Titles as Signals of Expertise: Information Aggregation and Hierarchy Depth

Zijian Qi

June 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, while the organization evaluates their expertise through an imperfect evaluation technology and summarizes the resulting information in a finite set of observable titles. We show that such title systems allow organizations to assign greater decision influence to more reliable contributors and thereby improve decision accuracy. More intensive evaluation and finer rank differentiation both improve information aggregation, but their marginal gains decline as the evaluation-and-title system becomes more refined. The gains from hierarchy are especially front-loaded: a small number of title categories captures much of the attainable improvement, while additional layers provide only limited further gains. As a result, when evaluation and hierarchy are costly, the optimal organization employs limited screening and a relatively shallow hierarchy.

Keywords: organizational design, hierarchy depth, information aggregation, expertise evaluation, weighted majority rule
JEL codes: D23, D83, D71

1 Introduction

Modern organizations increasingly make important decisions in teams whose members may have little direct experience with one another. In many workplaces, relatively short employment relationships give individuals limited time to learn about their coworkers' capabilities, as average job tenure in the U.S. private sector is under four years (U.S. Bureau of Labor Statistics, 2024). Collaboration also often extends across units and areas of specialization, bringing together individuals who may not have worked together before (Bloom and Van Reenen, 2010; Bloom, Sadun, and Van Reenen, 2012b). Furthermore, organizational restructuring can limit familiarity by changing who works with whom within the firm (Rajan and Wulf, 2006; Bloom, Sadun, and Reenen, 2012a). As a result, individuals often participate in important team decisions with limited direct knowledge of their coworkers' relevant expertise.

Decision making in such environments is particularly challenging. When collaborators lack reliable prior information about one another's competence, they cannot easily assess whose judgments are more informative. Yet many organizational decisions are complex and significant, 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, and vice president, provide one such device because they are publicly observable, organizationally endorsed, and often based on prior evaluations of performance and ability. Although titles also serve other purposes, including reporting relationships and career incentives, they can also act as coarse signals of expertise. In practice, when opinions conflict or evidence is noisy, higher ranked individuals often have greater influence over the collective decision. This creates an implicit aggregation rule in which rank affects the weight placed on different assessments. A natural design question follows: how finely should organizations differentiate titles, and how much information should they invest in doing so, 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. This paper highlights a different role of hierarchy: it can serve as an informational device that translates imperfect assessments of expertise into weights in collective decision-making.

If hierarchy plays this informational role, a natural question is how such a hierarchy should be designed. When titles are based on imperfect evaluations of expertise, the organization must decide how much information to collect about individual expertise and how finely to represent that information through formal ranks. More intensive evaluation can make titles more informative, while a more detailed title structure can preserve finer distinctions across individuals. At the same time, both choices may increase organizational complexity. The relevant tradeoff is therefore between the informational value of more precise titles and the cost of maintaining a more detailed evaluation-and-title system.

This paper develops a model of organizational decision-making under uncertainty in which workers receive signals whose reliability depends on their expertise, while expertise itself is not directly observable. The organization can evaluate workers through an imperfect technology and map evaluation outcomes into a finite set of observable titles. Titles therefore summarize the organization's information about the expected expertise of different workers. The organization then combines workers' signals into a collective decision, allowing the aggregation rule to depend on the titles workers hold. The analysis abstracts from other roles of titles, such as incentives, reporting relationships, delegation, and communication, in order to isolate their informational role in collective decision-making.

Building on standard results in the information aggregation literature (Nitzan and Paroush, 1982; Shapley and Grofman, 1984), the analysis characterizes the optimal aggregation rule for a fixed evaluation-and-title system as a title-based weighted majority rule, in which titles determine the relative influence of workers' signals. The paper then studies how decision quality changes with two organizational design margins: evaluation intensity, which determines how precisely the organization evaluates individual expertise, and hierarchy depth, which determines how finely evaluation outcomes are translated into rank categories. We show that both margins improve decision quality by helping the organization better distinguish differences in expertise, but their informational benefits are diminishing. These diminishing returns are especially sharp for hierarchy depth: a relatively small number of titles captures most of the attainable gains from information aggregation, while further differentiation adds complexity with little additional benefit. Consequently, when evaluation and hierarchy are costly, the optimal design combines a limited amount of evaluation with a relatively shallow hierarchy.

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.

The study of hierarchy has deep roots in organizational economics. Foundational 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). This literature highlights hierarchy as a structure for processing information and coordinating decisions across levels.

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). In this literature, hierarchy primarily organizes authority, incentives, and control.

Another related strand of the literature studies hierarchy as a mechanism for allocating problem-solving tasks and specialized knowledge. Garicano (2000) develops a canonical model of knowledge hierarchies, in which lower-level agents handle routine problems, while rare or complex problems are referred upward to agents with more specialized knowledge. 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). This literature views hierarchy as a way to route problems, economize on specialized knowledge, and organize expertise across levels.

Taken together, these perspectives emphasize hierarchy as a structure for coping with bounded rationality and communication constraints, allocating authority, providing incentives, monitoring agents, and organizing specialized knowledge. Our analysis abstracts from these channels and focuses instead on a different informational role of hierarchy: formal titles serve as coarse summaries of noisy evaluations of expertise and determine how dispersed individual assessments are aggregated into a collective decision. This perspective connects the paper to the literature on collective decision-making and heterogeneous information aggregation.

This literature shows that unequal influence can improve collective decision-making when individuals differ in expertise or signal quality. 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. Related results appear in models of voting and information aggregation with heterogeneous expertise (Condorcet, 1785; Austen-Smith and Banks, 1996). These results provide the informational foundation for our analysis.

Building on these insights, we study how formal titles can serve as the organizational basis for differential influence in collective decision-making. The organization chooses both how much information to generate about individual expertise and how finely to translate evaluation outcomes into hierarchical categories. Rather than taking the categories that determine decision weights as given, the model derives them from this evaluation-and-title system. In contrast to much of the hierarchy literature, hierarchy does not govern reporting lines, delegation, incentives, monitoring, or problem routing; communication is costless conditional on titles, and hierarchy affects decisions only through the information about expertise embedded in the title structure. By endogenizing evaluation intensity and hierarchy depth, we characterize when finer hierarchical distinctions improve decision accuracy and when their marginal informational value is outweighed by organizational costs.

The remainder of the paper is organized as follows. Section 2 presents the model and formalizes the evaluation-and-title system. Section 3 analyzes decision-making under a fixed evaluation-and-title system, derives the performance measure, and introduces benchmark cases. Section 4 studies how evaluation intensity and hierarchy depth affect performance and optimal organizational design. 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 $i = 1, 2, \dots, n$, 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 informational 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

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

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

The closed-form performance rate derived in Section 3 allows the organizational design problem to be evaluated numerically. For each pair (m, L), the assignment rule is chosen to maximize the performance rate, yielding $r^*(m, L)$. All performance measures and objective values reported below are evaluated under this optimized assignment rule. The main analysis focuses on how evaluation intensity m and hierarchy depth L affect performance and optimal organizational design. Their informational effects are first examined in the absence of organizational costs. Costs are then introduced separately for evaluation and hierarchy to study how they shape the optimal choices of m^* and L^*. The structure of $r^*(m, L)$ is also described to illustrate how evaluation outcomes are grouped into title categories.

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 allows us to isolate the informational roles 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 while holding evaluation intensity fixed at $m = 300$. Increasing the number of titles generates a sharp front-loaded improvement in performance: even a small number of additional categories substantially improves how decision authority is allocated. However, this effect saturates quickly. Beyond a modest hierarchy depth, the performance curve becomes essentially flat, so additional increases in L yield little improvement. This pattern indicates that hierarchy has large initial effects but rapidly diminishing returns: most of its informational value is realized early through coarse differentiation, and additional layers provide little incremental refinement for aggregation.

Figure 2 examines the role of evaluation intensity m when the hierarchy is fully refined,

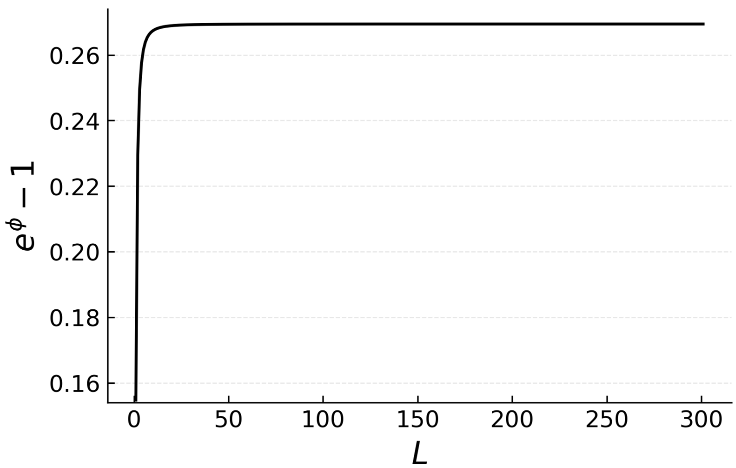


Figure 1: Performance rate as a function of hierarchy depth L for fixed evaluation intensity $m = 300$.

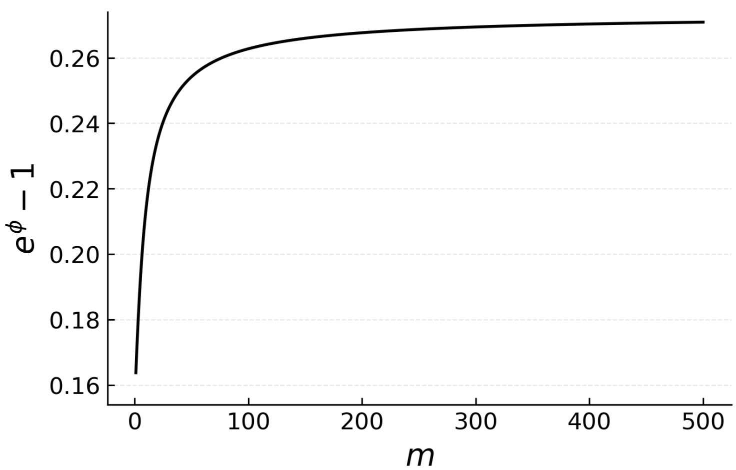
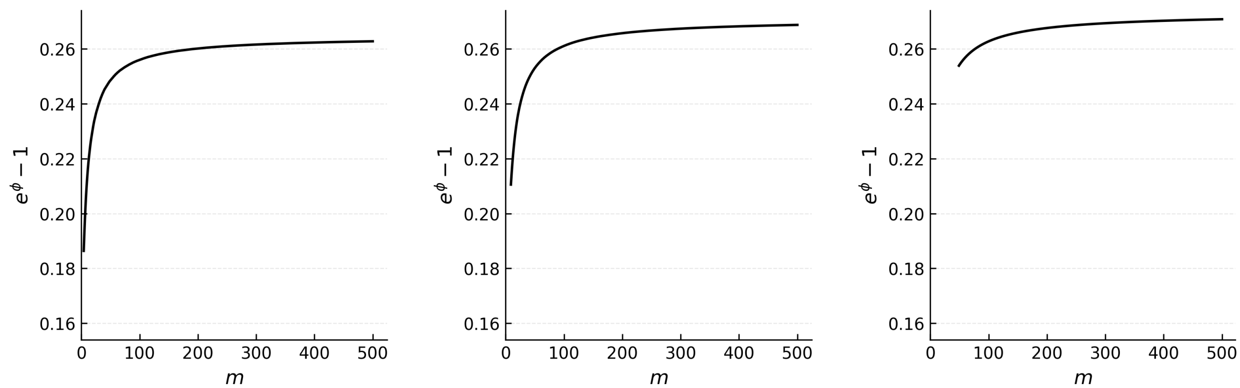


Figure 2: Performance rate as evaluation intensity m increases when hierarchy depth grows according to $L = m + 1$.

so that $L = m + 1$ and each test outcome corresponds to a distinct title. In this case, the evaluation-and-title system utilizes all available information generated by the tests. As m increases, the performance rate rises steadily and approaches the first-best benchmark derived in Section 3.3. The improvement is initially pronounced but tapers off more gradually than in the case of hierarchy depth, indicating that evaluation provides a smoother and more persistent source of informational gains.

Figure 3 separates the roles of evaluation intensity and hierarchy depth by holding L fixed and varying m. For all values of L, the performance rate is increasing and concave in m.



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

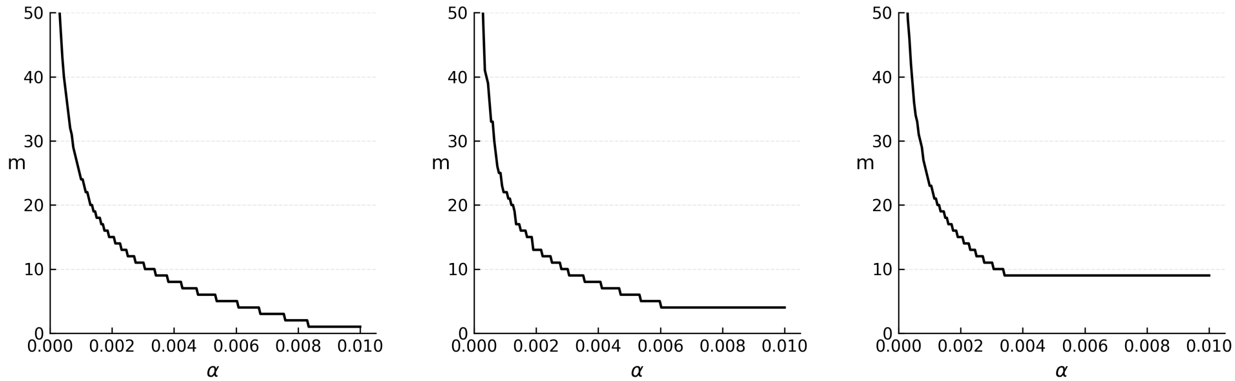
Figure 3: Performance rate as a function of evaluation intensity m for fixed hierarchy depths $L \in \{5, 10, 50\}$.

More informative evaluations improve the allocation of decision authority across titles, and while the marginal benefit declines, the adjustment is comparatively gradual. Differences across hierarchy depths are most pronounced at low levels of evaluation, but even as m becomes large, incremental improvements continue to accumulate, in contrast to the rapid flattening observed along the hierarchy dimension.

Taken together, these results show that evaluation intensity and hierarchy depth both improve decision performance, but their effects operate at different speeds. Hierarchy produces sharp, front-loaded gains that plateau quickly. Evaluation also exhibits diminishing returns, but its gains accumulate more gradually over a broader range of m. Thus, a small number of title categories captures much of the benefit from hierarchy, while additional evaluation can remain useful even after the hierarchy dimension has largely flattened. These patterns play a central role in the design problem studied below, where organizational costs limit the extent to which evaluation and hierarchy can be expanded.

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



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

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

tests. The cost function is given by

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

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 considers the case in which hierarchy expands with evaluation according to $L = m + 1$. The middle and right panels fix hierarchy depth at $L = 5$ and $L = 10$, respectively.

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 progressively outweighed by its direct cost. This leads to step wise reductions in m^* due to the discrete nature of the decision variable.

The comparison across hierarchy structures should be interpreted with two basic features in mind. When L is fixed, feasibility requires $m \geq L-1$, so the minimum value of m differs across panels. In particular, when $L=5$, the lower bound is $m=4$, and when $L=10$, the lower bound is $m=9$. By contrast, under $L=m+1$, the minimum feasible value is $m=0$. This accounts for the different lower endpoints observed in the figure when α becomes large.

Away from these lower bounds, the flexible hierarchy case typically supports a higher level of evaluation. When $L=m+1$, each additional test outcome corresponds to a distinct title, so the information generated by evaluation is fully utilized. When L is fixed, additional test outcomes are eventually pooled into existing title categories, which reduces the marginal value of further evaluation.

Overall, the main pattern is that the optimal evaluation intensity declines gradually as the marginal cost α increases. The organization scales back evaluation in a smooth manner over most of the cost range. Only when the feasibility constraint binds does the adjustment stop. This reflects that evaluation generates incremental improvements in information, so its value diminishes progressively rather than abruptly.

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

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

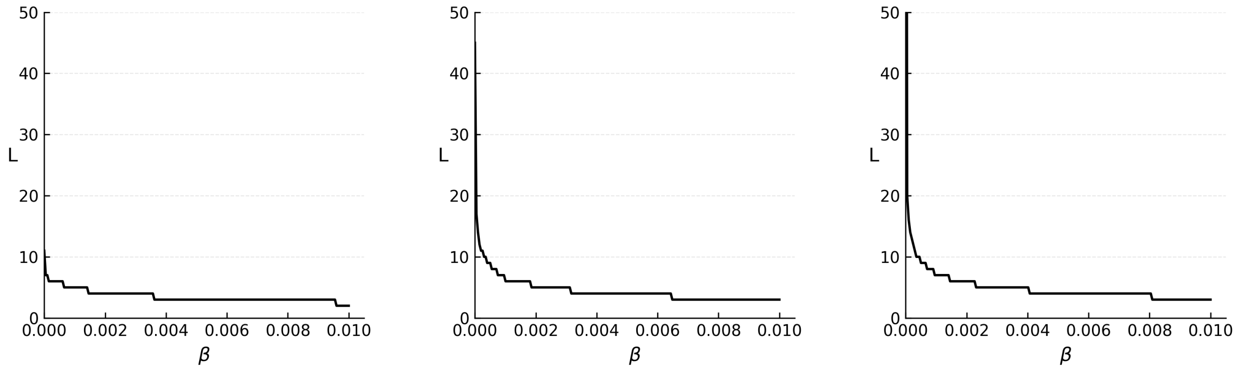


Figure 5: Optimal hierarchy depth L^* as a function of the marginal hierarchy cost β, for selected values of evaluation intensity m.

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. The left, middle, and right panels correspond to $m = 10$, $m = 50$, and $m = 300$, respectively. Across all cases, the optimal hierarchy depth declines as the marginal cost of hierarchy increases. When hierarchy is inexpensive, the organization adopts a deeper title structure in order to exploit finer distinctions in expected expertise. As hierarchy becomes more costly, the organization optimally compresses title categories and relies on a shallower hierarchy.

The figures also illustrate how evaluation intensity affects the usefulness of hierarchy. When evaluation is limited, as in the case $m = 10$, evaluation outcomes provide only coarse information about expertise. As a result, additional title categories have limited value because workers cannot be meaningfully separated into many distinct expertise groups. As evaluation becomes more informative, deeper hierarchies become more valuable because additional title categories can be used to encode finer distinctions in expected expertise. Consequently, the optimal hierarchy depth is generally higher for larger values of m.

At the same time, the adjustment in hierarchy depth is concentrated over a relatively narrow range of hierarchy costs. Most of the informational gains from hierarchy are realized through a relatively small number of title categories, so increases in β quickly reduce the value of maintaining additional layers. This pattern is consistent with the earlier result that performance gains from hierarchy are strongly front-loaded and flatten rapidly as L increases.

Overall, these results show that hierarchy is most valuable when evaluation generates sufficiently informative distinctions across workers. However, even in high-information environments, the optimal hierarchy remains relatively shallow once hierarchy costs become nontrivial.

4.4 Structure of the Optimal Title Assignment Rule

We next examine the structure of the optimal assignment rule $r^*(m, L)$. This subsection is mainly descriptive. Its purpose is to illustrate how evaluation outcomes are translated into title categories once m and L are fixed.

We focus on relatively coarse hierarchies, which are both empirically relevant and visually easier to interpret. Figure 6 reports the optimal assignment rules for $L = 5$, $L = 10$, and $L = 20$, holding evaluation intensity fixed at $m = 300$. The top row plots the mapping from evaluation outcomes j to titles. The bottom row reports the induced probability distribution across titles.

The figure shows that the optimal assignment rule partitions the evaluation distribution unevenly across titles. Lower titles pool broad ranges of evaluation outcomes, while higher titles correspond to progressively narrower intervals of test results. As hierarchy depth increases, the partition becomes finer, but the additional cuts occur primarily in the upper tail of the evaluation distribution.

The bottom row illustrates the corresponding distribution of workers across titles. In relatively coarse hierarchies, higher titles generally contain smaller probability mass. This occurs because extreme evaluation outcomes are relatively rare, so top titles are assigned only to a limited fraction of workers. At the same time, lower titles pool together a much larger set of evaluation outcomes and therefore contain a larger share of the population.

However, this relationship becomes less mechanical as hierarchy depth increases. When

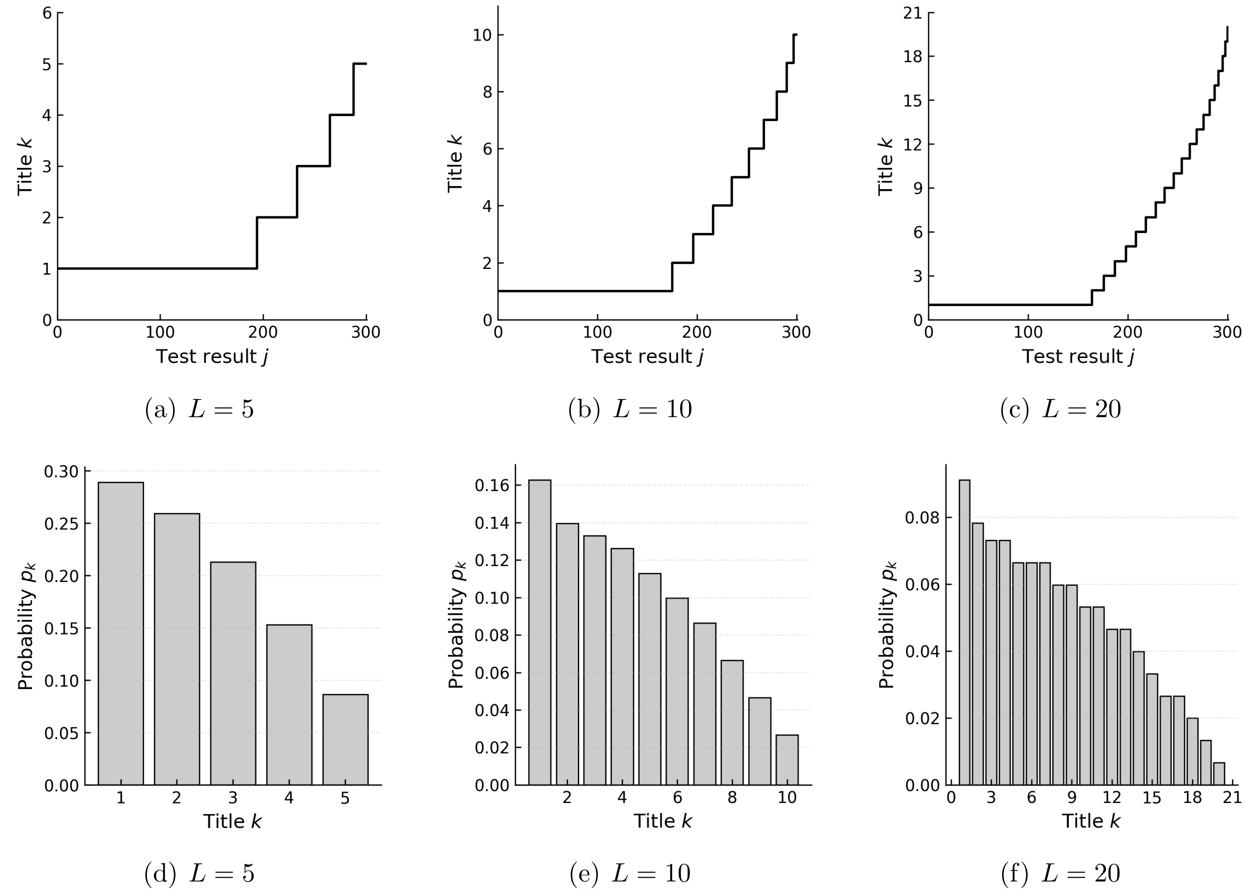


Figure 6: Top row: optimal assignment rules mapping evaluation outcomes to titles for $m = 300$. Bottom row: corresponding distribution of workers across titles. Columns vary hierarchy depth $L \in \{5, 10, 20\}$.

L is large, each title may correspond to only a narrow range of test outcomes, and the probability associated with a title depends more directly on the underlying distribution of expertise and the induced distribution of evaluation results. Thus, the probability mass across titles should not be interpreted as necessarily decreasing monotonically with title rank in general.

More broadly, the assignment rule allocates informational resolution unevenly across the expertise distribution. Additional title categories are used primarily to refine distinctions among workers with relatively high evaluation outcomes, while lower evaluation outcomes remain pooled into broader categories. This reflects diminishing returns to differentiation at lower levels of expected expertise, where finer distinctions have relatively limited value for information aggregation.

Overall, the assignment rule plays a supporting role in the design problem. For each pair (m, L), it determines how evaluation outcomes are translated into title categories. The main comparative statics are driven by evaluation intensity and hierarchy depth, while the assignment rule clarifies how the available title categories are allocated across the evaluation distribution.

5 Discussion

The numerical results reveal a common pattern across evaluation intensity and hierarchy depth. Both dimensions improve decision performance, and both exhibit diminishing marginal returns. The difference lies in how quickly these marginal gains decline. Performance gains from hierarchy are concentrated at low levels of L, while gains from evaluation accumulate more gradually as m increases. Thus, the results show a similar qualitative effect along the two dimensions, but a sharper flattening along the hierarchy dimension.

Evaluation intensity and hierarchy depth capture two different informational choices in the organization. Evaluation intensity determines the precision of information about individual expertise. A larger m makes evaluation outcomes more informative about agents' underlying reliability and allows the organization to form sharper posterior beliefs. Hierarchy depth determines how coarsely or finely this information is summarized at the organizational level. A larger L allows the organization to preserve more distinctions in posterior expertise when assigning titles and aggregation weights, while a smaller L compresses more heterogeneous evaluation outcomes into the same title category. Thus, evaluation intensity governs the quality of the information generated about agents, whereas hierarchy depth governs the resolution of the organizational summary through which that information enters collective decision-making.

The diminishing returns along both dimensions reflect how the evaluation-and-title system translates information into decision authority. Evaluation and hierarchy improve performance by helping the organization distinguish agents with different expected reliability. At low levels of precision or differentiation, these distinctions can be quite coarse, so additional information or additional title categories can substantially change how decision authority is allocated. As the system becomes more informative, however, the remaining distinctions are smaller. Further increases in m or L may still improve the induced grouping of agents and the associated weights, but they have less effect on the aggregate decision rule. This is why both dimensions improve performance, but with declining marginal gains.

The sharper flattening along the hierarchy dimension should be read as a quantitative feature of the numerical analysis. In these computations, a relatively small number of title categories is enough to capture much of the decision-relevant variation in expected expertise. Additional categories expand the organization's ability to differentiate agents, but they do not substantially change the title-level mean expertise, the induced aggregation weights, or the resulting performance rate. Thus, the front-loaded pattern for hierarchy suggests that coarse title differentiation can be highly effective in organizing information for collective decision-making.

The organizational implication is that hierarchy need not be very deep to be useful for information aggregation. A limited title structure can already allow the organization to distinguish broad differences in expertise and assign decision weights accordingly. Once these broad distinctions are in place, finer rank differentiation may have only a small effect on collective decision quality. This does not rule out deeper hierarchies for other organizational purposes, such as incentives, monitoring, communication, or career progression. It does suggest, however, that limited differentiation can be effective when the primary role of hierarchy is to organize heterogeneous expertise for collective decision-making.

6 Conclusion

This paper studies how hierarchy affects collective decision-making when individual expertise is heterogeneous and imperfectly observed. The organization evaluates agents through noisy signals, assigns them to title categories, and uses these titles to determine the weights placed on their private assessments. In this setting, hierarchy serves as an information structure: it summarizes heterogeneous expertise into a finite set of authority-relevant categories.

The analysis derives a closed-form performance rate for the optimal weighted decision rule and uses it to study the organization's design problem. Evaluation intensity and hierarchy depth both improve decision performance, but both exhibit diminishing marginal returns. The numerical results show that hierarchy can generate substantial gains with only a small number of title categories, while further differentiation adds progressively less to performance. When evaluation and hierarchy are costly, these diminishing returns lead to finite optimal choices of evaluation intensity and hierarchy depth.

The results provide an information-aggregation perspective on why limited hierarchical differentiation can be effective. Titles allow the organization to assign greater decision weight to agents with higher expected expertise, but a highly refined title structure is not always necessary for this purpose. Once broad differences in expertise are reflected in the aggregation weights, additional ranks may have limited value for decision quality. This mechanism complements other roles of hierarchy, such as incentives, monitoring, communication, and career progression.

The framework also suggests several natural directions for future research. One direction is to study dynamic title systems in which evaluations arrive over time and promotions are persistent, so that titles reflect both current information and past evaluation histories. Another is to allow more general evaluation technologies, such as project outcomes, peer assessments, or task-specific signals, rather than independent Bernoulli tests. A third direction is to consider richer cost structures, including nonlinear costs or joint costs of evaluation and hierarchy, to study how organizations choose evaluation intensity and hierarchy depth when the two design dimensions are cost-linked.

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 $k \in \{1, \ldots, L\}$, 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 $\xi_1 < \xi_2 < ... < \xi_L$. 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

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

Let $S_n = X_1 + X_2 + ... + X_n$, 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

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

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

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

Similarly,

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

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,

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

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,

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

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.