, s_n; K_1, \ldots, K_n)$.

Definition 1 (Performance rate)**.** *The performance rate is defined as*

$$\phi(m, L, r) \ := \ -\lim_{n\to\infty} \frac{1}{n} \log \Pr_n(m, L, r),$$

whenever the limit exists.

When the limit exists, the error probability declines exponentially with team size:

$$\Pr_n(m, L, r) \approx e^{-n\phi(m,L,r)} \quad \text{as } n \to \infty.$$

A larger value of $\phi(m, L, r)$ therefore corresponds to a faster decline in the probability of an incorrect collective decision as the organization grows.

The performance rate can be expressed directly in terms of the title distribution and the expertise associated with each title. The following proposition provides the resulting characterization.

Proposition 1 (Performance Rate under optimal title-based weights)**.** *Under the weighted majority rule in Lemma 1, the performance rate ϕ exists and satisfies*

$$\phi(m, L, r) = -\log\left[\sum_{k=1}^{L} 2p_k \sqrt{\xi_k(1-\xi_k)}\right].$$

Proposition 1 expresses decision performance as a function of the evaluation-and-title system (m, L, r). For any fixed system, the performance rate depends on the induced title probabilities p_k and the mean expertise levels ξ_k associated with each title.

The performance rate also yields a simple representation of the organization's expected payoff. Recall that the loss from an incorrect decision is $\kappa = \lambda^n$, where $\lambda \sim U[1, 2]$ is observed before the final action is taken. Under the weighted rule,

$$\mathbb{E}[\pi(d, \omega) \mid \lambda] = 1 - \lambda^n \Pr_n(m, L, r).$$

Using the definition of $\phi(m, L, r)$, this expression can be approximated as

$$\mathbb{E}[\pi(d, \omega) \mid \lambda] \approx 1 - e^{n(\log \lambda - \phi(m,L,r))}.$$

As $n \to \infty$, $\mathbb{E}[\pi(d, \omega) \mid \lambda]$ converges to 1 if $\log \lambda < \phi(m, L, r)$ and diverges to $-\infty$ if $\log \lambda > \phi(m, L, r)$. With the outside option normalized to zero, the organization therefore proceeds with the project if and only if $\lambda \leq e^{\phi(m,L,r)}$.

Since $\lambda \sim U[1, 2]$, the ex ante value of (m, L, r) is

$$\mathbb{E}_\lambda[\max\{\mathbb{E}[\pi(d, \omega) \mid \lambda], 0\}] = e^{\phi(m,L,r)} - 1.$$

In the large-organization limit, the project yields payoff 1 whenever it is undertaken.

The ex ante value therefore equals the probability that the project is accepted, which is $\Pr(\lambda \leq e^{\phi(m,L,r)}) = e^{\phi(m,L,r)} - 1$.

The organizational design problem therefore reduces to choosing (m, L, r) to maximize performance net of organizational costs:

$$\max_{m,L,r} \; e^{\phi(m,L,r)} - 1 - C(m, L).$$

3.3 Benchmarks

This subsection introduces benchmark cases that bound achievable decision performance. Because the organization's objective separates into the performance component $e^{\phi(m,L,r)} - 1$ and the organizational cost $C(m, L)$, the benchmarks focus on the attainable range of the performance term. Using the performance rate ϕ defined in Section 3.2, we characterize an upper bound ϕ^{FB} corresponding to full information about expertise and a lower bound ϕ^{NE} corresponding to the absence of evaluation and hierarchy differentiation.

First-best benchmark. Suppose individual expertise θ_i is publicly observed and can therefore be used directly in aggregation. In this case no evaluation or title system is required and no organizational cost is incurred. The organization assigns each signal the likelihood-ratio weight

$$w_i = \log\left(\frac{\theta_i}{1 - \theta_i}\right),$$

and aggregates signals using the corresponding optimal likelihood-ratio rule. Because expertise is observed, the aggregation rule fully exploits all available information about signal reliability.

Let ϕ^{FB} denote the resulting performance rate. Standard large-deviation arguments imply that

$$\phi^{FB} = -\log\left[\int_{1/2}^{1} 2\sqrt{\theta(1-\theta)}\, dF(\theta)\right],$$

where F is the distribution of expertise. When F is the uniform distribution on $[\frac{1}{2}, 1]$, the corresponding value of the objective is

$$U^{FB} = e^{\phi^{FB}} - 1 = \left[\int_{1/2}^{1} 4\sqrt{\theta(1-\theta)}\, d\theta\right]^{-1} - 1 = \frac{4}{\pi} - 1 \approx 0.273.$$

This case represents the maximal decision performance attainable in the model, since any feasible organizational design must rely on imperfect information about individual expertise.

No-evaluation benchmark. At the opposite extreme, consider a design with no evaluation and a single title level ($m = 0$, $L = 1$). In this case all workers receive the same title and are treated symmetrically in the aggregation rule. The mean expertise in the population is $\xi = \mathbb{E}[\theta_i]$, and the performance rate reduces to

$$\phi^{NE} = -\log\left[2\sqrt{\xi(1-\xi)}\right].$$

When $\theta_i \sim U[\frac{1}{2}, 1]$, the mean expertise is $\xi = 3/4$, which implies

$$U^{NE} = \left[2\sqrt{\xi(1-\xi)}\right]^{-1} - 1 = \frac{2}{\sqrt{3}} - 1 \approx 0.155.$$

This case represents decision-making without evaluation or hierarchical differentiation and provides a natural lower bound on attainable performance.

Together, these benchmarks define the range of attainable performance in the model. The no-evaluation case provides a natural lower bound, while the first-best benchmark represents the maximal performance achievable when expertise is fully observed. Evaluation and hierarchy improve decision performance by moving the organization away from the no-evaluation benchmark and toward the first-best bound, but cannot exceed it. The organizational design problem studied in this paper therefore concerns how much of this performance gap can be closed, and at what organizational cost.

4 Results

This section studies how the organization's design choices affect decision performance. The analysis in Section 3 provides a closed-form expression for the performance rate $\phi(m, L, r)$, which allows the organizational design problem to be evaluated numerically. While the assignment rule r is formally part of the design problem, it is not the primary object of comparative statics in the analysis. For each pair (m, L), we therefore compute numerically the assignment rule $r^*(m, L)$ that maximizes performance. All performance measures and objective values reported in this section correspond to this optimized assignment rule. The results below then examine how performance varies with evaluation intensity m and hierarchy depth L, and how organizational costs shape the optimal design (m^*, L^*) when the assignment rule is chosen optimally.

4.1 Performance Rate under Evaluation and Hierarchy

We begin by examining how the evaluation-and-title system affects decision performance in the absence of organizational costs, so that $C(m, L) = 0$. This exercise isolates the informational role of evaluation intensity m and hierarchy depth L in determining the performance rate $\phi(m, L, r^*)$.

Figure 1 plots the performance rate as a function of hierarchy depth L holding evaluation intensity fixed at $m = 300$. Increasing the number of titles initially improves performance substantially because the hierarchy allows the organization to differentiate workers with different expected expertise levels. However, the gains diminish quickly. Beyond a modest number of hierarchical levels, additional titles generate only small improvements in the performance rate.

Figure 2 examines the role of evaluation intensity. In this case hierarchy depth grows with evaluation according to $L = m + 1$, so that each possible test outcome corresponds to a distinct title. This specification allows the evaluation-and-title system to use all available in-

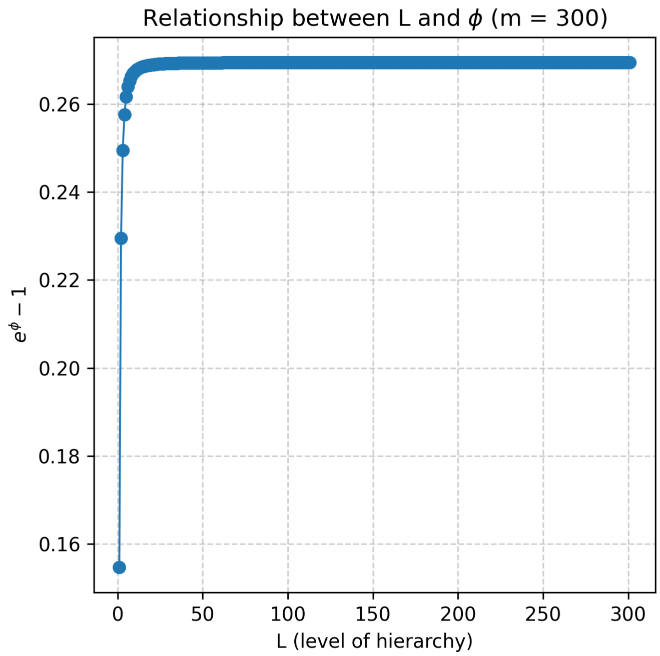


Figure 1: Performance rate as a function of hierarchy depth L for fixed evaluation intensity $m = 300$.

formation generated by the tests. As evaluation becomes more informative, the performance rate rises and approaches the first-best benchmark derived in Section 3.3. The increase is steep at low levels of testing and then flattens rapidly.

Figure 3 separates the effects of evaluation and hierarchy by holding hierarchy depth fixed. The figure plots the performance rate as a function of evaluation intensity for $L \in \{5, 10, 50\}$. Across all cases the performance rate is increasing and concave in m. More informative evaluations improve the allocation of decision authority, but the marginal value of additional testing declines quickly. Even relatively coarse hierarchies capture a large share of the attainable performance once evaluation becomes moderately informative.

Taken together, these results show that both evaluation intensity and hierarchy depth increase the performance rate by improving how information is aggregated. However, the informational gains from both dimensions exhibit strong diminishing returns. These patterns suggest that once organizational costs are introduced, the optimal evaluation-and-title system will involve finite evaluation intensity and a relatively shallow hierarchy.

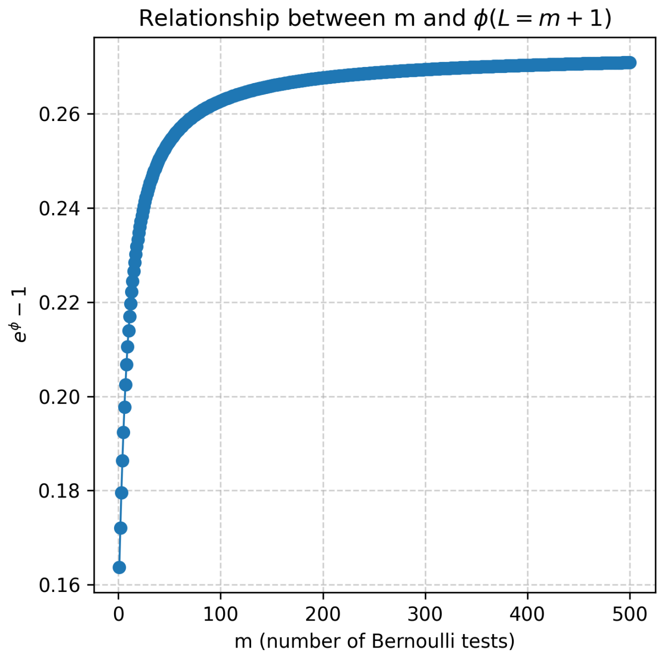
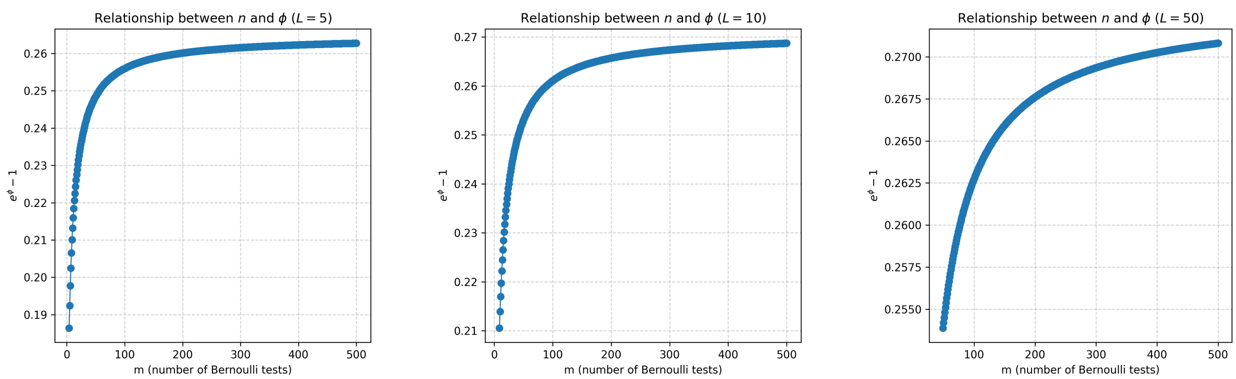


Figure 2: Performance rate as evaluation intensity m increases when hierarchy depth grows according to $L = m + 1$.



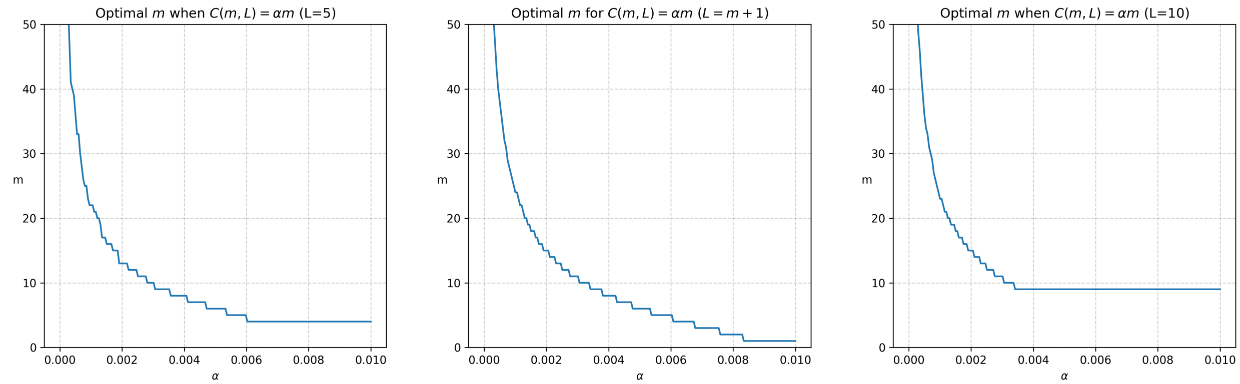
(a) $L = 5$ (b) $L = 10$ (c) $L = 50$

Figure 3: Performance rate as a function of evaluation intensity m for fixed hierarchy depths $L \in \{5, 10, 50\}$.

4.2 Optimal Evaluation Intensity under Evaluation Costs

We now introduce organizational costs and study how they shape the optimal amount of evaluation. Throughout this subsection, the cost of evaluation is linear in the number of tests,

$$C(m, L) = \alpha m,$$



(a) $L = 5$ (b) $L = m + 1$ (c) $L = 10$

Figure 4: Optimal evaluation intensity m^* as a function of the marginal testing cost α under linear evaluation costs.

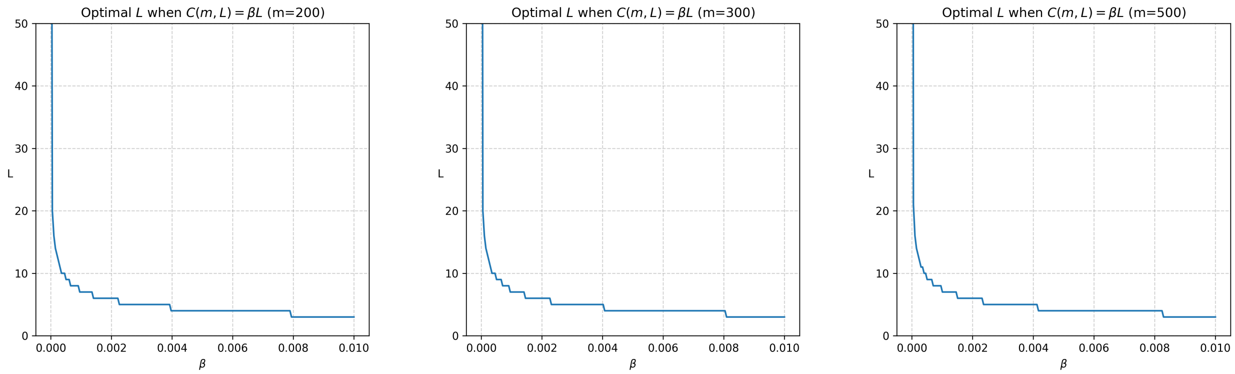
where $\alpha > 0$ measures the marginal cost of administering an additional test. For each value of α, the organization chooses m to maximize

$$U(m, L) = e^{\phi(m,L,r^*)} - 1 - \alpha m,$$

taking the hierarchy depth L as given. For each (m, L), the assignment rule r is optimized.

Figure 4 plots the optimal number of tests m^* as a function of α under three organizational environments. The left panel fixes hierarchy depth at $L = 5$, the middle panel considers the case in which hierarchy expands with evaluation according to $L = m+1$, and the right panel fixes hierarchy depth at $L = 10$.

Across all cases, the optimal evaluation intensity is decreasing in the marginal cost α. When testing is inexpensive, the organization invests heavily in evaluation to improve the allocation of decision authority across titles. As α increases, the marginal informational benefit of additional testing is outweighed by its direct cost, leading to discrete reductions in m^*. These stepwise patterns reflect both the integer nature of the design problem and the diminishing returns to evaluation documented in Section 4.1.



(a) $m = 200$ (b) $m = 300$ (c) $m = 500$

Figure 5: Optimal hierarchy depth L^* as a function of the marginal hierarchy cost β, for selected values of evaluation intensity m.

4.3 Optimal Hierarchy Depth under Hierarchy Costs

We next examine how the optimal depth of hierarchy responds to organizational costs. In this subsection, the cost function takes the form

$$C(m, L) = \beta L,$$

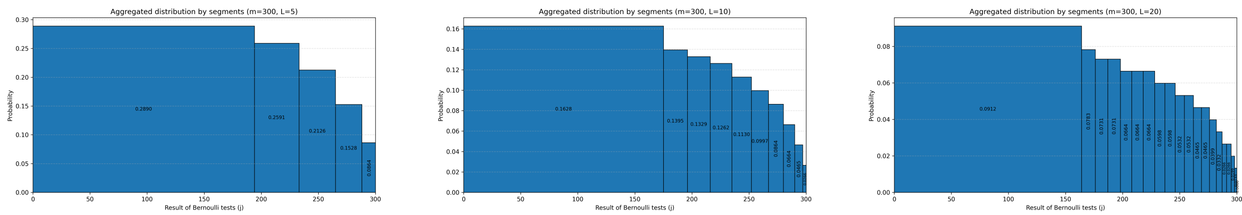
where $\beta > 0$ captures the marginal cost of maintaining an additional hierarchical level. Evaluation intensity m is fixed, and for each (β, m) the organization chooses the hierarchy depth L to maximize

$$U(m, L) = e^{\phi(m,L,r^*)} - 1 - \beta L.$$

with the assignment rule r optimized for each (m, L).

Figure 5 plots the optimal hierarchy depth L^* as a function of β for different fixed values of evaluation intensity m. Across all cases, the optimal hierarchy depth declines sharply as the marginal cost of hierarchy increases. When β is small, the organization adopts a relatively deep hierarchy in order to exploit fine distinctions in expected expertise. As β rises, the informational gains from additional levels are quickly dominated by organizational costs, leading to discrete reductions in L^*.

Two patterns are worth emphasizing. First, the optimal hierarchy depth is highly sen-



(a) $L = 5$ (b) $L = 10$ (c) $L = 20$

Figure 6: Aggregated assignment rules for $m = 300$ under different hierarchy depths.

sitive to even modest increases in β: most of the reduction in L^* occurs at low values of the marginal cost. Second, higher evaluation intensity supports deeper hierarchies, but only when the cost of hierarchy is sufficiently low. Once β exceeds a moderate threshold, the optimal structure collapses to a shallow hierarchy regardless of how informative evaluation is.

Together, these results highlight that hierarchy depth is the most fragile dimension of organizational design. While evaluation intensity can be adjusted smoothly in response to costs, the optimal number of hierarchical levels responds discretely and declines rapidly as coordination and communication costs increase.

4.4 Structure of the Optimal Title Assignment Rule

We focus on coarser hierarchies that are empirically more relevant and visually informative. The figures below display the aggregated assignment rules for $L = 5$, $L = 10$, and $L = 20$, holding evaluation intensity fixed at $m = 300$.

Figure 6 shows that, for $L = 5$, $L = 10$, and $L = 20$, the optimal assignment rule groups wide ranges of test outcomes into a small number of title categories. Lower titles pool a large mass of relatively uninformative evaluations, while higher titles are reserved for increasingly extreme test outcomes. As L increases, the partition becomes progressively finer, but the additional cuts occur primarily in the upper tail of the evaluation distribution.

Two patterns emerge. First, optimal assignment rules are monotone in evaluation outcomes: higher test scores are always mapped to weakly higher titles. Second, hierarchy depth primarily affects how finely the upper tail of the evaluation distribution is separated. Additional title categories are used to refine distinctions among high-performing workers, while lower-performing workers remain pooled into broad categories. This reflects diminishing returns to differentiation at low levels of expected expertise, where additional information has little effect on decision quality.

Overall, these results confirm that the role of the assignment rule is largely supportive. Once the aggregation rule is optimally chosen, the main determinants of performance are evaluation intensity and hierarchy depth. The precise shape of the assignment rule matters primarily insofar as it governs how finely high levels of expertise are distinguished, which explains why changes in r have smaller quantitative effects than changes in m or L.

5 Discussion

This paper studies hierarchy as a mechanism for allocating decision authority when individual expertise is heterogeneous and imperfectly observed. Titles summarize noisy evaluations of expertise and determine how individual assessments are weighted in collective decisions. The analysis highlights how organizational performance depends on three design choices: evaluation intensity, hierarchy depth, and the assignment of evaluations to titles.

A central finding is that both evaluation and hierarchy improve decision quality, but with strongly diminishing returns. More intensive evaluation sharpens posterior beliefs about expertise, while additional title categories allow the organization to translate these beliefs into finer distinctions in decision authority. However, once the most informative distinctions are already captured, further refinement yields little additional benefit. As a result, performance converges quickly toward an upper bound well below the first-best benchmark. This explains why relatively simple organizational structures can perform nearly as well as much more complex ones.

The results also reveal an important asymmetry between evaluation intensity and hierarchy depth. Evaluation can be adjusted relatively smoothly in response to costs, whereas hierarchy depth responds discretely. Small increases in the marginal cost of hierarchy lead to sharp reductions in the optimal number of title categories. This reflects the fact that hierarchy primarily refines distinctions among high-expertise workers: when hierarchy is costly, the organization optimally collapses many intermediate ranks while preserving a small number of top positions. In this sense, hierarchy depth is the most fragile dimension of organizational design.

The analysis further clarifies the role of the assignment rule. While the optimal rule is monotone and respects evaluation rankings, its detailed shape plays a secondary role once aggregation weights are optimally chosen. Changes in performance are driven mainly by how many distinctions the organization makes, not by the exact placement of thresholds. This helps explain why the numerical structure of assignment rules varies across environments without substantially affecting performance. From a design perspective, this suggests that organizations can rely on relatively simple and robust assignment procedures without sacrificing much efficiency.

Several limitations of the analysis point to directions for future work. First, the model is static and abstracts from learning and promotion over time. In practice, titles evolve as workers accumulate performance histories, and hierarchical structures may serve additional incentive or career concerns. Second, the evaluation technology is stylized as a sequence of independent Bernoulli tests. Other screening mechanisms—such as peer evaluation, project outcomes, or task-specific signals—may interact differently with hierarchy. Finally, the model treats communication costs in reduced form through the cost of hierarchy, rather than modeling explicit information loss across layers.

Despite these simplifications, the framework captures a fundamental tradeoff in organizational design. Hierarchy improves decision-making by concentrating authority among more reliable contributors, but only up to the point where organizational costs outweigh informational gains. The results provide a rationale for why many organizations adopt shallow hierarchies with limited differentiation, and why attempts to excessively refine internal rank structures often yield little improvement in decision quality.

6 Conclusion

This paper studies how organizations should design internal hierarchies when individual expertise is heterogeneous, imperfectly observed, and must be aggregated into a collective decision. Titles arise as a coarse summary of noisy evaluations of expertise and serve to allocate decision authority among contributors. By embedding hierarchy into a formal information-aggregation framework, the analysis clarifies how evaluation, rank differentiation, and collective decision-making interact.

The model delivers three main insights. First, both evaluation intensity and hierarchy depth improve decision accuracy by concentrating influence on more reliable contributors, but the gains exhibit strong diminishing returns. As a result, relatively simple organizational structures capture most of the achievable improvements in decision quality. Second, hierarchy depth responds sharply to organizational costs. Even modest increases in the marginal cost of hierarchy lead to substantial reductions in the optimal number of title categories, whereas evaluation intensity adjusts more smoothly. This makes hierarchy depth the most fragile dimension of organizational design. Third, while the optimal assignment of evaluations to titles is monotone and intuitive, its detailed shape plays a secondary role once aggregation weights are optimally chosen. Performance is driven primarily by how many distinctions the organization makes, rather than by the precise placement of thresholds.

Together, these results provide a disciplined explanation for why many organizations rely on shallow hierarchies with limited rank differentiation, even when decision-making is complex and expertise is unevenly distributed. Hierarchy improves decision quality by filtering noise and concentrating authority, but only up to the point where organizational costs outweigh informational gains. Beyond that point, additional layers contribute little and may be counterproductive.

More broadly, the framework highlights the informational role of hierarchy distinct from its incentive or control functions. By viewing titles as a mechanism for compressing information about expertise and structuring influence, the analysis offers a new perspective on organizational design in environments where direct knowledge of ability is limited and collaboration is transient.

References

Aghion, Philippe and Jean Tirole (1997), "Formal and real authority in organizations." *Journal of Political Economy*, 105, 1–29.

Alonso, Ricardo, Wouter Dessein, and Niko Matouschek (2008), "When does coordination require centralization?" *American Economic Review*, 98, 145–179.

Alonso, Ricardo and Niko Matouschek (2008), "Optimal delegation." *Review of Economic Studies*, 75, 259–293.

Austen-Smith, David and Jeffrey S. Banks (1996), "Information aggregation, rationality, and the condorcet jury theorem." *American Political Science Review*, 90, 34–45.

Baker, George, Robert Gibbons, and Kevin J. Murphy (1999), "Informal authority in organizations." *Journal of Law, Economics, and Organization*, 15, 56–73.

Baker, George, Robert Gibbons, and Kevin J. Murphy (2002), "Relational contracts and the theory of the firm." *Quarterly Journal of Economics*, 117, 39–84.

Beggs, Alan (2001), "Queues and hierarchies." *Review of Economic Studies*, 68, 297–322.

Bloom, Nicholas, Raffaella Sadun, and John Van Reenen (2012a), "The organization of firms across countries." *Quarterly Journal of Economics*, 127, 1663–1705.

Bloom, Nicholas, Raffaella Sadun, and John Van Reenen (2012b), "The organization of firms across countries." *Quarterly Journal of Economics*, 127, 1663–1705.

Bloom, Nicholas and John Van Reenen (2010), "Why do management practices differ across firms and countries?" *Journal of Economic Perspectives*, 24, 203–224.

Bolton, Patrick and Mathias Dewatripont (1994), "The firm as a communication network." *Quarterly Journal of Economics*, 109, 809–839.

Condorcet, Marquis de (1785), "An essay on the application of probability theory to plurality decision-making."

Dessein, Wouter (2002), "Authority and communication in organizations." *Review of Economic Studies*, 69, 811–838.

Garicano, Luis (2000), "Hierarchies and the organization of knowledge in production." *Journal of Political Economy*, 108, 874–904.

Garicano, Luis and Andrea Prat (2011), "Organizational economics with cognitive costs." Available at SSRN: 1846254.

Garicano, Luis and Esteban Rossi-Hansberg (2006), "Organization and inequality in a knowledge economy." *Quarterly Journal of Economics*, 121, 1383–1435.

Holmstrom, Bengt (1982), "Moral hazard in teams." *Bell Journal of Economics*, 324–340.

Marschak, Jacob and Roy Radner (1972), *Economic Theory of Teams*. Yale University Press.

McAfee, R. Preston and John McMillan (1991), "Optimal contracts for teams." *International Economic Review*, 32, 561–577.

Mookherjee, Dilip (1984), "Optimal incentive schemes with many agents." *Review of Economic Studies*, 51, 433–446.

Nitzan, Shmuel and Jacob Paroush (1982), "Optimal Decision Rules in Uncertain Dichotomous Choice Situations." *International Economic Review*, 23, 289.

Radner, Roy (1993), "The organization of decentralized information processing." *Econometrica*, 61, 1109–1146.

Rajan, Raghuram G. and Julie Wulf (2006), "The flattening firm: Evidence from panel data on the changing nature of corporate hierarchies." *Review of Economics and Statistics*, 88, 759–773.

Rantakari, Heikki (2008), "Governing adaptation." *Review of Economic Studies*, 75, 1257–1285.

Sah, Raaj Kumar and Joseph E. Stiglitz (1986), "The architecture of economic systems: Hierarchies and polyarchies." *American Economic Review*, 76, 716–727.

Sah, Raaj Kumar and Joseph E. Stiglitz (1988), "Committees, hierarchies and polyarchies." *Economic Journal*, 98, 451–470.

Shapley, Lloyd and Bernard Grofman (1984), "Optimizing group judgmental accuracy in the presence of interdependencies." *Public Choice*, 43, 329–343.

Simon, Herbert A (1947), *Administrative behavior*. Macmillan. Originally published 1947; revised editions 1957, 1976, 1997.

U.S. Bureau of Labor Statistics (2024), "Employee tenure summary." `https://www.bls.gov/news.release/tenure.nr0.htm`. Median tenure for wage and salary workers in the private sector is under four years.

Van Zandt, Timothy (1997), "Decentralized information processing in the theory of organizations." *Contemporary Economic Development Reviewed*, 4.

A Proofs

Proof of Proposition 1. From the main text, we know that $\phi(m, L, r)$ under a fixed evaluation-and-title system is

$$\phi(m, L, r) = -\lim_{n\to\infty} \frac{1}{n} \log \Pr_n(m, L, r) \quad \text{where } \Pr_n(m, L, r) = \Pr(d \neq \omega \mid m, L, r).$$

The key is to pin down $\Pr(d \neq \omega \mid m, L, r)$ under the weighted majority rule stated in Lemma 1. Recall that for each category $k \in \{1, \dots, L\}$, the population share p_k, the posterior density f_k, and the corresponding mean ability ξ_k, defined by

$$p_k = \Pr(K_i = k), \qquad f_k(\theta) = f(\theta \mid K_i = k), \qquad \xi_k = \mathbb{E}[\theta_i \mid K_i = k].$$

Without loss of generality, we assume that $\omega = 1$ and $\xi_1 < \xi_2 < \dots < \xi_L$. Now denote $w_k = \log \frac{\xi_k}{1-\xi_k}$, and X_i as the random variable of the weighted vote of agent i, where

$$X_i = \begin{cases} w_{K_i} & s_i = 1 \\ -w_{K_i} & s_i = -1 \end{cases}.$$

We can further write X_i as

$$X_i = \begin{cases} w_1 & \text{with probability } p_1\xi_1 \\ -w_1 & \text{with probability } p_1(1-\xi_1) \\ w_2 & \text{with probability } p_2\xi_2 \\ -w_2 & \text{with probability } p_2(1-\xi_2) \\ \dots & \dots \\ w_L & \text{with probability } p_L\xi_L \\ -w_L & \text{with probability } p_L(1-\xi_L) \end{cases}.$$

Let $S_n = X_1 + X_2 + \dots + X_n$, then the decision is $d = 1$ if $S_n > 0$, $d = -1$ if $S_n < 0$, and random when $S_n = 0$. Now by the weighted majority rule stated in Lemma 1, we know that

$$\Pr(d \neq \omega) = \Pr(S_n < 0) + \frac{1}{2}\Pr(S_n = 0).$$

As we will show in the later proof, $\lim_{n\to\infty} \frac{1}{n} \log \Pr(S_n < 0) = \lim_{n\to\infty} \frac{1}{n} \log \Pr(S_n \leq 0)$. Thus, we can conclude that

$$\phi(m, L, r) = -\lim_{n\to\infty} \frac{1}{n} \log \Pr(S_n < 0).$$

Moment generating function

Calculate the moment generating function:

$$M(t) = \mathbb{E}[e^{tX}] = \sum_{k=1}^{L} p_k \left(\xi_k e^{w_k t} + (1-\xi_k)e^{-w_k t}\right).$$

Lemma 2. *$M(t)$ is strictly convex, and so is $\log M(t)$.*

Proof. Taking derivatives, we obtain

$$M'(t) = \sum_{k=1}^{L} p_k w_k \Big(\xi_k e^{w_k t} - (1-\xi_k)e^{-w_k t}\Big),$$

$$M''(t) = \sum_{k=1}^{L} p_k w_k^2 \Big(\xi_k e^{w_k t} + (1-\xi_k)e^{-w_k t}\Big) > 0.$$

Thus $M(t)$ is strictly convex. Moreover,

$$\frac{d \log M(t)}{dt} = \frac{M'(t)}{M(t)}, \qquad \frac{d^2 \log M(t)}{dt^2} = \frac{M''(t)M(t) - (M'(t))^2}{(M(t))^2}.$$

Let

$$a_k = \sqrt{p_k \Big(\xi_k e^{w_k t} + (1-\xi_k)e^{-w_k t}\Big)}, \qquad b_k = w_k a_k.$$

By the Cauchy–Schwarz inequality,

$$M''(t)M(t) = \Big(\sum_{k=1}^{L} a_k^2\Big)\Big(\sum_{k=1}^{L} b_k^2\Big) \geq \Big(\sum_{k=1}^{L} a_k b_k\Big)^2.$$

We also have

$$\Big(\sum_{k=1}^{L} a_k b_k\Big)^2 = \Big(\sum_{k=1}^{L} p_k w_k \big(\xi_k e^{w_k t}+(1-\xi_k)e^{-w_k t}\big)\Big)^2 > \Big(\sum_{k=1}^{L} p_k w_k \big(\xi_k e^{w_k t}-(1-\xi_k)e^{-w_k t}\big)\Big)^2 = (M'(t))^2.$$

Therefore,

$$\frac{d^2 \log M(t)}{dt^2} = \frac{M''(t)M(t) - (M'(t))^2}{(M(t))^2} > 0,$$

so $\log M(t)$ is strictly convex. *Q.E.D.*

Lemma 3. *The range of* $\frac{d \log M(t)}{dt} = \frac{M'(t)}{M(t)}$ *is* $(-w_L, w_L)$.

Proof. We have already established that $\frac{d^2 \log M(t)}{dt^2} > 0$, so $\frac{M'(t)}{M(t)}$ is strictly increasing in t. Moreover,

$$\lim_{t\to-\infty} \frac{M'(t)}{M(t)} = \lim_{t\to-\infty} \frac{\sum_{k=1}^{L} p_k w_k \big(\xi_k e^{w_k t} - (1-\xi_k)e^{-w_k t}\big)}{\sum_{k=1}^{L} p_k \big(\xi_k e^{w_k t} + (1-\xi_k)e^{-w_k t}\big)} = -w_L,$$

and

$$\lim_{t\to\infty} \frac{M'(t)}{M(t)} = \lim_{t\to\infty} \frac{\sum_{k=1}^{L} p_k w_k \big(\xi_k e^{w_k t} - (1-\xi_k)e^{-w_k t}\big)}{\sum_{k=1}^{L} p_k \big(\xi_k e^{w_k t} + (1-\xi_k)e^{-w_k t}\big)} = w_L.$$

Therefore, the range of $\frac{M'(t)}{M(t)}$ is $(-w_L, w_L)$. *Q.E.D.*

Rate function

Denote the rate function as

$$I(x) := \sup_{t\in\mathbb{R}} tx - \log M(t).$$

Now let

$$g(x,t) = tx - \log M(t) = tx - \log\left[\sum_{k=1}^{L} p_k\Big(\xi_k e^{w_k t} + (1-\xi_k)e^{-w_k t}\Big)\right],$$

then $I(x) = \sup_{t\in\mathbb{R}} g(x,t)$.

When $x < -w_L$, we have $g(x,t) \to \infty$ as $t \to -\infty$, so $I(x) = \infty$.

When $x > w_L$, we have $g(x,t) \to \infty$ as $t \to \infty$, so $I(x) = \infty$.

When $x = -w_L$,

$$\begin{aligned} e^{g(x,t)} &= \frac{e^{-w_L t}}{\sum_{k=1}^{L} p_k\big(\xi_k e^{w_k t} + (1-\xi_k)e^{-w_k t}\big)} \\ &= \frac{1}{p_L\xi_L e^{2w_L t} + p_L(1-\xi_L) + \sum_{k=1}^{L-1} p_k\big(\xi_k e^{(w_k+w_L)t} + (1-\xi_k)e^{(w_L-w_k)t}\big)}. \end{aligned}$$

Thus, $e^{g(x,t)}$ is decreasing in t, so the supremum is attained as $t \to -\infty$, and

$$I(-w_L) = -\log\big(p_L(1-\xi_L)\big).$$

When $x = w_L$,

$$\begin{aligned} e^{g(x,t)} &= \frac{e^{w_L t}}{\sum_{k=1}^{L} p_k\big(\xi_k e^{w_k t} + (1-\xi_k)e^{-w_k t}\big)} \\ &= \frac{1}{p_L\xi_L + p_L(1-\xi_L)e^{-2w_L t} + \sum_{k=1}^{L-1} p_k\big(\xi_k e^{-(w_L-w_k)t} + (1-\xi_k)e^{-(w_L+w_k)t}\big)}. \end{aligned}$$

Thus, $e^{g(x,t)}$ is increasing in t, so the supremum is attained as $t \to \infty$, and

$$I(w_L) = -\log\big(p_L\xi_L\big).$$

When $x \in (-w_L, w_L)$, since $g(x,t)$ is concave in t, the supremum is attained at t^* satisfying

$$\frac{\partial g(x,t)}{\partial t} = x - \frac{M'(t)}{M(t)} = 0.$$

Since $\frac{M'(t)}{M(t)}$ is strictly increasing in t, there is a unique solution t^*. Thus, the rate function is

$$I(x) = \begin{cases} xt^* - \log M(t^*) & x \in (-w_L, w_L), \\ -\log\left(p_L(1-\xi_L)\right) & x = -w_L, \\ -\log\left(p_L \xi_L\right) & x = w_L, \\ \infty & x \notin [-w_L, w_L]. \end{cases}$$

Lemma 4. *t^* is continuous and increasing in x.*

Proof. $t^*(x)$ is solution to $x = \frac{M'(t)}{M(t)}$. $\frac{M'(t)}{M(t)}$ is continuous and strictly increasing in t, and the range is $(-w_L, w_L)$. So for each $x \in (-w_L, w_L)$, there is a unique $t^* \in \mathbb{R}$ such that $x = \frac{M'(t^*)}{M(t^*)}$. Also, $\lim_{t\to-\infty} \frac{M'(t)}{M(t)} = -w_L$, $\lim_{t\to\infty} \frac{M'(t)}{M(t)} = w_L$. Thus, $t^* \to -\infty$ as $x \to -w_L$, and $t^* \to \infty$ as $x \to w_L$.

By implicit function theorem, we have

$$\frac{dt^*}{dx} = \frac{1}{\frac{\partial^2 g(x,t)}{\partial t^2}} = \frac{1}{-\frac{M''(t)M(t)-(M'(t))^2}{(M(t))^2}} > 0.$$

So t^* is increasing in x. *Q.E.D.*

Lemma 5. *$I(x)$ is well-defined and continuous in x on $[-w_L, w_L]$.*

Proof. When $x \in (-w_L, w_L)$, $I(x)$ is well-defined since t^* is unique. Also, $I(x)$ is well-defined when $x = -w_L$ and $x = w_L$. Thus $I(x)$ is well-defined.

When $x \in (-w_L, w_L)$, we have t^* satisfies $x = \frac{M'(t^*)}{M(t^*)}$. Since t^* is a unique map between x

and t^*, we can write $I(x)$ as:

$$I(\frac{M'(t)}{M(t)}) = t \cdot \frac{M'(t)}{M(t)} - \log M(t).$$

Since $M'(t)$ and $M(t)$ are continuous in t, $I(\frac{M'(t)}{M(t)})$ is continuous in t. Thus, $I(x)$ is continuous on $(-w_L, w_L)$.

Let $t \to -\infty$,

$$\begin{aligned}
\lim_{x \to -w_L^+} e^{I(x)} &= \lim_{t \to -\infty} e^{t \cdot \frac{M'(t)}{M(t)} - \log M(t)} \\
&= \lim_{t \to -\infty} \frac{e^{t \cdot \frac{M'(t)}{M(t)}}}{M(t)} \\
&= \lim_{t \to -\infty} \frac{e^{t \cdot \frac{M'(t)}{M(t)}}}{\sum_{k=1}^{L} p_k(\xi_k e^{w_k t} + (1-\xi_k) e^{-w_k t})} \\
&= \lim_{t \to -\infty} \frac{1}{\sum_{k=1}^{L} p_k(\xi_k e^{(w_k - \frac{M'(t)}{M(t)})t} + (1-\xi_k) e^{(-w_k - \frac{M'(t)}{M(t)})t})}
\end{aligned}$$

Since $\lim_{t \to -\infty} \frac{M'(t)}{M(t)} = -w_L$, so for an $w^* \in (-w_L, -w_{L-1})$, there is a $\underline{t}$ such that when $t < \underline{t}$, $\frac{M'(t)}{M(t)} \in (-w_L, w^*)$.

Thus, we have

$$0 \le \lim_{t \to -\infty} e^{(-w_k - \frac{M'(t)}{M(t)})t} \le \lim_{t \to -\infty} e^{(-w_k - w^*)t} = 0, \quad k = 1, 2, ..., L-1.$$

Similarly,

$$0 \le \lim_{t \to -\infty} e^{(w_k - \frac{M'(t)}{M(t)})t} \le \lim_{t \to -\infty} e^{(w_k - w^*)t} = 0, \quad k = 1, 2, ..., L.$$

Thus, we have:

$$\lim_{x \to -w_L^+} e^{I(x)} = \lim_{t \to -\infty} \frac{1}{p_L(1-\xi_L) e^{(-w_L - \frac{M'(t)}{M(t)})t}}$$

The key is the term $(-w_L - \frac{M'(t)}{M(t)})t$.

$$\begin{aligned}
\lim_{t\to-\infty}(-w_L - \frac{M'(t)}{M(t)})t &= \lim_{t\to-\infty}\left(\frac{\sum_{k=1}^{L} p_k w_k((1-\xi_k)e^{-w_k t} - \xi_k e^{w_k t})}{\sum_{k=1}^{L} p_k(\xi_k e^{w_k t} + (1-\xi_k)e^{-w_k t})} - w_L\right)t \\
&= \lim_{t\to-\infty}\left(\frac{\sum_{k=1}^{L} p_k w_k((1-\xi_k)e^{(w_L-w_k)t} - \xi_k e^{(w_L+w_k)t})}{\sum_{k=1}^{L} p_k(\xi_k e^{(w_L+w_k)t} + (1-\xi_k)e^{(w_L-w_k)t})} - w_L\right)t \\
&= \lim_{t\to-\infty}\left(\frac{\sum_{k=1}^{L-1} p_k(w_k - w_L)(1-\xi_k)e^{(w_L-w_k)t} + \sum_{k=1}^{L} p_k(w_k + w_L)\xi_k e^{(w_L+w_k)t}}{\sum_{k=1}^{L} p_k(\xi_k e^{(w_L+w_k)t} + (1-\xi_k)e^{(w_L-w_k)t})}\right)t \\
&= \lim_{t\to-\infty}\frac{\sum_{k=1}^{L-1} p_k(w_k - w_L)(1-\xi_k)te^{(w_L-w_k)t} + \sum_{k=1}^{L} p_k(w_k + w_L)\xi_k te^{(w_L+w_k)t}}{\sum_{k=1}^{L} p_k(\xi_k e^{(w_L+w_k)t} + (1-\xi_k)e^{(w_L-w_k)t})}
\end{aligned}$$

All terms except $(1-\xi_L)e^{(w_L-w_L)t}$ will go to zero as $t \to -\infty$. Thus,

$$\lim_{t\to-\infty}\left(-w_L - \frac{M'(t)}{M(t)}\right)t = 0.$$

Thus, we have:

$$\lim_{x\to-w_L^+} e^{I(x)} = \frac{1}{p_L(1-\xi_L)}.$$

$I(x)$ is continuous at $-w_L$. Similarly, we can prove that $I(x)$ is continuous at w_L. Thus, $I(x)$ is continuous on $[-w_L, w_L]$. *Q.E.D.*

Lemma 6. *Let $x^0 = \sum_{k=1}^{L} p_k w_k(2\xi_k - 1))$. When $x < x^0$, I is decreasing; when $x > x^0$, I is increasing. I achieve its minimum at x^0.*

Proof. As before, we can write $I(x)$ as:

$$I(x) = g(x, t^*(x)).$$

By envelope theorem, we can get:

$$I'(x) = g_x(x, t^*(x)) = t^*(x)$$

Since $t^*(x)$ is strictly increasing in x, we need to find the point where $t^* = 0$. When $t^* = 0$, we have:

$$x^0 = \frac{M'(0)}{M(0)} = \frac{\sum_{k=1}^{L} p_k w_k(\xi_k - (1-\xi_k))}{\sum_{k=1}^{L} p_k w_k(\xi_k + 1 - \xi_k)} = \sum_{k=1}^{L} p_k w_k(2\xi_k - 1)$$

Thus, when $x < x^0$, $I'(x) = t^*(x) < 0$, so I is decreasing; when $x > x^0$, $I'(x) = t^*(x) > 0$, so I is increasing. Since I is continuous, I achieve its minimum at x^0. *Q.E.D.*

Applying Cramér's theorem

Cramer's theorem says that $\frac{S_n}{n}$ satisfies a large deviation principle with rate function $I(x)$.

Now we know that $I(x)$ is decreasing when $x < x^0$, and

$$x^0 = \sum_{k=1}^{L} p_k w_k(2\xi_k - 1) > 0.$$

Thus,

$$\inf_{x<0} I(x) = \inf_{x\leq 0} I(x) = I(0).$$

Under the optimal aggregation rule with weights $w_k = \log\left(\frac{\xi_k}{1-\xi_k}\right)$, we have $t^*(0) = -\frac{1}{2}$. This can be verified by

$$\begin{aligned} M'\left(-\tfrac{1}{2}\right) &= \sum_{k=1}^{L} p_k w_k\left(\xi_k e^{w_k(-\frac{1}{2})} - (1-\xi_k)e^{-w_k(-\frac{1}{2})}\right) \\ &= \sum_{k=1}^{L} p_k w_k\left[\xi_k\left(\frac{\xi_k}{1-\xi_k}\right)^{-\frac{1}{2}} - (1-\xi_k)\left(\frac{1-\xi_k}{\xi_k}\right)^{-\frac{1}{2}}\right] \\ &= \sum_{k=1}^{L} p_k w_k\left(\sqrt{\xi_k(1-\xi_k)} - \sqrt{\xi_k(1-\xi_k)}\right) = 0. \end{aligned}$$

Therefore,

$$\begin{aligned} I(0) = &\ -\tfrac{1}{2}\cdot 0 - \log M\left(-\tfrac{1}{2}\right) \\ = &\ -\log \sum_{k=1}^{L} p_k \Big(\xi_k e^{w_k(-\frac{1}{2})} + (1-\xi_k)e^{-w_k(-\frac{1}{2})}\Big) \\ = &\ -\log \sum_{k=1}^{L} p_k \left[\xi_k \left(\frac{\xi_k}{1-\xi_k}\right)^{-\frac{1}{2}} + (1-\xi_k)\left(\frac{1-\xi_k}{\xi_k}\right)^{-\frac{1}{2}}\right] \\ = &\ -\log\left[\sum_{k=1}^{L} 2p_k\sqrt{\xi_k(1-\xi_k)}\right]. \end{aligned}$$

By Cramér's theorem,

$$\lim_{n\to\infty}\frac{1}{n}\log \Pr(S_n < 0) = \lim_{n\to\infty}\frac{1}{n}\log \Pr(S_n \le 0) = \log\left[\sum_{k=1}^{L} 2p_k\sqrt{\xi_k(1-\xi_k)}\right].$$

Thus, we conclude that

$$\phi = -\log\left[\sum_{k=1}^{L} 2p_k\sqrt{\xi_k(1-\xi_k)}\right].$$

Q.E.D.