

# On Profit Sharing and Hierarchies in Organizations

Swaprava Nath	Indian Institute of Science, Bangalore
Balakrishnan Narayanaswamy	IBM India Research Lab
Kundan Kandhway	Indian Institute of Science, Bangalore
Bhushan Kotnis	Indian Institute of Science, Bangalore
David C. Parkes	Harvard School of Engineering and Applied Sciences

## Abstract

We study the effect of hierarchies on the performance of an organization that exhibits both profit sharing and information sharing. We adopt a server-queue model of effort and intra-organizational competition and quantify the performance of an organization in terms of the overall efficiency and risk in meeting target production levels. On the one hand, we show that profit sharing leads to free riding at the Nash equilibrium and reduces overall effort. Introducing a form of information asymmetry, we then quantitatively establish the trade-off between free riding and the positive effects of information sharing in hierarchies. We formulate an optimal hierarchy design problem that captures both effects, and provide a simple algorithm that gives near optimal designs. We provide an analysis of the productive value of optimized hierarchies, considering their ability to attain target production levels with low risk.

## 1 Introduction

The organization of economic activity as a means for the efficient allocation of scarce resources is a cornerstone of economic theory. Our economy is dominated by large firms, which face the problem of organizing the activities of their employees (or human resources) to maximize profits, forming a miniature economy of their own. Large firms usually have a hierarchical organization and the economic significance of this structure has been the focus of much research [8, 9, 7, e.g.]. In this paper, we focus on the interaction between hierarchical structure and other common aspects of organizational interactions, namely profit sharing [5], information exchange [2], and influence [12]. We take a multi-agent view of an organization, modeling it as a collection of agents with disparate information and non-aligned interests. In this setting, understanding incentives and the effect of asymmetric information become central [7].

We adopt a game theoretic perspective to construct quantitative models of organizations, and quantify the effects of asymmetric information and opportunistic behavior. We propose a queuing-theoretic model of task arrival and processing, and a natural model of profit sharing that provides a direct reward for completing a task and an indirect reward if subordinates complete a task. We then model employee competition for tasks based on the cost of the effort involved and the possible rewards. In addition to considerations that follow from boundedly rational models of managers [10, 13] with limited information about opportunities, we show how common organizational features, in particular profit sharing and intra-organizational competition, can reduce overall efficiency due to a free-riding phenomenon.

In an environment with profit sharing and without the benefits of information sharing that can accrue within an organization, our first result (Theorem 1) quantifies the effect of free riding on reducing the efficiency of an organization. In particular, the total output of an organization decreases when compared to a flat organization. We then introduce a model of information asymmetry and consider the positive effects of information sharing within an organization, where managers are able to improve the estimates subordinates have for the rewards for task completion. That is, managers share a form of ‘business intelligence’ about the value of tasks to the organization and their potential rewards.

Based on this queuing and game theoretic model, we study the trade-offs between free riding and the positive effects of information sharing. In particular we seek to answer quantitative questions of the form ‘how well should management communicate to overcome the free riding effects’? We show that information sharing can provide benefit even in the presence of free riding, and quantify the magnitude of both effects (Theorem 2). The simulation results serve to illustrate that the performance of the organization improves when the hierarchy is designed considering free riding, as opposed to constructions that do not model these effects.

**Paper outline:** We first introduce our general model, which includes rewards, information asymmetry and information sharing. In Section 3, we address the perfect information scenario where agents observe the true reward they obtain from task completion, and describe the free riding that occurs when managers share in the profits of employees. In Section 4, we study the effect of information asymmetry, and quantify how information sharing and propagation improves the effectiveness of agent effort. In Section 5, we combine the two, and discuss how the optimization of information propagation in hierarchies with free riding results in interesting network structures with improved efficiency.

## 2 The Model

We first develop a model of task (or work) arrival and processing in an organization using a server-queue model. The employees and managers of the organization are modeled as strategic autonomous agents who choose actions to maximize their own payoffs, given the organizational structure and with the possibility of indirect rewards for the tasks completed by others. The agents may have asymmetric information and can communicate when they are connected.

### 2.1 The Task Arrival and Processing Model

We model an organization as a task processing unit where the tasks arrive at a Poisson rate  $\lambda$ , and wait until served. For example, the tasks could be projects that need to be executed, or even small modules that need to be developed or completed as part of a larger contract. Other situations where such a model would be applicable arise for example in sales organizations for which agents get commissions [4].

Let the set of employees be denoted by  $N = \{1, 2, \dots, n\}$ , and connected over a hierarchy, represented by a directed tree  $T$ . We assume that if an individual performs a task she gets a direct reward and the nodes along the path from the individual to the root receive indirect incentives. These indirect rewards model bonuses or the ‘credit’ a manager receives when a task is completed by her employee.

Each node  $i$  attempts to capture a task from the work queue according to a Poisson process with rate  $\lambda_i$ . The rate  $\lambda_i$  corresponds to the effort that the agent expends. The cost of maintaining an attempt rate  $\lambda_i$  is  $C\lambda_i$ , where  $C > 0$  is a known constant. Such a linear increase model of cost has been adopted in literature on public goods network [3]. If the total reward of a task is  $R$ , corresponding

to the total bonus or credit that the organization allocates to a task, we assume that the node that executes it gets a direct reward of  $\gamma R$  ( $\gamma \in (0, 1)$ ), and the rest shared as indirect incentives among its predecessors in the organizational tree. The true reward of a task,  $R$ , is stochastic and distributed as  $F_R$ .

In our initial analysis, we assume that every agent knows the realized reward of a task and in the subsequent analysis we assume every agent has a noisy estimate of the reward of a task, corresponding to information asymmetry between agents in an organization, where some agents can better estimate the true value of a task and thus expend their effort more effectively. To keep the model simple, every agent is assumed to have the same intrinsic ability to complete a task, all tasks carry the same reward and that the time to complete a task is small compared to the inter-attempt time of an agent. Some of these assumptions can be relaxed with considerable increase in notational complexity though our basic model is sufficient to capture the effects of profit sharing and information asymmetry.

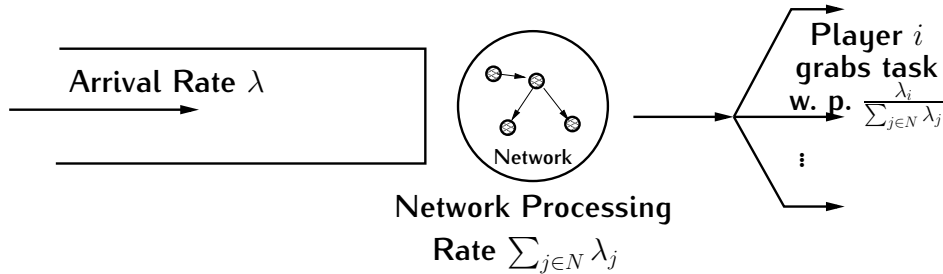


Figure 1: Server queue model for the task arrival to the network.

We model the task arrival and departure in a server-queue model, as shown in Figure 1, where the arrival rate is  $\lambda (> 0)$  and the consolidated service rate of the entire network is  $\sum_{j \in N} \lambda_j$ , using the simple fact that the superposition of Poisson processes is Poisson with the rate being the sum of the rates [1]. When agent  $i$  tries to capture a job from the task queue, the probability that she succeeds is given by  $\lambda_i / \sum_{j \in N} \lambda_j$ . This can be shown using the fact that the inter-attempt times are exponentially distributed for a Poisson process, and exponential random variables exhibit memoryless property [14]. This basic model can be extended using other queuing theoretic models of job arrival and task completion, though that is beyond the scope of this paper.

## 2.2 Agent Utility and Reward Sharing

Organizations have profit sharing schemes where a manager is rewarded for a successful execution of a project by his team. Hence all agents connected in the hierarchy have two parts to their net payoffs. If node  $j$  executes a task, she receives a ‘direct’ reward, and each node  $i$  on the directed path from the root to  $j$  receives an ‘indirect’ reward. Since efforts are costly ( $C\lambda_i$ ), each agent decides on an effort level to maximize her net payoff, i.e., (direct + indirect) reward - cost. The choice of attempt rate  $\lambda_i \in S_i = [0, \infty)$  is a strategy for the agent, and this induces a game between the nodes. If there is a large indirect reward a node can (and will) reduce efforts to reduce costs. In the following, we derive an quantitative expression for the expected utility for an agent  $i$ , when there are explicit direct and indirect rewards.

The indirect incentives are shared with other players as specified by a *reward sharing matrix*  $\Delta = [\delta_{ij}]$  where  $\delta_{ij}$  is the fraction of  $j$ ’s reward received by  $i$  when  $j$  executes a task. We assume,  $\delta_{ii} = \gamma > 0$ ,  $\forall i \in N$ . Moreover, since the network is a directed tree, if  $j$  does not appear in subtree of  $i$  then  $\delta_{ij} = 0$ . Our analysis extends to more general organizational structures, such as ‘matrix’ type organizations where employees may report to multiple managers. We call the matrix  $\Delta$  *monotone non-increasing* if  $\delta_{ij} \geq \delta_{ij'}$ , whenever the hop-distance  $\text{dist}_T(i, j) \leq \text{dist}_T(i, j')$ , and *anonymous* if the

$\delta_{ij}$  depends only on the distance of  $i$  and  $j$  on  $T$ ,  $\text{dist}_T(i, j)$ , and not on the identities of the nodes. To ensure that the total reward is bounded above by  $R$ , we assume  $\sum_{k \in N} \delta_{kj} \leq 1$ ,  $\forall j \in N$ .

An example of a monotone non-increasing and anonymous reward sharing scheme is a *geometric reward sharing* scheme. Node  $i$ , a predecessor of  $j$  in the tree  $T$ , gets a fraction of reward  $\delta^{\text{dist}_T(i, j)}$ ,  $\delta \in (0, 1)$ , when  $j$  grabs the task. Node  $j$  gets  $\gamma$  fraction for her own effort, and  $\delta + \gamma \leq 1$  is required to balance the budget. For concreteness, we will use this reward sharing scheme for the simulations.

Let us denote with  $T_i$  the subtree rooted at  $i$ . Combining the direct and indirect rewards and costs, the expected utility of agent  $i \in N$  is given by,

$$u_i(\lambda_i, \lambda_{-i}) := \lambda \gamma R \frac{\lambda_i}{\sum_{k \in N} \lambda_k} + \lambda R \sum_{j \in T_i \setminus \{i\}} \delta_{ij} \frac{\lambda_j}{\sum_{k \in N} \lambda_k} - C \lambda_i, \quad (1)$$

where  $\lambda_{-i} = (\lambda_1, \dots, \lambda_{i-1}, \lambda_{i+1}, \dots, \lambda_n)$ . In the above equation, the first term on the RHS denotes the expected utility of agent  $i$  due to her own effort  $\lambda_i$ , the second term represents the indirect utility coming from the efforts  $\lambda_j$  of all the nodes  $j \in T_i \setminus \{i\}$ , and the third term captures the cost an agent pays to maintain the effort level. The function  $u_i$  can be shown to be concave in  $\lambda_i$ , we skip the proof due to space constraints.

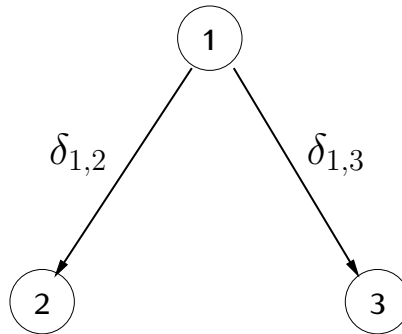


Figure 2: Node 1 is rewarded indirectly by 2 and 3.

Depending on the efforts exerted by the agents, the system could operate in one of the following two zones:

**Zone 1:** When  $\sum_{j \in N} \lambda_j > \lambda$ , the task queuing process is a positive recurrent Markov chain and all tasks will be served, and hence this is a desirable zone for the organization to operate. Let us denote this zone by  $Z_1 = \{\boldsymbol{\lambda} : \sum_{j \in N} \lambda_j > \lambda\}$ .

**Zone 2:** When  $\sum_{j \in N} \lambda_j \leq \lambda$ , the queuing process is a null or transient Markov chain and with high probability (probability approaching unity) the queue will be non-empty and growing at a steady state [14]. This is an undesirable state for an organization since certain incoming projects will never be served. Let us denote this zone by  $Z_2 = \{\boldsymbol{\lambda} : \sum_{j \in N} \lambda_j \leq \lambda\}$ .

Simply stated, the desirable state for the organization as a whole is in Zone  $Z_1$ , with all incoming customers served. To illustrate the utilities and the reward sharing among the agents, let us consider a small hierarchy  $T^{(3)}$  with nodes 1, 2, and 3, connected as shown in Figure 2. Node 2 and 3 gets only direct reward due to their own effort, but 1 gets both direct and indirect rewards.

## 2.3 Information Asymmetry

In the full version of our model, we model agents as having only noisy estimates of the reward associated with a task. Agent  $i$  observes the reward as  $R_i = R + \eta_i$ , which is a noisy version of the actual reward. The noise is independent across agents, and is additive to true reward  $R$ , and is distributed according to a Normal distribution:  $\eta_i \sim \mathcal{N}(0, \sigma_i^2)$ . Agents vary in their intrinsic variance  $\sigma_i^2$ , reflecting their ability to track the reward accurately. The parameter  $\sigma_i$  is assumed private to agents. For this reason, agent  $i$  knows that agent  $j$  observes a realization of  $R_j = R + \eta_j$ , but does not know  $\sigma_j$  or the value of  $R_j$ . In this paper, we assume that agent  $i$  picks its action after observation  $R_i$  is realized; i.e., the employees put their efforts after a project arrives.

We allow information to be shared between agents. We model information propagation as the change in the variance of the individuals, corresponding to a change in the uncertainty or business insight that an employee has through communication with her manager. We assume the variances evolve over time, and the dynamics depend on how agents are connected in the hierarchy and how well they communicate. With a little abuse of notation, we use the lowercase  $t$  as the index for time, and the uppercase  $T$  to denote the hierarchy. Let  $\sigma_j^2(t, T)$  be the instantaneous variance of node  $j$  on tree  $T$ , with  $\sigma_j^2(0, T) = \sigma_j^2$ , the intrinsic variances, for all  $T$ . The rate of change of agent  $j$ 's variance is proportional to the instantaneous difference of variance between her parent  $p(j)$  and herself.<sup>1</sup> Formally,

$$\frac{d\sigma_j^2(t, T)}{dt} = \mu(C_{p(j)}(T))(\sigma_{p(j)}^2(t, T) - \sigma_j^2(t, T)). \quad (2)$$

The set of children of node  $i$  on tree  $T$  is denoted by  $C_i(T) := \{k \in N : (i, k) \in T\}$ . The communication function  $\mu$  maps the set of children of the parent  $p(j)$ , given by  $C_{p(j)}(T)$ , to  $\mathbb{R}$ , and is a monotone decreasing function. Where the tree  $T$  is understood, we leave this silent to simplify notation.

To get a feel for this information propagation model, let us consider a case where the parent  $p(j)$  has a smaller instantaneous variance than  $j$ , i.e., the manager knows the business potential more accurately. Under this model, agent  $j$  also understands the business better due to her connection with a better informed manager. Similarly, if the instantaneous variance of the manager was larger, then the employees managed by him would be worse informed about the business, which would reflect in their variance, even though their intrinsic variances were low.

The monotone decreasing function  $\mu$  models that if a manager manages a large team, his effort per employee would be small. For the simulations, we adopt a sigmoid function:

$$\mu(C_{p(j)}) = a \cdot \frac{1}{1 + e^{s(|C_{p(j)}| - b)}}, \quad (3)$$

where the parameters  $a \in [0, 1]$ ,  $s \in \mathbb{R}_{>0}$ , and  $b \in \mathbb{N}$  denote the amplitude, steepness, and breadth of communication respectively. A typical plot of  $\mu$  with  $a = 1, s = 3, b = 4$  is given in Figure 3. Parameter  $a$  controls the amplitude of the function,  $s$  captures how steeply the curve falls, and  $b$  captures how many children can the parent support without significant fall in  $\mu$ .

## 2.4 Agent Model

The agents are assumed to be strategic, and they choose their effort  $\lambda_i$ 's to maximize their payoff, given by Equation (1). If the reward is perfectly known to the agents, they maximize this utility function by choosing appropriate effort in a pure-strategy Nash equilibrium.

---

<sup>1</sup>This is similar to the information ‘osmosis’ model [6].

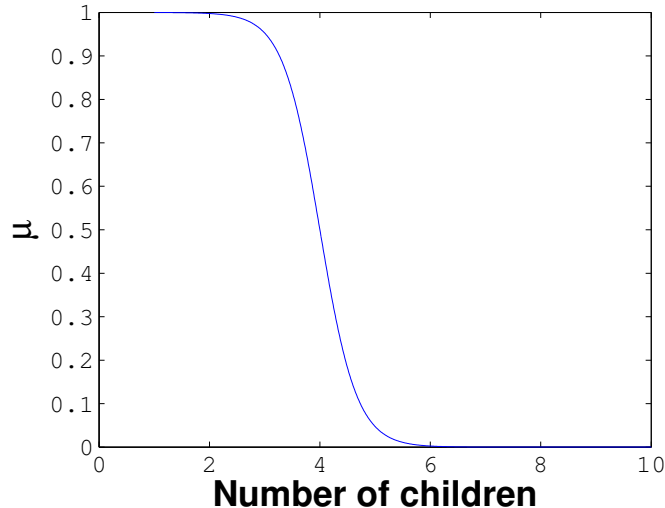


Figure 3: Function  $\mu$ , with  $a = 1, s = 3, b = 4$ .

When  $R$  is observed with noise, according to the information asymmetry model (§2.3), we assume that each agent  $i$  estimates the reward perceived by the other agents. That is, each agent  $i$  tries to estimate  $R_j$  given their privately observed  $R_i, \forall j \neq i$ . The two random variables are related as,  $R_i = R_j + \eta_i - \eta_j$ . Since both  $\eta_i$  and  $\eta_j$  are zero mean Gaussians with variances  $\sigma_i^2$  and  $\sigma_j^2$  respectively,  $R_i$  given  $R_j$  is distributed as  $\mathcal{N}(R_j, \sigma_i^2 + \sigma_j^2)$ . Hence the *maximum likelihood* (MLE) estimate of  $R_j$  after observing  $R_i$  is given by  $R_i$  itself.

In this setting, we adopt a simple model of the agent  $i$ 's behavior, in which the agent acts in a certainty-equivalent way and adopts this estimate  $R_i$  in place of  $R$ , including for the purpose of modeling the reward perceived by other agents. In particular, agent  $i$  adopts as its strategy the effort level that it would adopt in the pure strategy Nash equilibrium in a subjective model of the game, in which it assumes that every other agent adopts as the reward for the current task the MLE estimate of agent  $i$ .<sup>2</sup> In this paper, we refer to this behavioral model as the *MLE Best Response (MBR)* model.

## 2.5 Performance Metric

To ensure that all incoming tasks are served in the long run, the sum of the best response efforts should be at least the incoming rate of tasks. Let us denote the best response effort of agent  $i$  under the MBR model by  $\lambda_i^{MBR}$ . The sum effort is a random variable under the information asymmetry model of the agents. We define the *risk* of the organization as the probability that the sum effort does not meet the arrival rate:

$$\text{Risk} := \mathbb{P} \left( \sum_{i \in N} \lambda_i^{MBR} \leq \lambda \right) \quad (4)$$

## 3 Drawback: Free Riding

In this section, we analyze the effect of reward sharing on the effort level of the agents in a model with no information asymmetry, such that all agents observe  $R$  perfectly. The *effort sharing function*

---

<sup>2</sup>This model of agent behavior is related to formal models in behavioral game theory; e.g., the *Quantal Level  $k$  (QL- $k$ ) model* [11]. In these models, agents make (different, depending on their level) simplifying assumptions about the reasoning of other agents. In our MLE best response model, agents make simplifying assumptions about the beliefs of other agents about the reward of a task.

$f$  is a recursive function, which is computed at the leaf first and then is computed on the nodes above till the root, for a given tree  $T$  and a reward sharing matrix  $\Delta$ , defined as follows. Let us denote the set of single-rooted directed trees with  $n$  nodes by  $\mathcal{T}$ .

**Definition 1 (Effort Sharing Function)** *An effort sharing function is a mapping  $f : \mathcal{T} \times [0, 1)_{n \times n} \rightarrow [0, 1]$  given by the recursive formula,*

$$f(T_i, \Delta) = \max \left\{ 0, 1 - \frac{1}{\gamma} \sum_{j \in T_i \setminus \{i\}} \delta_{ij} \cdot f(T_j, \Delta) \right\}, \quad (5)$$

where  $T_i$  is the subtree rooted at  $i$ , which includes  $i$ .

This function is maximum when  $i$  is a leaf, and decreases as we move towards the root. In Figure 4, we plot  $f(T_i, \Delta)$  as a function of the size of the subtree  $T_i$ . The function decreases as the size of the subtree  $T_i$  increases. The following theorem states that the equilibrium effort level of each node in a hierarchy is proportional to this function, leading to a smaller effort levels of nodes near the root.

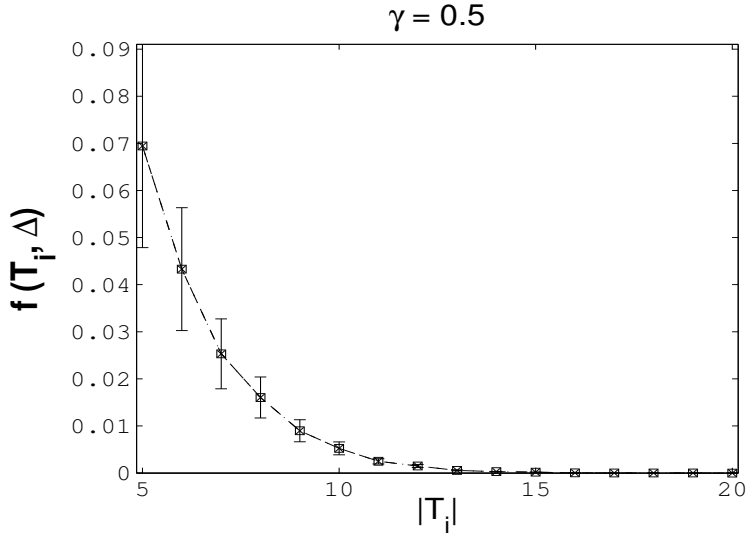


Figure 4: Plot of  $f(T_i, \Delta)$  ( $\Delta$  averaged) versus  $|T_i|$ .

**Theorem 1 (Nash Equilibrium Characterization)** *Under the perfect observation of the reward  $R$ , the Nash equilibrium effort profile  $\lambda^*$ , is uniquely given by,*

$$\lambda_i^* = \frac{\lambda \gamma R}{C} \left( \frac{\sum_{j \in N} f(T_j, \Delta) - 1}{(\sum_{j \in N} f(T_j, \Delta))^2} \right) f(T_i, \Delta), \forall i \in N, \quad (6)$$

and lies in  $Z_1 = \left\{ \lambda : \sum_{j \in N} \lambda_j > \lambda \right\}$ , if and only if,

$$\frac{\gamma \cdot R}{C} > \frac{\sum_{j \in N} f(T_j, \Delta)}{\sum_{j \in N} f(T_j, \Delta) - 1} \quad (7)$$

**Proof:** ( $\Rightarrow$ ) Substituting (6) in  $\sum_{j \in N} \lambda_j > \lambda$  we obtain  $\frac{\gamma \cdot R}{C} > \frac{\sum_{j \in N} f(T_j, \Delta)}{\sum_{j \in N} f(T_j, \Delta) - 1}$ .  
( $\Leftarrow$ ) We show this in two steps.

**Step 1:** Unique PSNE effort profile: If a PSNE effort profile  $(\lambda_i^*, \lambda_{-i}^*)$  exists in the given game, then it must satisfy,  $u_i(\lambda_i^*, \lambda_{-i}^*) \geq u_i(\lambda_i, \lambda_{-i}^*), \forall \lambda_i \in S_i = [0, \infty), \forall i \in N$ . This implies,  $\lambda_i^* = \arg \max_{\lambda_i \in S_i = [0, \infty)} u_i(\lambda_i, \lambda_{-i}^*), \forall i \in N$ .

Thus in order to find the Nash equilibrium we have to solve the following optimization problem for each  $i \in N$ .

$$\left. \begin{array}{l} \max_{\lambda_i} \quad u_i(\lambda_i, \lambda_{-i}^*) \\ \text{s.t.} \quad \lambda_i \geq 0, \end{array} \right\} \Rightarrow \left. \begin{array}{l} \min_{\lambda_i} \quad -u_i(\lambda_i, \lambda_{-i}^*) \\ \text{s.t.} \quad -\lambda_i \leq 0. \end{array} \right\} \quad (8)$$

Due to concavity of  $u_i$ , this is a convex optimization problem with linear constraints, which can be solved using KKT theorem. At the minimizer  $\lambda_i^*$  of problem (8),  $\exists \mu \in \mathbb{R}$  such that, (i)  $\mu \geq 0$ , (ii)  $-\frac{\partial}{\partial \lambda_i} u_i(\lambda_i^*, \lambda_{-i}^*) - \mu = 0$ , (iii)  $-\mu \lambda_i^* = 0$ , (iv)  $-\lambda_i^* \leq 0$ .

Case 1:  $\mu > 0 \Rightarrow \lambda_i^* = 0$  and in this case  $\frac{\partial}{\partial \lambda_i} u_i = -\mu \leq 0$ .

Case 2:  $\mu = 0 \Rightarrow \frac{\partial}{\partial \lambda_i} u_i(\lambda_i^*, \lambda_{-i}^*) = 0$  and in this case  $\lambda_i^* \geq 0$ . This leads us to,

$$\left( \sum_{j \in N} \lambda_j^* \right)^2 = \frac{\lambda \gamma R}{C} \left( \sum_{j \neq i} \lambda_j^* - \sum_{j \in T_i \setminus \{i\}} \lambda_j^* \delta_{ij} \right). \quad (9)$$

The above expression is obtained by differentiating  $u_i$ .

For a given tree and its equilibrium profile  $\lambda^*$ , let us substitute  $x$  for  $\sum_{j \in N} \lambda_j^*$ , then manipulation of (9) leads to,

$$\lambda_i^* + \sum_{j \in T_i \setminus \{i\}} \lambda_j^* \delta_{ij} = x - \frac{x^2 C}{\lambda \gamma R}, \quad \forall i \in N. \quad (10)$$

We do another variable substitution to denote the RHS of Eqn (10) by  $y$  ( $\geq 0$  since LHS is  $\geq 0$ ). That is,

$$y = x - \frac{x^2 C}{\lambda \gamma R}. \quad (11)$$

**Claim:**  $\lambda_i^* = y f(T_i, \Delta), \forall i \in N$ .

**Proof:** We prove this claim via induction on the levels of  $T$ . Let the depth of  $T$  be  $D$ . From Eqn (10),  $\lambda_i^* + \sum_{j \in T_i \setminus \{i\}} \delta_{ij} \lambda_j^* = y, \forall i \in N$ . From Cases 1 and 2 above,

$$\lambda_i^* = \max \left( 0, y - \sum_{j \in T_i \setminus \{i\}} \delta_{ij} \lambda_j^* \right) \quad \forall i \in N. \quad (12)$$

*Step 1:* For an arbitrary node  $j$  at level  $D$ , from (12),  $\lambda_j^* = y$ . Hence, the proposition is true as  $f(T_j, \Delta) = 1$  for a leaf. Now, select an arbitrary node  $i$  (which is not a leaf) at level  $D - 1$ . From (12) we get,  $\lambda_i^* = \max(0, y - \sum_{j \in T_i \setminus \{i\}} \delta_{ij} y) = y \max(0, 1 - \sum_{j \in T_i \setminus \{i\}} \delta_{ij} 1) = y f(T_i, \Delta)$ .

*Step 2:* Let  $\lambda_j^* = y f(T_j, \Delta)$  be true for all nodes  $j$  upto level  $D - l$ . Consider an arbitrary node  $i$  at level  $D - l - 1$ . From (12) and (5),

$$\lambda_i^* = \max \left( 0, y - \sum_{j \in T_i \setminus \{i\}} y \cdot f(T_j, \Delta) \delta_{ij} \right) = y f(T_i, \Delta) \quad (13)$$

which concludes the induction.

To find an expression for PSNE we now evaluate  $y$ . The sum of efforts of all the players is defined as  $x$ . Hence,  $x = y \sum_{j \in N} f(T_j, \Delta)$ . Substituting for  $y$  from (11) in this expression and solving for  $x$  yields,

$$x = \sum_{j \in N} \lambda_j^* = \frac{\lambda \gamma R}{C} \left( \frac{\sum_{j \in N} f(T_j, \Delta) - 1}{\sum_{j \in N} f(T_j, \Delta)} \right). \quad (14)$$

Using (11) we get,

$$y = \frac{\lambda \gamma R}{C} \left( \frac{\sum_{j \in N} f(T_j, \Delta) - 1}{\left( \sum_{j \in N} f(T_j, \Delta) \right)^2} \right). \quad (15)$$



Combining (15) and the claim above, the PSNE is given by,

$$\lambda_i^* = \frac{\lambda \gamma R}{C} \left( \frac{\sum_{j \in N} f(T_j, \Delta)^{-1}}{(\sum_{j \in N} f(T_j, \Delta))^2} \right) f(T_i, \Delta), \forall i \in N.$$

KKT equations led to a unique solution of the optimization problem, hence PSNE is unique.

**Step 2:**  $\sum_{j \in N} \lambda_j > \lambda$ : We use Eqn (6) to compute the sum  $\sum_{j \in N} \lambda_j^*$ , and use the fact that  $\gamma R/C > \frac{\sum_{j \in N} f(T_j, \Delta)}{\sum_{j \in N} f(T_j, \Delta)^{-1}}$  to get  $\sum_{j \in N} \lambda_j^* > \lambda$ .

### 3.1 Example

To get a feel for how free riding can affect the total effort in the system, let us look at the specific example of of Figure 5. We consider the geometric reward sharing with factor  $\delta$  (see §2.2), and plot the equilibrium sum effort level with increasing  $\delta$ . If  $\delta \approx 0.4$ , and no free riding happens (that is, pretend  $\delta = 0$  in the figure), in order to reach a sum effort level of 8 the organization would have needed a reward of  $R = 10$ . However, because of free riding phenomenon, to attain the same sum effort level, one needs to put  $R = 20$ , see the circled curve on the top at  $\delta \approx 0.4$ .

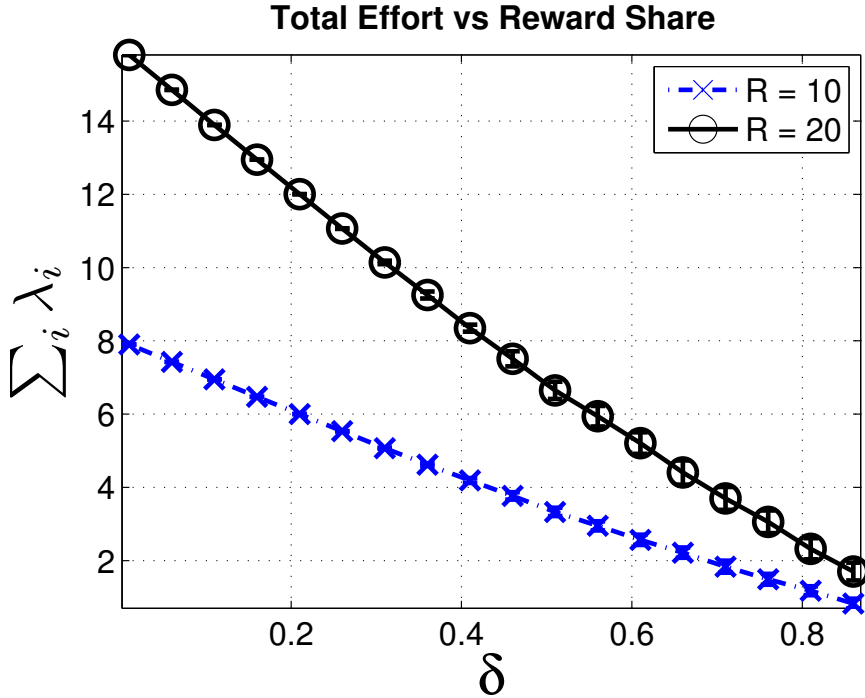


Figure 5: Plot of  $\sum_{i \in N} \lambda_i^*$  versus the geometric reward share factor,  $\delta$ .

## 4 Benefit: Information Sharing

In this section, we discuss the benefit of placing asymmetrically informed nodes in a hierarchy, for now studying this in the absence of free riding behavior. The goal is to understand one of the ‘pure’ advantages of having a hierarchy. In particular, while agents still aim to maximize their payoffs, we restrict them to ignore the reward share coming from the subtree below them (and thus the indirect reward).

We assume that the reward  $R$  is sufficient to guarantee  $\gamma R/C > \frac{n}{n-1}$  ( $n$  is the number of nodes) with high probability (i.e. approaches 1 in the limit). Since we ignore indirect rewards in this section,

agent  $i$  perceives the utility:

$$u_i(\lambda_i, \lambda_{-i}) = \lambda\gamma R_i \frac{\lambda_i}{\sum_{k \in N} \lambda_k} - C\lambda_i.$$

The above utility is a modification of Equation (1), which does not contain passive reward terms, due to our assumption that the nodes do not free ride. Also the reward  $R$  is replaced by  $R_i$  because of the MBR agent behavior model.

Taking the partial derivative of the above equation w.r.t.  $\lambda_i$  and equating it to zero, and solving the set of equations similar to the ones in §3, we get,

$$\lambda_i^{MBR} = \frac{\lambda\gamma R_i}{C} \frac{n-1}{n^2} \Rightarrow \sum_{i \in N} \lambda_i^{MBR} = \frac{\lambda\gamma \sum_i R_i}{C} \frac{n-1}{n^2}.$$

Notice that,  $R_i$ 's are conditionally independent given  $R$ , and are distributed as  $\mathcal{N}(0, \sigma_i^2(t))$ . Hence, the sum effort  $\sum_{i \in N} \lambda_i^{MBR}$ , is also distributed as a Gaussian random variable given  $R$ , with mean  $m_{\text{total}}$  and variance  $\sigma_{\text{total}}^2(t)$ , such that,

$$\begin{aligned} m_{\text{total}} &= \frac{\lambda\gamma R}{C} \frac{n-1}{n}, \\ \sigma_{\text{total}}^2(t) &= \left(\frac{\lambda\gamma}{C}\right)^2 \left(\frac{n-1}{n^2}\right)^2 \sum_i \sigma_i^2(t). \end{aligned} \quad (16)$$

Since,  $\gamma R/C > \frac{n}{n-1}$ , it implies,  $m_{\text{total}} > \lambda$ . Hence, using the upper bound on the tail distribution of the standard normal distribution, we can show the following bound on the risk:

$$\mathbb{P}\left(\sum_{i \in N} \lambda_i^{MBR} \leq \lambda\right) = O\left(\frac{\sigma_{\text{total}}(t)}{\epsilon} e^{-\epsilon^2/2\sigma_{\text{total}}^2(t)}\right), \quad (17)$$

where,  $\epsilon = m_{\text{total}} - \lambda > 0$ . To minimize risk, we want the RHS of the above equation to be small. Let us define the *risk bound* ignoring free riding (IFR) as,

$$\rho_{\text{IFR}}(t) := \frac{\sigma_{\text{total}}(t)}{\epsilon} e^{-\epsilon^2/2\sigma_{\text{total}}^2(t)}. \quad (18)$$

The risk is bounded by a constant factor of the above quantity. This is a function of the total variance  $\sigma_{\text{total}}^2(t)$ , which has been calculated ignoring the free riding effect in this section, and hence this expression does not consider free riding effect. Since  $\rho_{\text{IFR}}(t)$  is a function of time, we should consider its rate of fall. However, if the magnitude of the risk bound remains large, then the rate alone is not a good measure of how the actual risk falls. Therefore, a good metric to quantify both the rate and magnitude of the risk is the fractional rate,  $\frac{1}{\rho_{\text{IFR}}(t)} \frac{d\rho_{\text{IFR}}(t)}{dt}$ .

For this reason, we formulate a hierarchy design problem as the problem of finding the tree  $T_{\min}$  that minimizes the fractional rate of fall of the risk bound at time  $t = 0$ . Formally, denoting the set of all single rooted trees with  $n$  nodes by  $\mathcal{T}$ , we get the following *Optimization problem Ignoring Free Riding (OPT-IFR)*:

$$T_{\min} \in \arg \min_{T \in \mathcal{T}} \frac{1}{\rho_{\text{IFR}}(t)} \frac{d\rho_{\text{IFR}}(t)}{dt} \Bigg|_{t=0} \quad \text{OPT-IFR} \quad (19)$$

The fractional rate of fall of  $\rho_{\text{IFR}}(t)$  at time  $t = 0$  can be shown to be:

$$\frac{1}{\rho_{\text{IFR}}(t)} \frac{d\rho_{\text{IFR}}(t)}{dt} \Big|_{t=0} = \frac{1}{\sigma_{\text{total}}(0)} \left( 1 + \frac{\epsilon^2}{\sigma_{\text{total}}^2(0)} \right) \cdot \frac{d\sigma_{\text{total}}^2(t)}{dt} \Big|_{t=0}$$

The rate of fall in the total variance is given by,

$$\begin{aligned} \frac{d\sigma_{\text{total}}^2(t)}{dt} \Big|_{t=0} &\propto \sum_{i \in N} \frac{d\sigma_i^2(t)}{dt} \Big|_{t=0} \\ &= \sum_{i \in N} \mu(C_{p(i)}) (\sigma_{p(i)}^2(t) - \sigma_i^2(t)) \Big|_{t=0} \\ &= \sum_{i \in N} \mu(C_{p(i)}) (\sigma_{p(i)}^2 - \sigma_i^2). \end{aligned} \tag{20}$$

The first proportionality comes from the Equation (16). The second equality comes due to the information propagation model (Equation 2). The parent of  $i$  is denoted by  $p(i)$  (and unique due to the tree structure).

### 4.1 Example

Under this model, the benefit of the hierarchy comes from the fact that it helps reduce the uncertainty levels of the employees, and thereby reduces the risk. Figure 6 shows a typical hierarchy formed using Equation (19) with 6 nodes with a typical set of intrinsic variances. We will show in the next section that the hierarchy changes when we consider both information propagation and free riding of the nodes jointly into account.

**Ignoring FR, Rate of Fall = -0.3466,  $\delta = 0.25$**

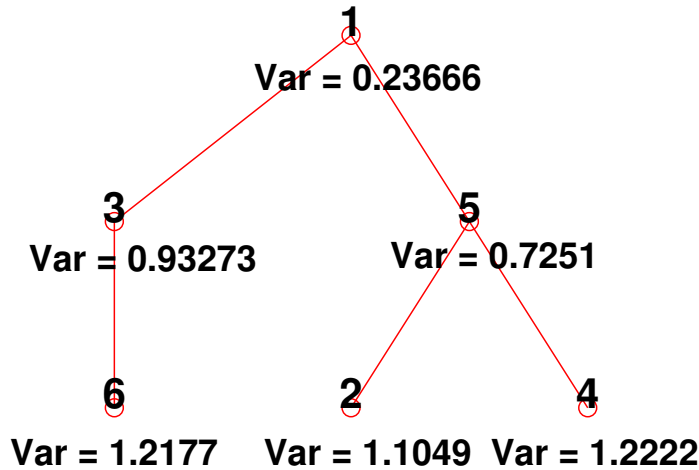


Figure 6: Hierarchy formed when free riding is not a concern.

## 5 Trade-off between Free-riding and Information Sharing

In the previous two sections, we have addressed separately the issues of free-riding and information sharing. In an actual hierarchical organization, both these effects take place simultaneously, making it necessary to consider both aspects while designing a hierarchy.

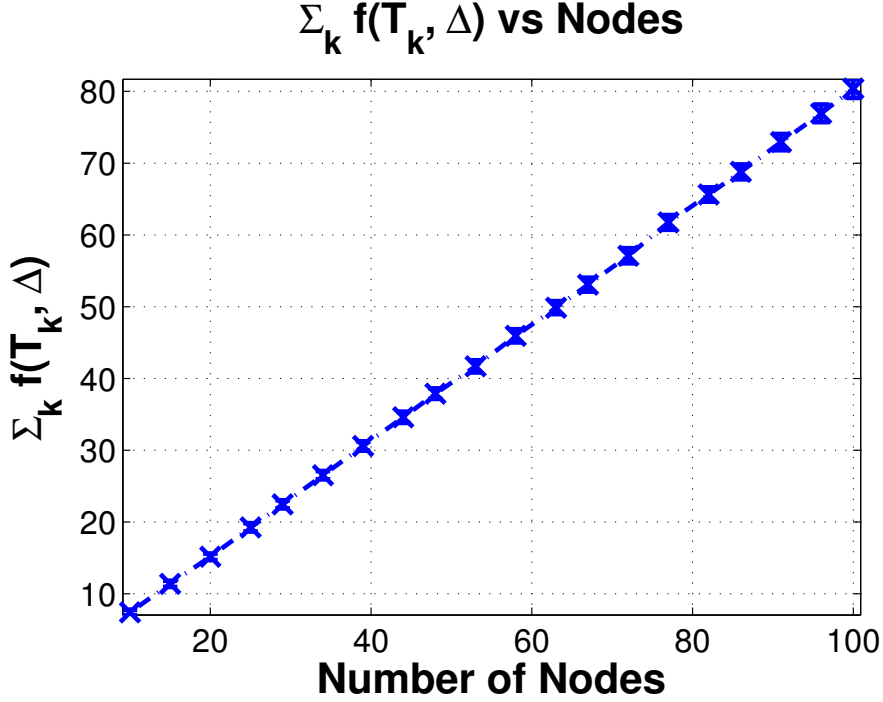


Figure 7: Plot of  $\sum_{i \in N} f(T_i, \Delta)$  ( $\Delta$  averaged) vs  $n$ .

We consider hierarchies with  $n$  nodes. The sum,  $\sum_{i \in N} f(T_i, \Delta)$  is increasing in  $n$  (see Fig. 7). Hence,  $\frac{\sum_{i \in N} f(T_i, \Delta)}{\sum_{i \in N} f(T_i, \Delta) - 1} \downarrow n$  and approaches 1 in the limit. If we assume that the potential reward  $R$  is distributed such that with high probability (w.h.p.),  $\gamma R/C \geq (1 + \epsilon)$ , where  $\epsilon > 0$  is any fixed number, then for large enough  $n$ , the parameters will satisfy Eq. (7) w.h.p.

**Lemma 1** For a population of agents with MBR behavioral model, the MBR effort of agent  $i$  is given by,

$$\lambda_i^{MBR} = \frac{\lambda \gamma R_i}{C} \left( \frac{\sum_{j \in N} f(T_j, \Delta) - 1}{(\sum_{j \in N} f(T_j, \Delta))^2} \right) f(T_i, \Delta), \quad \forall i \in N, \forall T. \quad (21)$$

**Proof:** By definition of the MBR model (§2.4), an agent  $i$  makes the MLE estimate of the reward observed by agent  $j$  given her own observed reward  $R_i$ , which is  $R_i$ . Hence, the utility of agent  $i$  is given by Equation (1), with  $R$  replaced by  $R_i$ .

$$u_i(\lambda_i, \lambda_{-i}) = \lambda \gamma R_i \frac{\lambda_i}{\sum_{k \in N} \lambda_k} + \lambda R_i \sum_{j \in T_i \setminus \{i\}} \delta_{ij} \frac{\lambda_j}{\sum_{k \in N} \lambda_k} - C \lambda_i. \quad (22)$$

To maximize this w.r.t.  $\lambda_i$ , we follow a similar analysis technique as in the proof of Theorem 1, by writing down the constrained optimization problem and the KKT conditions, and solving the set of equations we get the expression of Equation (21).

Note: Lemma 1 does not claim the equilibrium result given by Theorem 1, but uses the same analysis technique to solve the set of equations and find the best response effort of agent  $i$ , given by  $\lambda_i^{MBR}$ , that maximizes the utility in Equation (22).

The following theorem presents the distribution of the total effort of the nodes in a hierarchy, and gives the expressions for mean and variance.

**Theorem 2** *The sum effort of a population of agents with MBR behavior under the information asymmetry and reward sharing model on a hierarchy  $T$ , is distributed as  $\mathcal{N}(m_{\text{total}}, \sigma_{\text{total}}^2(t))$ , where for large  $n$ ,  $m_{\text{total}} \approx \frac{\lambda\gamma R}{C}$ , and  $\sigma_{\text{total}}^2(t) = \Theta\left(\frac{1}{n^2} \sum_{i \in N} f^2(T_i, \Delta) \sigma_i^2(t)\right)$ .*

**Proof:** From Lemma 1 we get,

$$\begin{aligned} \sum_{i \in N} \lambda_i^{MBR} &= \sum_{i \in N} \frac{\lambda\gamma R_i}{C} \left( \frac{\sum_{j \in N} f(T_j, \Delta) - 1}{(\sum_{j \in N} f(T_j, \Delta))^2} \right) f(T_i, \Delta) \\ &= \sum_{i \in N} \frac{\lambda\gamma}{C} \left( \frac{\sum_{j \in N} f(T_j, \Delta) - 1}{(\sum_{j \in N} f(T_j, \Delta))^2} \right) f(T_i, \Delta) \cdot (R + \eta_i) \end{aligned} \quad (23)$$

Conditioned upon the true reward  $R$ , this sum is also distributed as Gaussian with mean and variance as follows.

$$\begin{aligned} m_{\text{total}} &:= \mathbb{E} \left[ \sum_{i \in N} \lambda_i^{MBR}(T, R_i) \mid R \right] \\ &= \frac{\lambda\gamma R}{C} \frac{\sum_{j \in N} f(T_j, \Delta) - 1}{\sum_{j \in N} f(T_j, \Delta)} \end{aligned} \quad (24)$$

$$\begin{aligned} \sigma_{\text{total}}^2(t) &:= \text{Var} \left[ \sum_{i \in N} \lambda_i^{MBR}(T, R_i) \mid R \right] \\ &= \left( \frac{\lambda\gamma}{C} \right)^2 \left( \frac{\sum_{j \in N} f(T_j, \Delta) - 1}{(\sum_{j \in N} f(T_j, \Delta))^2} \right)^2 \sum_{i \in N} f^2(T_i, \Delta) \sigma_i^2(t) \end{aligned} \quad (25)$$

For large enough  $n$ ,  $\frac{\sum_{j \in N} f(T_j, \Delta) - 1}{\sum_{j \in N} f(T_j, \Delta)} \approx 1$ , and,

$$\left( \frac{\sum_{j \in N} f(T_j, \Delta) - 1}{(\sum_{j \in N} f(T_j, \Delta))^2} \right)^2 = \Theta(1/n^2).$$

Hence, we have proved the theorem.

To minimize the risk of the organization, the approach of minimizing the risk bound (Equation (18)) is followed. However, the risk bound considering free riding (CFR) is different from that of Equation (19) since the total variance  $\sigma_{\text{total}}^2(t)$  now considers the free riding effect and is different (given by Equation (25) as opposed to Equation (16)). The modified risk bound is:

$$\rho_{\text{CFR}}(t) := \frac{\sigma_{\text{total}}(t)}{\epsilon} e^{-\epsilon^2/2\sigma_{\text{total}}^2(t)}. \quad (26)$$

The modified design goal is to identify the hierarchy that minimizes the fractional fall rate of the risk bound at time  $t = 0$ . Hence the *Optimization problem Considering Free Riding (OPT-CFR)* is given by:

$$T_{\min} \in \arg \min_{T \in \mathcal{T}} \frac{1}{\rho_{\text{CFR}}(t)} \frac{d\rho_{\text{CFR}}(t)}{dt} \Big|_{t=0} \quad \text{OPT-CFR} \quad (27)$$

Following similar steps to the analysis following Equation (19), and using the expression of  $\rho_{\text{CFR}}(t)$  from Equation (26), we can show that the RHS of the above optimization problem is given by,

$$\frac{1}{\rho_{\text{CFR}}(t)} \frac{d\rho_{\text{CFR}}(t)}{dt} \Big|_{t=0} = \frac{1}{\sigma_{\text{total}}(0)} \left( 1 + \frac{\epsilon^2}{\sigma_{\text{total}}^2(0)} \right) \cdot \frac{d\sigma_{\text{total}}^2(t)}{dt} \Big|_{t=0}$$

## Considering FR, Rate of Fall = $-0.39759$ , $\delta = 0.25$

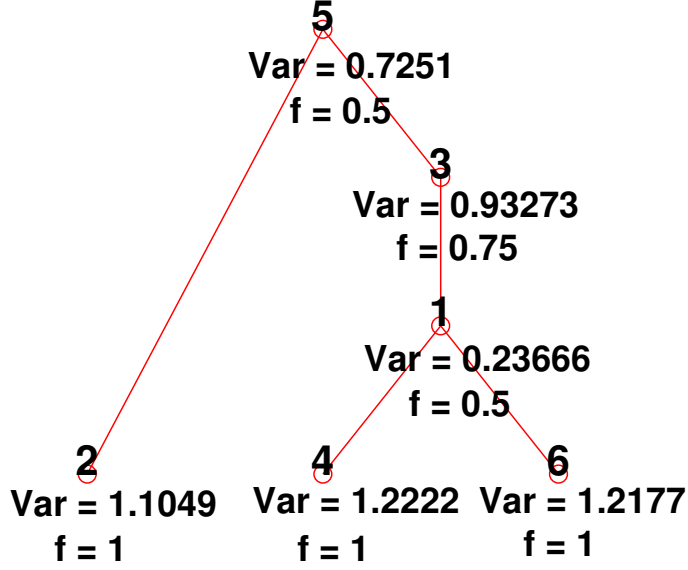


Figure 8: Hierarchy formed considering free riding.

Taking the derivative of the total variance given by Theorem 2, and following the steps as in Equation (25), we have:

$$\left. \frac{d\sigma_{\text{total}}^2(t)}{dt} \right|_{t=0} \propto \sum_{i \in N} f^2(T_i, \Delta) \mu(C_{p(i)}) (\sigma_{p(i)}^2 - \sigma_i^2). \quad (28)$$

We now see that the term accounting for the free riding effect,  $f^2(T_i, \Delta)$ , appears in the sum in the above equation, which makes it different from Equation (20). The rate of change of the total variance is now a weighted sum of the squared effort sharing function  $f$ , and therefore the output of OPT-CFR will potentially be a different hierarchy from the OPT-IFR. In the following section we show the difference between the hierarchies found by these two optimization problems.

### 5.1 Example

To show how the structure of the hierarchy changes when one designs a network considering the free riding effect, we take the same set of nodes as in Figure 6, but now solve the optimization problem given by Equation (27), and find the hierarchy as shown in Figure 8.

We compare the performances of the OPT-IFR and OPT-CFR. Figure 9 plots the fractional fall rate w.r.t. the amplitude factor of the function  $\mu$  and shows that the fall rate is faster in OPT-CFR. Figure 10 shows that the difference in the fall rate between OPT-CFR and OPT-IFR is around 10% for 5 nodes and increases with the number of nodes.

## 6 An Algorithm for Finding the Near Optimal Hierarchy

Equation (27) gives a method of jointly optimizing over the free riding and information propagation phenomena in hierarchies. It is natural to ask how one can solve the optimization problem to find the optimal hierarchy. An exhaustive search algorithm would iterate over the space of all single rooted

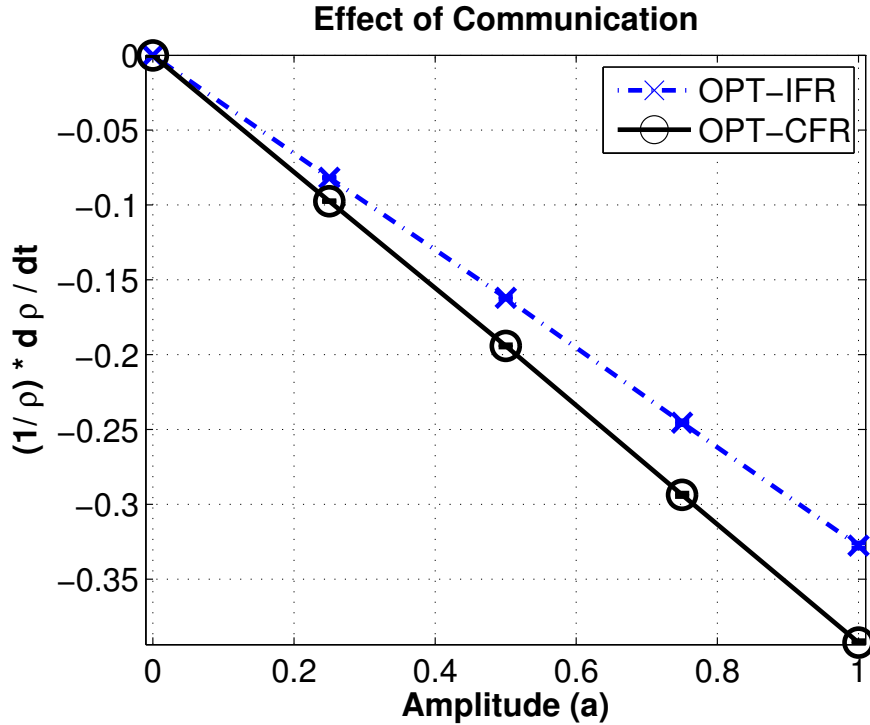


Figure 9: Absolute difference in the fractional decay rate of risk bound in OPT-CFR with OPT-IFR.

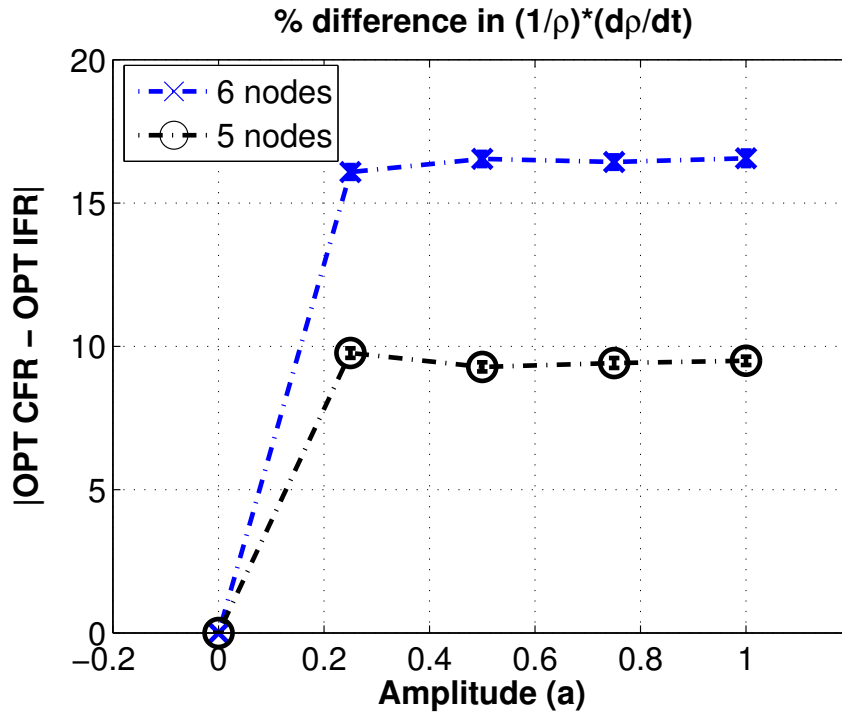


Figure 10: Percentage difference in the fractional decay rate of risk bound in OPT-CFR with OPT-IFR.

directed trees with  $n$  nodes. To find the optimal hierarchies in the simulations of this paper, we have used the exhaustive search.

However, a simple *sequential greedy* algorithm, given in tabular form in Algorithm 1, runs in  $O(n^3)$  time and gives reasonably good approximation to the optimal hierarchy. Figure 11 shows the performance comparison of the rate of fall of the fractional risk bound given by the greedy algorithm with that of the optimal hierarchy for 6 nodes with varying average noise variance of the population.

---

**Algorithm 1** Sequential Greedy

---

- 1: **Input:** A set of  $n$  nodes with intrinsic variances  $\sigma_i^2$ 's
  - 2: **Output:** A tree connecting the nodes
  - 3: **Step 1:** Sort nodes in ascending order of variances
  - 4: **Step 2:**
  - 5: **for** Both configuration of *the first two nodes* **do**
  - 6:   **for** Nodes  $j \leftarrow 3$  to  $n$  **do**
  - 7:     Sequentially add  $j$  as a child of one of the previous nodes, such that the total variance of the subnetwork after adding  $j$  is minimum
  - 8:   **end for**
  - 9: **end for**
  - 10: **Step 3:** Output the tree giving the lowest total variance with  $n$  nodes
- 

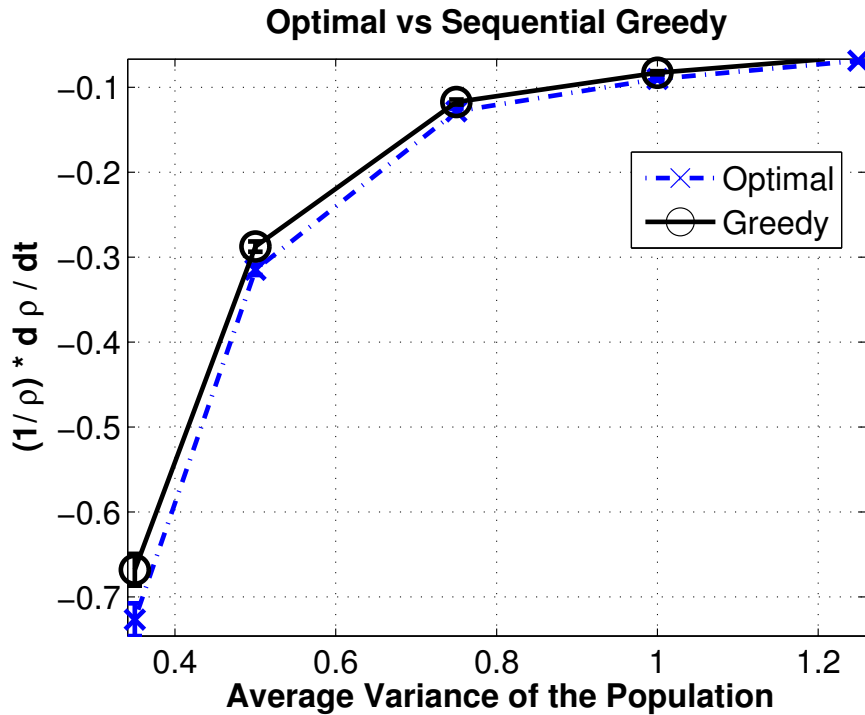


Figure 11: Fractional decay rate of the risk bound  $\rho_{\text{CFR}}(t, T)$  comparison between the Optimal and Sequential Greedy algorithms (Algorithm 1). The x-axis shows different average variances of the population. Number of nodes = 6.

## 7 Conclusions

We have studied organizational hierarchies to understand the effects of competition, information asymmetry, and reward sharing on the performance of organizations, for a simple queuing-theoretic model. Our analysis has highlighted a trade-off between the inefficiencies that come from free riding and the benefits that come from information sharing, and thus improving the accuracy of agent models for the reward of tasks. We have demonstrated the effect of considering both phenomena on the design of organizational hierarchies. There are many future avenues to this research, including: (a) extending the model to consider the cost of hiring an employee, (b) looking to capture competition amongst agents local to each other on the hierarchy rather global competition, and (c) considering a model with heterogeneous worker abilities.



## References

- [1] D. P. Bertsekas and R. Gallager. *Data Networks*. Prentice Hall, Englewood Cliffs, 1992.
- [2] G. D. Bhatt. Knowledge management in organizations: examining the interaction between technologies, techniques, and people. *Journal of knowledge management*, 5(1):68–75, 2001.
- [3] Y. Bramoullé and R. Kranton. Public Goods in Networks. *Journal of Economic Theory*, 135(1):478 – 494, 2007.
- [4] K. Eisenhardt. Agency-and institutional-theory explanations: the case of retail sales compensation. *Academy of Management journal*, 31(3):488–511, 1988.
- [5] B. A. Gerhart, H. B. Minkoff, and R. N. Olsen. Employee Compensation: Theory, Practice, and Evidence. *CAHRS Working Paper Series*, page 194, 1995.
- [6] M. K. Gibson and M. J. Papa. The Mud, the Blood, and the Beer Guys: Organizational Osmosis in Blue-Color Work Groups. *Journal of Applied Communication Research*, 28:68–88, 2000.
- [7] D. Mookherjee. Incentives in Hierarchies. preliminary version, prepared for Handbook of Organizational Economics, 2010.
- [8] R. Radner. Hierarchy: The Economics of Managing. *Journal of Economic Literature*, pages 1382–1415, 1992.
- [9] E. Ravasz and A.-L. Barabási. Hierarchical organization in complex networks. *Physical Review E*, 67(2):026112, 2003.
- [10] H. Simon. Bounded Rationality and Organizational Learning. *Organization Science*, 2:125 – 134, 1991.
- [11] D. O. Stahl and P. W. Wilson. Experimental Evidence on Players’ Models of Other Players. *Journal of Economic Behavior & Organization*, 25:309 – 327, 1994.
- [12] N. M. Tichy, M. L. Tushman, and C. Fombrun. Social Network Analysis for Organizations. *Academy of Management Review*, pages 507–519, 1979.
- [13] O. E. Williamson. The economics of organization: the transaction cost approach. *American Journal of Sociology*, 87:548 – 577, 1981.
- [14] R. W. Wolff. *Stochastic Modeling and the Theory of Queues*. Prentice Hall, Englewood Cliffs, New Jersey 07632, 1989.